Jornadas de Seguimiento de Proyectos en Tecnologías Informáticas

José María Troya y Sascha Ossowski
(editores)

Prólogo

Las convocatorias de proyectos del Plan Nacional de Ciencia y Tecnología incluyen la obligación por parte de la Subdirección de Proyectos de Investigación de efectuar el seguimiento y evaluación de los resultados de cada proyecto subvencionado. Siguiendo el espíritu de esta norma, el Programa Nacional de Tecnologías Informáticas (TIN) realiza periódicamente unas jornadas de seguimiento de proyectos que tienen como fin tanto la mencionada evaluación como la difusión de las actividades en este área a otros científicos y al sector industrial.

Las presentes actas recopilan los informes remitidos por los coordinadores e investigadores responsables de los proyectos que participaron en las jornadas de seguimiento del año 2007. Ofrecen una panorámica de la investigación nacional a través del amplio abaníco de proyectos concedidos en la convocatoria 2004. Las jornadas se celebraron en Móstoles (Madrid) los días 15 y 16 de Febrero de 2007.

Deseamos agradecer su participación a los miembros del Comité Técnico que han constituido los dos paneles en que se ha organizado el seguimiento. Dicho Comité estuvo compuesto por diferentes investigadores provenientes tanto del sector académico como industrial, incluyendo algunos miembros de otros países. Asimismo, queremos dar las gracias a todos los miembros del comité organizador local, cuyo trabajo ha sido clave para la realización de las jornadas.

Febrero 2007

José María Troya y Sascha Ossowski
Programa Nacional de Tecnologías Informáticas

Organización

Las Jornadas de Seguimiento de Proyectos en Tecnologías Informáticas han sido organizadas por el equipo de gestión del Programa Nacional de Tecnologías Informáticas (TIN) de la Subdirección General de Proyectos de Investigación del Ministerio de Educación y Ciencia. Queremos dar las gracias a la Universidad Rey Juan Carlos por ceder los espacios para las jornadas, y al comité local de organización, presidido por el profesor Holger Billhardt, por el apoyo logístico recibido.

Programa Nacional de Tecnologías Informáticas

José María Troya Linero (Gestor del Programa TIN)
Sascha Ossowski (Colaborador del Programa TIN)
Ramón López de Arenosa (Jefe del Dep. de Tecnologías de la Producción y las Comunicaciones)

Miembros de los paneles

Antonio Brogi (Universidad de Pisa)
Francisco J. Garijo (Telefónica I+D)
Carlos A. Iglesias Fernández (Germinus XXI)
Paolo Petta (Instituto de Investigación Austríaco para la IA)
Juan José Villanueva Pipaon (Universidad Autónoma de Barcelona)
Francisco Tirado Fernández (Universidad Complutense de Madrid)

Índice

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Técnicas emergentes de minería de datos para la extracción de conocimiento de grandes volúmenes de información: aplicación a datos científicos e industriales TIN2004-00159

Jesús S. Aguilar-Ruiz[*]
Universidad Pablo de Olavide

José C. Riquelme[**]
Universidad de Sevilla

Abstract

The main aim of this project is to design a set of tools to extract knowledge from numerical information, which is provided by companies that usually handle huge amount of examples with high dimensionality. This aim consists mainly in addressing three research directions: 1) to improve current machine learning algorithms with medium-size datasets; 2) to extract useful knowledge from large-size datasets by using the previous experience and research results; 3) finally, to apply the tools developed in the first two items to three real-world problems provided by local companies, which the research group keeps in contact with for the last years: an metallurgical process industry, a biotechnology company, and the Spanish electricity market.

Keywords: Data Mining, Machine Learning, Bioinformatics

1 Project goals

Two main goals are going to be addressed, which can be categorized in two research directions: a) Development of new techniques and methodologies for the extraction of knowledge from data; b) Application of such techniques to real-world problems.

Goals in machine learning and data mining: We aim for finding solutions to extract information and knowledge from data with a great size and dimensionality:

- Methods to reduce the data size and dimensionality by means of editing or feature selection techniques based on heuristics with low computational cost.
- To improve the encoding and evaluation of Evolutionary Algorithms for the extraction of decision rules. In this way, we could increase the data volume handled by previous tools.
- To develop a software tool for the extraction and visualization of decision rules from data streams.

[*] Email: aguilar@upo.es
[**] Email: riquelme@us.es

Targets:

- Applications to industrial data: those provided by Atlantic Copper, located at the industrial estate of Huelva, which devotes itself to produce copper.
- Application to biological data: provided by NBT, coming from microarray gene expression analysis, containing hundreds of examples and several thousands of feature.
- Application to predictions from time series of demand and price of electricity.

2 Level of success

2.1 Evolutionary Computation

One of the most influential factors in the quality of the solutions found by an evolutionary algorithm is a correct coding of the search space and an appropriate evaluation function of the potential solutions. In this research, the coding of the search space for the obtaining of those decision rules is approached, i.e., the representation of the individuals of the genetic population and also the design of specific genetic operators. Our approach, called "natural coding", uses one gene per feature in the dataset (continuous or discrete). The examples from the datasets are also encoded into the search space, where the genetic population evolves, and therefore the evaluation process is improved substantially. Genetic operators for the natural coding are formally defined as algebraic expressions. Experiments show that as the genetic operators are better guided through the search space, the number of rules decreases considerably while maintaining the accuracy. The knowledge models obtained are simpler, with very few decision rules, and therefore easier to understand, which is an advantage in many domains. Main results: [3], [11].

With respect to speeding-up the EA evaluation, we have worked in a novel data structure (called EES) with two goals: to incorporate knowledge to the EA to be used by genetic operators and to reduce the time complexity of the EA by means of a fast evaluation of examples from the dataset. EES incorporates knowledge in the early stage, when the structure is built before running the EA, by organizing the discrete attributes and by obtaining potentially good intervals for continuous attributes. The main feature of the ESS is the possibility of accessing the information from the dataset through attributes instead of through examples. This means that EES organizes the information in such a way that it is not necessary to process all the examples to evaluate individuals (candidate decision rules) from the genetic population generated by a supervised learning system. Results show that EES achieves to reduce the computational cost about 50%, maintaining the quality of decision rules generated. Therefore, the approach guarantees the accuracy of the solutions taking less computational resources. Main publications: [1], [29].

2.2 Feature Selection

We present a new heuristic to select relevant feature subsets in order to further use them for the classification task. Our method is based on the statistical significance of adding a feature from a ranked-list to the final subset. The technique extracts the best non-consecutive genes from the ranking, trying to statistically avoid the influence of unnecessary genes in further

classifications. Our approach, named BIRS, uses a very fast search through the attribute space and any classifier can be embedded into it as evaluator. The efficiency and effectiveness of our technique has been demonstrated through extensive comparative study with other representative heuristics. Our approach shows an excellent performance, not only at identifying relevant features, but also with respect to the computational cost. This approach has been applied to select genes in microarray data with excellent results: our technique chooses a small subset of genes from the original set (0.0018% on average) with predictive performance similar to others [10], [18], [24].

2.3 Data Streams

Mining data streams is a challenging task that requires online systems based on incremental learning approaches. In this field we have developed a classification system based on decision rules that may store up–to–date border examples to avoid unnecessary revisions when virtual drifts are present in data. Consistent rules classify new test examples by covering and inconsistent rules classify them by distance as the nearest neighbour algorithm. In addition, the system provides an implicit forgetting heuristic so that positive and negative examples are removed from a rule when they are not near one another. Thus, our system (called FACIL) is an incremental rule learner with partial instance memory based on moderate generalization and example nearness. Our proposal is not based on a window policy but examples are rejected when they do not describe a decision boundary. On the contrary, FACIL builds and refines inconsistent rules simultaneously without adversely affecting the learning efficiency and avoiding unnecessary revisions when virtual drifts are present in data. Experimental results show an excellent performance of our approach as a general purpose classification method [2], [14].

2.4 Visual Data Mining

Visualization has become an essential support throughout the KDD process in order to extract hidden information from huge amount of data. Visual data exploration techniques provide the user with graphic views or metaphors that represent potential patterns and data relationships. However, an only image does not always convey high–dimensional data properties successfully. From such data sets, visualization techniques have to deal with the curse of dimensionality in a critical way, as the number of examples may be very small with respect to the number of attributes. In this field, we have developed a visual exploration technique (called VETIS) that automatically extracts relevant attributes and displays their ranges of interest in order to support two data mining tasks: classification and feature selection. Through different metaphors with dynamic properties, the user can re-explore meaningful intervals belonging to the most relevant attributes, building decision rules and increasing the model accuracy interactively. VETIS means an approach for visualization of multidimensional data sets with numerical attributes based on feature selection via segmentation. In addition, VETIS allows interactive mining of flexible decision rules through successive exploration levels in which the user decides the support and the purity a segment must fulfil to be taken as condition belonging to the antecedent of a rule. Results are very interesting as the tool is versatile and allows building a model with the accuracy level needed [4], [6], [9].

2.5 Application to Electrical Systems

In this area we have worked in two directions: prices and demand forecasting and scheduling of the energy production. In the Time Series forecasting field we have proposed a simple methodology based on the Weighted Nearest Neighbours technique to forecast hourly prices in deregulated electricity markets. Prediction results corresponding to the market of Spain for the entire year 2002 are reported, yielding an average monthly error which is close to 8%. The performance of the proposed method is also compared with that of other techniques such as ANN, Neuro-Fuzzy systems, GARCH and ARIMA (with and without Wavelet transform). In view of the results previously reported in the literature, it can be stated that the proposed method performs satisfactorily, particularly when the uncertainty of prices associated with the Spanish system are taken into account [8], [13].

In the other hand, we have developed an evolutionary technique applied to the optimal short-term scheduling (24 h) of the electric energy production. The equations that define this problem lead to a non-convex non-linear programming problem with a high number of continuous and discrete variables. Consequently, the resolution of the problem based on combinatorial methods is rather hard. The required heuristics, introduced to assure the feasibility of the constraints, are analyzed, along with a brief description of the proposed GA. The GA is used to compute the optimal on/off status of thermal units and the fitness function is obtained by solving a quadratic programming problem by means of a standard non-linear Interior Point method. The results from real-world cases based on the Spanish power system were reported, which showed the good performance of the proposed algorithm, given the complexity and dimensionality of the problem [12].

2.6 Bioinformatics

In this area, we address the biclustering of gene expression data with evolutionary computation. Our approach is based on evolutionary algorithms, which have been proven to have excellent performance on complex problems, and searches for biclusters following a sequential covering strategy. The goal is to find biclusters of maximum size with mean squared residue lower than a given value. In addition, we pay special attention to the fact of looking for high-quality biclusters with large variation, i.e., with a relatively high row variance, and with a low level of overlapping among biclusters. In general, our approach shows an excellent performance at finding patterns in gene expression data [7], [26], [27].

Also we have worked on the identification of two types of patterns in biclustering (shifting and scaling) and demonstrated that the mean squared residue is very useful to search for shifting patterns, but it is not appropriate to find scaling patterns because even when we find a perfect scaling pattern the mean squared residue is not zero. In addition, we have provided an interesting result: the mean squared residue is highly dependent on the variance of the scaling factor, which makes possible that any algorithm based on this measure might not find these patterns in data when the variance of gene values is high. The main contribution of this research is to prove that the mean squared residue is not precise enough from the mathematical point of view in order to discover shifting and scaling patterns at the same time [5], [17].

3 Result Indicators

3.1 Training researchers

From the beginning of the project several researchers have been incorporated: Beatriz Pontes Balanza (2006), Isabel Nepomuceno Chamorro (2005), Juan Antonio Nepomuceno Chamorro (2006), Norberto Díaz Díaz (2005) y Francisco Martínez Álvarez (2006). The first four of them are currently faculty of the Department of Lenguajes y Sistemas Informáticos of the Universidad de Sevilla. All of them have been finished the teaching period of the PhD program.

In the years 2005 and 2006 three PhD. thesis have been defended:

- Title: Técnicas Avanzadas de Predicción y Optimización aplicadas a Sistemas de Potencia
  PhD student: Alicia Troncoso Lora. Date: Junio 2005
  Supervisors: JC Riquelme, JL Martínez (Dpto. Ingeniería Eléctrica de la U. de Sevilla)
  Award "Fundación Sevillana de Electricidad"
- Title: Heurísticas de Selección de Atributos para Datos de Gran Dimensión
  PhD student: Roberto Ruiz Sánchez. Date: Junio 2006
  Supervisors: JC Riquelme, JS Aguilar Ruiz
  European Doctor Mention
- Title: Aprendizaje Incremental y Modelado Visual de Reglas de Decisión
  PhD student: Francisco Ferrer Troyano. Date: Julio 2006
  Supervisors: JC Riquelme, JS Aguilar Ruiz

3.2 Publication record

During the period 2005-2006 the many research results have been published in several international conferences and journals. It is worth to mention the fifteen articles published in journals of the *Journal Citation Report*. Among these journals are BioInformatics (IF: 6.019), IEEE Trans. on Evolutionary Computation (IF:3.257), Pattern Recognition (IF:2.176), IEEE Trans. on SMC-B (IF:1.052), IEEE TKDE (IF:1.758), IJSEKE (IF:0.350), IEEE Trans. on SMC-C (IF:0.482), JUCS (IF:0.456), EJOR (IF:0.828), AI Comm (0.612), etc. Thus, the group has participated in 18 international conferences whose proceedings were edited by outstanding editorials, such as Springer-Verlag, IEEE Press or ACM Press.

3.2.1 International Journals (included in JCR)

[1] Giráldez R, JS Aguilar-Ruiz, Riquelme JC, "Knowledge-based Fast Evaluation for Evolutionary Learning", IEEE Systems, Man and Cybernetics Part C, Vol. 35, nº 2, pp.254-261, IEEE Press, 2005.
[2] Ferrer-Troyano F, Aguilar-Ruiz JS, Riquelme JC, "Incremental Rule Learning and Border Examples Selection from Numerical Data Streams", J. of Universal Computer Science, Vol. 11, nº 8, pp. 1426-1439, Springer, 2005.
[3] Giráldez, R. "Improving the performance of Evolutionary Algorithms for Decision Rules Learning", AI Communications, Vol. 18 (1), pp. 63-65, 2005.
[4] Ferrer-Troyano F, Aguilar-Ruiz JS, Riquelme JC, "Connecting Segments for Visual Data

Exploration and Interactive Mining of Decision Rules", J. of Universal Computer Science, Vol. 11, nº 11, pp. 1835-1848, Springer, 2005.
[5] Aguilar-Ruiz J, "Shifting and Scaling Patterns from Gene Expression Data", BioInformatics, Vol. 20 pp. 3840 - 3845, 2005.
[6] Álvarez JL, Mata J, Riquelme JC, "Data Mining for the management of software development project", IJ of Software Eng. and Know. Eng. Vol. 14, nº 6, pp. 665-695, World Scientific Pub, 2005.
[7] Aguilar-Ruiz JS, F Divina, "Biclustering of expression data with evolutionary Computation", IEEE Transactions on Know. & Data Eng., Vol. 18 (5), pp. 590-602, 2005.
[8] Troncoso, A, "Advances in optimization and prediction techniques: real world applications", AI Communications, Vol. 19 (3), pp. 295-297, 2006.
[9] Riquelme JC, Polo M, Aguilar-Ruiz JS, Piattini M, Ferrer F, Ruiz F, "A comparison of effort estimation methods for 4GL programs: experiences with Statistics and Data Mining", IJ of Software Eng. and Know. Eng., Vol. 16, nº 1, pp. 127-140, World Scientific Pub, 2006.
[10] Ruiz R, Riquelme JC, Aguilar-Ruiz JS, "Incremental wrapper-based gene selection from microarray expresion data for cancer classification", Pattern Recognition, Vol. 39 (12), pp. 2383-2392, Elsevier Science, 2006.
[11] Aguilar-Ruiz JS, Giráldez R, Riquelme JC, "Natural Encoding for Evolutionary Supervised Learning", IEEE Transactions on Evolutionary Computation, pendiente de publicación, 2007.
[12] Troncoso A, Riquelme JC, Aguilar-Ruiz JS, Riquelme JM, "Evolutionary Techniques applied to the optimal short-term scheduling of the electrical energy production", European J. of Operational Research, pendiente de publicación Elsevier, 2007.
[13] Troncoso A, Riquelme JM et al., "Electricity market price forecasting based on weighted nearest neighbors techniques", IEEE Transactions on Power Systems, pendiente de publicación, 2007.
[14] Ferrer-Troyano FJ, Aguilar-Ruiz JS, Riquelme JC, "Data streams hybrid classification via border examples selection under rule induction", Applied Soft Computing, aceptado, pendiente de publicación, 2007.
[15] Aroba J, JA Grande, JM Andujar, ML de la Torre, JC Riquelme, "Application of fuzzy logic and data mining techniques as tools for qualitative interpretation of acid mine drainage processes", Environmental Geology, aceptado, pendiente de publicación, 11 páginas, 2007.

3.2.2 International Conferences

[16] Diaz-Diaz N, D Rodriguez, I Nepomuceno, J.S. Aguilar-Ruiz "Neighborhood-based clustering of gene-gene interactions", LNAI, Vol. 4224, pp. 1111-1120, IDEAL 2006.
[17] Giráldez R, B. Pontes, J.S. Aguilar-Ruiz, "Shifting patterns discovery in microarrays with evolutionary algorithms", LNAI, Vol. 4252, pp. 1264-1271, KES 2006.
[18] Ruiz R., B. Pontes, R. Giráldez, J.S. Aguilar-Ruiz, "Gene ranking from microarray data for cancer classification - a machine learning approach", LNAI Vol. 4252, pp. 1272-1280, KES, 2006.
[19] Aguilar-Ruiz J.S., JA Nepomuceno, N. Diaz-Diaz, I Nepomuceno, "A measure for data set editing by ordered projections", LNCS, Vol. 4031, pp. 1349-1358, IEA-AIE, 2006.
[20] Ferrer-Troyano F., JC Riquelme, J.S. Aguilar-Ruiz, "Data streams classification by incremental rule learning with parameterized generalization", SAC, ACM Press, pp. 657-661, 2006.
[21] Aguilar-Ruiz JS, D. Rodriguez, DA Simovici, "Biclustering of gene expression data based on local nearness", Extraction et Gestion des Connaissances EGC, pp. 681-692, 2006.

[22] Aguilar-Ruiz J.S., F Divina, "Evolutionary biclustering of microarray data", LNCS Vol. 3449, pp.1-10, EvoBio 2005.
[23] Diaz-Diaz N., J.S. Aguilar-Ruiz, JA Nepomuceno, García J, "Feature selection based on bootstrapping", CIMA, Advanced Computing in Biomedicine, IEEE Press, pp. 1-6, 2005.
[24] Ruiz R., J.S. Aguilar-Ruiz, JC Riquelme, N. Diaz-Diaz, "Analysis of feature rankings for classification", LNCS Vol. 3646, pp. 362-372, IDA, 2005.
[25] Giráldez R., J.S. Aguilar-Ruiz, "Feature influence for evolutionary learning", GECCO, pp. 1139-1145, ACM Press, 2005.
[26] Aguilar-Ruiz J.S., F. Divina, "GA-based approach to discover meaningful biclusters", GECCO, pp. 473-474 ACM Press, 2005.
[27] Aguilar-Ruiz J.S., F. Divina, "Evolutionary Computation for biclustering of gene expression", SAC, ACM Press, pp. 959-960, 2005.
[28] Ruiz R., JC Riquelme, J.S. Aguilar-Ruiz, "Heuristic search over a ranking for feature selection", LNCS Vol. 3512, pp. 742-749, IWANN 2005.
[29] Giráldez R., N. Diaz-Diaz, I. Nepomuceno, J.S. Aguilar-Ruiz, "An approach to reduce the cost of evaluation in evolutionary learning", LNCS, Vol. 3512, pp. 804-811, IWANN 2005.
[30] Ferrer-Troyano F., J.S. Aguilar-Ruiz, JC Riquelme, "Incremental rule learning based on example nearness from numerical data streams", SAC, ACM Press pp. 568-572, 2005.
[31] Pachón V, Mata J, Roche F, Riquelme JC, "Application of a KDD methodology to a real industrial process", WAC World Automation Congress, pp. 495-500, 2005.
[32] Blanquero R., E. Carrizosa, N. Díaz-Díaz et al. "Classification with VNS", Mini EURO Conference on VNS, 2005.
[33] Nepomuceno I, JA Nepomuceno, R Ruiz, JS Aguilar-Ruiz "Databases Reduction Simultaneously by Ordered Projections", DS, pp. 337-341, 2006.
[34] Pontes B, F. Divina, R. Giráldez, JS Aguilar-Ruiz, "Virtual error: a new measure for evolutionary biclustering", accepted, EvoBio 2007.
[35] Martínez F, A Troncoso, JC Riquelme, JM Riquelme, "Discovering patterns in electricity price using clustering techniques", accepted, ICREPQ 2007.

3.2.3 National Conferences

[36] Troncoso A, JC Riquelme, "Predicción de Series Temporales basada en Vecindad", Simposio de Inteligencia Computacional, SICO'2005, pp. 105-112.
[37] Pontes B, DS Rodriguez-Baena, N Díaz-Díaz, R Giráldez, "Análisis de Datos de Expresión Genética", Jornadas de Automática, 2006.
[38] Blanquero R, E Carrizosa, N Díaz-Díaz, "Un algoritmo heurístico para Clasificación Semisupervisada", Congreso Nacional de Estadística e Investigación Operativa, 2006.
[39] Divina F, Aguilar-Ruiz JS, N Díaz-Díaz, "Biclustering de datos de expresión genómica con computación evolutiva multiobjetivo", MAEB, pp. 229, 2005.
[40] Aguilar-Ruiz JS, N Diaz-Diaz, "Selección de atributos relevantes basada en bootstrapping", TAMIDA, 2005.
[41] Nepomuceno I, JA Nepomuceno, R Ruiz, "RESOP: Un método para la reducción de Bases de Datos", JISBD, pp. 262-270, 2006.

3.3 Participation in other projects

At European level, we are participating in the KDNet network and its renewal KDUbiq. Also we

are working in the Spanish-Portuguese Integrated Action "Knowledge Discovery in Data Streams", together with the University of Oporto, which is aimed at the study of new methodologies for data streams. Another international project is the "Advanced Data - and Knowledge - Driven Methods for State Failure Risk Assessment" project supported by NATO, within the "Security through Science" framework. The Rudjer Boskovic Instituce (Croatia) and the University of Ulster (UK) also participate in this project, which aims at studying the existing relations between economic and social-demographic variables of a country and its political stability.

At national level the group collaborates with the groups belonging to the Spanish Data Mining and Learning Network, coordinated by J. Riquelme. Currently this network is applying for its third renewal. This application supported by more than 30 national groups and by almost 300 researchers, which practically constitutes the totality of Spanish researchers working in this research field. The collaborations with the universities of Granada (F. Herrera and A. Vila), Malaga (R. Morales) and Cordoba (C. Hervás) are especially consolidated. We also collaborate with the MINDAT-Plus project for the application of the research lines of each group to real-world problems.

Besides of this project, at a regional level, the members of this project are currently in the TIC-200 group (coordinated by J. Aguilar) and TIC-134 (coordinated by J. Riquelme) of the Andalusian Research Program. In the bioinformatics field there have been collaborations with José Olivé, Hilario Ramírez y Armando Blanco of the University of Granada, for the proposal of the Master in Bioinformatics in the Andalusia context.

3.4 International relations

Several members of the group have done research visits to foreign universities:

- J. S. Aguilar has done several research visits (May-Aug 2004 and Jun-Sep 20005) at the University of Massachusetts in Amherst and Boston, collaborating with Prof. Paul Cohen and Dan Simovici, working on data streams analysis and bioinformatics, respectively.
- A. Troncoso has spent two months at the U. of Colorado (Jul-Aug 06), collaborating with Prof. Manuel Laguna and working in optimization techniques as Scatter Search Algorithms.
- R. Ruiz was six weeks in the University of Reading (November and December 2005) with Prof. Vassil Alexandrov, working on parallelization of data mining algorithms.

At international level the group has participated in the organization of various tracks and workshops in several conferences. In collaboration with J. Gama of the University of Oporto, the group organized the data streams tracks of SAC since 2004. With P.R. Cohen (University of Southern California) and with S. Muthukrishnan (Rutgers University), the group participated in three ECML: 2004 to 2006. The group participated also in the ICML 2006 held at the Carnegie Mellon University of Pittsburg, with J. Gama and J. Roure. Next March, the group is participating, together with Y. Lechvallier, at the First European Workshop on Data Streams Analysis in Naples. All these collaborations have already succeeded co-authoring papers with P.R. Cohen and D. Simovici (University of Massachussets), F. Azuaje (University of Ulster), E. Marchiori (Free University of Amsterdam), J. Bacardit (University of Nottingham), etc. Moreover the collaborations resulted also in the edition of special issues of journals, like Intelligent Data Analysis, New Generation Computing, JUCS, etc.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Advanced Methods and Tools for the Production of Next-Generation Industrial Software TIN2004-00779

Jaime Gómez
Universidad de Alicante
c/ San Vicente s/n 03690 Alicante, Spain

Abstract

Effectiveness and the productivity in the current development of software needs innovative methods, techniques and tools to allow the applicability of theoretical solutions to industrial context. The results obtained in the last years evidence the wide distance between the state of the art and the real practice in the development of next generation software systems. With this project we intend to face these challenges by producing methodologies and tools to make possible the integration of the critical activities in the development of industrial software. Within this project, we are defining models, languages and tools for the design and development of application software in open systems, characterizing their applicability, defining the appropriate methodologies for each case, and building tools to support and allow the easy application of such methodologies. As a concretion of the contributions of this project we are building tools of great practical interest in two different application domains: Web Engineering and Data Warehouses.
Keywords: Conceptual Modelling, Web Engineering, Model-Driven Development, Data Warehouses, Document Modelling.

1 Project Aims

The METASIGN Project is a non-coordinated project leaded by Dr. Jaime Gómez, from University of Alicante. According to its initial plan the objetives of the project were categorized in the following three main tracks:

- Executable Web Conceptual Models
- Models and Tools for the Design of Datawarehouses
- Digital Document Modelling

The general objetive is to produce methodologies and tools to make possible the integration of critical activities in the development of industrial software. With this purpose, we aim at a precise objetive: to facilitate the construction of next generation software developed under industrialization

levels similar to those that have been reached in the production environment of other engineering disciplines.

Following the commented structure, we are going to introduce the particular, concrete objetives associated to the three sections of the project. In all cases, the required activities are scheduled to be accomplished in the three years of project duration. As the initial schedule has been basically respected, more details can be obtained in the original project description [52].

Executable Web Conceptual Models

The objetive of this project part is to define a software production environment to generate web applications from conceptual models. In concrete, the project focus on the following aspects:

- To improve web conceptual modelling methods to support the specification of new relevant properties (i.e. personalization).
- To define an action language (with a precise semantics) to improve the automatization of the software production process.
- To identify new execution models for web applications.
- To understand how to design and develop web applications for the most relevant software architectures (possibly applying the MDA standard).

Models and Tools to Manage Information from Datawarehouses

In this context, the Project objetive is to improve the conceptual design phase in the development of Datawarehouses. This type of systems are specially relevant for decission support in big–size companies. The concrete objetives can be structured according to the following items:

- To improve the conceptual design in Datawarehouses (DW) in order to get more expresiveness.
- To define a guideline for DW based on the Rational Unified Process.
- To improve the techniques used to automate ETL processes starting from hetereogeneus data sources.
- To develop some relevant case studies with an industrial focus in mind as a previus step to apply the techniques in the EPOs of this project.

Digital Document Modelling

The Project objetive in this context is to develop a methodological proposal to design the structure of complex documents like those used in Digital Libraries, Help-Desks and so on. To achieve this central objetive, the Project focus on the following specific aspects:

- To study relation between semistructure data models and the UML.
- To integrate UML and XML standards in an interactive and dynamic environment that allows the definition of different typs of semistructured documents.
- To define an automatization process to elaborate those digital documents.
- To study the integration of those documents with DW.

2 Project accomplishments

The sucess level achieved in the Project is high. It is being really very productive, and it is mainly being developed according to the initial schedule. There are not remarkable problems, beyond the fact of being out of money for travelling too soon, due to the big success in the acceptance rate of papers in relevant national and international conferences. The level of cooperation with industry is even higher than expected, whan is again a strong reason for satisfaction.

Next, we highline the main achieved scientific and technological results, following the same Project structure used in the previous section. In the next section, we will introduce the concrete production indicators to justify and measure the success we are talking about.

Executable Web Conceptual Models

We have been working on extensions to Object-Oriented, Model-Based Code Generation Techniques, starting from the OO-H approach developed in Alicante. As a result of that some extensions has been produced:

- Inclusion of personalization support in the method by means of a new Personalization Rule Modelling Language (PRML).
- An MDA approach to Generate Web Applications following the MDA approach (WebSA).

Both extensions have been extensively and successfully published, they have generated both a rich industrial relationships and a rich set of academic results. The most relevant publications are cited in [ref].

Models and Tools to Manage Information from Datawarehouses

Two main results in this area:

A process, a method and a prototype for DW modelling have been defined. Generically called DW4UML, its extends the GOLD model by guiding the process of generating a DW model. This model is converted into a DW conceptual schema following a precise set of steps to translate modelling primitives into their corresponding Conceptual Modeling counterpart. The process is partially automated. It has generated a rich set of academic results and a very promising industrial relation.

A process and a method to model ETL processes in DW. In this case, we provide the necessary mechanisms for an easy and quick specification of the common operations defined in ETL processes such as, the integration of different data sources, the transformation between source and target attributes and the generation of surrogate keys. As an advantage of this proposal we provide a seamless integration of the design of the ETL processes with the DW conceptual schema. This process is partially automated, and it has generated a rich set of academic results.

Digital Document Modelling

This Project section has generateg a proposal to adapt modeling methods originally meant for business applications, on the design of the complex markup vocabularies used for XML Web-content production. We have been working on integrating these technologies to create a dynamic and interactive environment for the design of document markup schemes. The task focuses on the analysis, design and maintenance of XML vocabularies based on UML. It considers the automatic generation of Schemas and DTDs from a visual UML model of the markup vocabulary, as well as pieces of software, like input forms. Additionally, we are integrating these UML design capabilities with other handy tools like automatic Schema simplification and multilingual markup.

The proposal has been applied successfully to develop real applications in the context of the Miguel de Cervantes Digital Library. The next step will be to try to generate a tool that implements this environment for document modelling. This part is generating an interesting set of academic results, and an incipient industrial relationships with the industrial partners of the Project.

3 Project-acomplishment metrics

The Project results indicators are going to be structures following the same items that are used along this report. Its presentation is strongly connected to the previous section information. The most relevant resutls will be introduced (researchers in formation, publications in journals and LNCS based conferences, technological transfer, international innitiatives and collaborations with other research groups). The full set of publications incluidng other conferences publications can be seen in the final results summary report [52].

PhD Thesis

Sergio Luján. "DataWarehouse Design with UML". Advisor: Juan Trujillo. Julio 2005.
Santiago Melia. "WebSA:an MDA-based method for Web Applications". Advisor: Jaime Gómez. Febrero 2007.

Two more PhD thesis are about to be defended:
Irene Garrigós. "PRML: a Personalization Rule Modelling Language". Advisor: Jaime Gómez.
Jose-Norberto Mazón. "Transformations for Datawarehouses with MDA". Advisor: Juan Trujillo.

Publications (2005-2007)

The following is a summary of the 51 publications (international workshops and national conferences not included) obtained during the years of the project.

Books (edited)	7
Book chapters	4 (published by Springer, Addison.Wesley, etc.)
Articles in International Journals	17 (**11 of them included in the JCR**)
Paper in International Conferences	23 (published in LNCS, IEEE CS Press, etc.)

Technology transfer and Collaboration with other research groups

This project has greatly helped establishing strong links with other research groups . We want to remark specially the interaction with the groups of Nora Koch (LMU, Munich), Piero Fraternali (Politecnico de Milano, Italy), Olga De Troyer/Geer.Jan Houben (VUB, Brussels) and Gerti Kappel (TuWien, Austria). The participation in the Program Committee of some of the most relevant international conferences on Web Engineering and DW is another consequence of all the work done (including WWW, ICWE, EDOC, ER, DOLAP, DAWAK among others).

The industrial liason has been very important with Diputación de Alicante, Caja de Ahorros del Mediterráneo, COITI and COGAA. These R&D contract prove the interest of the involved companies, and are for us a clear indicator of success.

4 References

[1] Meliá, S; Gómez, J.;Koch, N. Improving Web Design Methods with Architecture Modeling. EC-Web 2005. LNCS 3590.Pags 53-64. 2005.
[2] Bia, A.; Gómez, J. UML for Document Modeling: Designing Document Structures for Massive and Systematic Production of XML-based Web Contents. MoDELS 2005. LNCS 3713. Pags 648-660. 2005.
[3] Bia, A.; Gómez, J. DiCoMo: An Algorithm Based Method to Estimate Digitization Costs in Digital Libraries. ECDL 2005. LNCS 3652. Pags 519-520. 2005.
[4] Garrigós, I. ,Casteleyn, S., Gómez, J. A Structured Approach to Personalize Websites using the OO-H Personalization Framework. APWeb 2005. LNCS 3399. Pags 695-706. 2005.
[5] Meliá, S; Kraus, A.;Koch, N. MDA Transformations Applied to Web Application Development. ICWE 2005. LNCS 3579. Pags 465-471. 2005.
[6] Meliá, S.; Gómez J. Applying Transformations to Model Driven Development of Web Applications. ER Workshops 2005. LNCS 3770. Pags 63-73. 2005
[7] Casteleyn S., Garrigós, I., De Troyer, O.. Automatic Runtime Validation and Correction of the Navigational Design of Web Sites. APWeb 2005. LNCS 3399. Pags 453-463. 2005.
[8] Gómez, J.; Bia, A.;Párraga, A. Tool Support for Model-Driven Development of Web Applications. WISE 2005. LNCS 3806. Pags 721-730. 2005.
[9] Gómez, J. Designing Data-Intensive Web Applications. The Computer Journal 49(1):127. 2006
[10] Comai, S.: Matera, M.; Cachero, C. Editorial message: special issue on web technologies and applications. Journal of Web Engineering 5(1):1-2. 2006

[11] Cachero C.; Genero M.; Calero C.; Meliá S. Quality-driven Automatic Transformation of Object-Oriented Navigational Models. ER Workshops 2006. LNCS 4231.Pags 329-338. 2006.

[12] Meliá, S.; Gómez J. The WebSA Approach: Applying Model-Driven Engineering to Web Applications. Journal of Web Engineering 5(2): 121-149. 2006.
[13] Kozareva Z., Ferrández O., Montoyo A., Muñoz R., Suárez A. and Gómez J. Combining data-driven systems for improving Named Entity Recognition. Data & Knoledge Engineering (in Press).
[14] Vicedo J.L., Gómez J. TREC: Experiment and Evaluation in Information Retrieval. Journal of the American Society for Information Science and Technology 2006. (In Press).
[15] Gómez, J.; Bia, A.;Párraga, A. Tool Support for Model-Driven Development of Web Applications. International Journal of Information Technology and Web Engineering (In Press).
[16] A Min Tjoa, Juan Trujillo. Data Warehousing and Knowledge Discovery. Lectura Notes in Computer Science. Vol. 3589. Páginas: 538. ISSN: 0302-9743.
[17] Jacky Akoka, Stephen W. Liddle, Il-Yeol Song, Michela Bertolotto, Isabelle Comyn-Wattiau, Samira Si-Said Cherfi, Willem-Jan van den Heuvel, Bernhard Thalheim, Manuel Kolp, Paolo Bresciani, Juan Trujillo, Christian Kop, Heinrich C. Mayr: Perspectives in Conceptual Modeling, ER 2005 Workshops AOIS, BP-UML, CoMoGIS, eCOMO, and QoIS. Klagenfurt, Austria, 24-28 de Octubre de 2005. Lecture Notes in Computer Science. Vol. 3770. Páginas 476. ISSN: 0302-9743..
[18] Juan Trujillo y Il-Yeol Song. ACM 8th International Workshop on Data Warehousing and OLAP, DOLAP 2005. Celebrado junto con ACM 14th Conference on Information and [19] Knowledge Management (CIKM 2005). Bremen, Alemania, 4-5 de Noviembre de 2005. Páginas 114. ISBN: 1-59593-162-7.
[20] Serrano, M., Calero, C., Trujillo, J., Luján-Mora, S., Piattini, M. Metrics for Datawarehouses Conceptual Models. In: Metrics for Software Conceptual Models, London: Imperial College Press, pp. 207-235. ISBN: 1-86094-497-3.
[21] Il-Yeol Song, Kurt Yano, Juan Trujillo, Sergio Luján-Mora. A Taxonomic Class Modeling Methodology for Object-Oriented Analysis. In J. Krogstie, T. Halpin, K. Siau (Eds.), Information Modeling Methods and Methodologies, Hershey, PA: Idea Group Publishing, p. 216-240. ISBN: 1-59140-376-6.
[22] Juan Trujillo, Sergio Luján-Mora, Il-Yeol Song. Design and Representation of Multidimensional Models with UML and XML Technologies. In K. Siau (Ed.), Advanced Topics in Database Research, vol. 4, Hershey, PA: Idea Group Publishing, p. 50-95. ISBN: 1-59140-471-1.
[23] Trujillo, J., Lujan-Mora, y Song, I. Applying UML for designing multidimensional databases and OLAP applications. Advanced Topics in Database Research., pp. 13-36. ISBN: 1-59140-063-5.
[24] Rodolfo Villarroel, Eduardo Fernández-Medina, Juan Trujillo, y Mario Piattini. Un Profile de UML para Diseñar Almacenes de Datos Seguros (A UML Profile for Designing Secure Data Warehouses) IEEE América Latina. Vol 3, Issue 1. http://www.ewh.ieee.org/reg/9/etrans/Marzo2005/March05.htm. ISSN: 1548-0992. Marzo de 2005.
[25] Zubcoff, J.J., Trujillo, J. Extending the UML for designing Association Rule Mining Models for Data Warehouses. 7th International Conference on Data Warehousing and Knowledge Discovery (DaWak'05), Agosto 22-26, 2005. Lecture Notes in Computer Science. Vol. 3589, pp. 11-21. Springer-Verlag.
[26] Berenguer, G., Romero R., Trujillo, J., Serrano, M., Piattini, M. A set of quality indicators and their corresponding quality metrics for conceptual models for Data Warehouses. 7th International

Conference on Data Warehousing and Knowledge Discovery (DaWak'05), Agosto 22-26, 2005. Lecture Notes in Computer Science. Vol. 3589, pp. 95-104. Springer-Verlag.
[27] Tho Manh Nguyen, A. Min Tjoa, Juan Trujillo: Data Warehousing and Knowledge Discovery: A Chronological View of Research Challenges. 7th International Conference on Data Warehousing and Knowledge Discovery (DaWak'05), Agosto 22-26, 2005. Lecture Notes in Computer Science. Vol. 3589, pp. 530-535. Springer-Verlag.
[28] A Min Tjoa, Juan Trujillo. Data Warehousing and Knowledge Discovery. Lectura Notes in Computer Science. Vol. 4081. Páginas: 575. ISSN: 0302-9743.
[29] John F. Roddick, V. Richard Benjamins, Samira Si-Said Cherfi, Roger H. L. Chiang, Christophe Claramunt, Ramez Elmasri, Fabio Grandi, Hyoil Han, Martin Hepp, Miltiadis D. Lytras, Vojislav B. Misic, Geert Poels, Il-Yeol Song, Juan Trujillo, Christelle Vangenot: Advances in Conceptual Modeling - Theory and Practice, ER 2006 Workshops BP-UML, CoMoGIS, COSS, ECDM, OIS, QoIS, SemWAT, Tucson, AZ, USA, November 6-9, 2006. Lecture Notes in Computer Science. Vol. 4231. Páginas 454. ISSN: 0302-9743.
[30] Eduardo Fernández-Medina, Juan Trujillo, Rodolfo Villarroel and Mario Piattini, Access control and audit model for the multidimensional modeling of data warehouses, Decision Support Systems, 42(3): 1270-1289 (2006)
[31] Sergio Luján-Mora, Juan Trujillo and Il-Yeol Song, A UML profile for multidimensional modeling in data warehouses, Data & Knowledge Engineering, 59(3): 725-769 (2006). ISSN: 0169-023X.
[32] Rodolfo Villarroel, Eduardo Fernández-Medina, Mario Piattini, Juan Trujillo: A UML 2.0/OCL Extension for Designing Secure Data Warehouses. Journal of Research and Practice in Information Technology 38(1): (2006).
[33] Sergio Luján-Mora, Juan Trujillo. Applying the UML and the Unified Process to the design of Data Warehouses. Journal of Computer Information Systems (JCIS), XLVI(5), p. 30-58, Special issue 2006. ISSN: 0887-4417.
[34] Sergio Luján-Mora, Juan Trujillo. Physical Modeling of Data Warehouse Using UML Component and Deployment Diagrams: Design and Implementation Issues. Journal of Database Management (JDM), 17(2), p. 12-42, Abril-Junio 2006. ISSN: 1063-8016.
[35] Mazón J-N., Trujillo J., Serrano M., Piattini M. Applying MDA to the development of data warehouses. Computing Reviews. December 2006. ISSN 1530-6586.
[36] Mazón J-N.; Trujillo J.; Lechtenbörger, J. A Set of QVT Relations to Assure the Correctness of Data Warehouses by Using Multidimensional Normal Forms. 25th International Conference on Conceptual Modeling (ER2006). Tucson, Arizona, USA. Noviembre 2006. Lecture Notes in Computer Science Vol. 4215, pp. 385-398. ER 2006.
[37] Mazón, J-N.; Trujillo, J. Enriching Data Warehouse Dimension Hierarchies by Using Semantic Relations. 23rd British National Conference on Databases (BNCOD 2006). 18 - 20 July, 2006. Belfast, Northern Ireland. Lecture Notes in Computer Science, Vol. 4042, pp. 278-281.
[38] Mazón, J-N.; Pardillo, J.; Trujillo,J. Applying Transformations to Model Driven Data Warehouses. 8th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2006). Cracovia (Polonia). 4-8 Septiembre, 2006. Lecture Notes in Computer Science, Vol. 4081, pp. 13-22.
[39] Pardillo, J.; Mazón, J-N.; Trujillo,J. Implementación de Profiles UML 2.0 en herramientas CASE. I Conferencia Internacional sobre Ciencias y Tecnologías Multidisciplinares de la Información. Octubre 2006, Mérida, pp. 560-564. ISBN: 84-611-3105-3.
[40] Jose Zubcoff, C. Ojeda-Martínez, J. Bayle-Sempere, A. Forcada, C. Valle, J.L. Sánchez-Lizaso, P. Sánchez-Jerez, Juan Trujillo. A data mining approach to assess the effects of some management

tools on coastal Mediterranean fisheries. I Conferencia Internacional sobre Ciencias y Tecnologías Multidisciplinares de la Información. Octubre 2006, Mérida, pp. 8-11. ISBN: 84-611-3105-3.

[41] Jose Zubcoff, Juan Trujillo. Un modelo conceptual para clustering en datos estructurados. I Conferencia Internacional sobre Ciencias y Tecnologías Multidisciplinares de la Información. Octubre 2006, Mérida, pp. 67-71. ISBN: 84-611-3105-3.
[42] Rodolfo Villarroel, Emilio Soler, Eduardo Fernández-Medina, Mario Piattini, Juan Trujillo: Representing levels of abstraction to facilitate the Secure Multidimensional Modeling. First International Conference on Availability, Reliability and Security, ARES 2006, The International Dependability Conference - Bridging Theory and Practice, April 20-22 2006, Vienna University of Technology, Austria. IEEE Computer Society 2006. pp. 678-684.
[43] Emilio Soler, Rodolfo Villarroel, Juan Trujillo, Eduardo Fernández-Medina, Mario Piattini: Representing Security and Audit Rules for Data Warehouses at the Logical Level by Using the Common Warehouse Metamodel. First International Conference on Availability, Reliability and Security, ARES 2006, The International Dependability Conference - Bridging Theory and Practice, April 20-22 2006, Vienna University of Technology, Austria. IEEE Computer Society 2006. pp. 914-921.
[44] Rodolfo Villarroel, Emilio Soler, Eduardo Fernández-Medina, Juan Trujillo, Mario Piattini: Using UML Packages for Designing Secure Data Warehouses. Computational Science and Its Applications - ICCSA 2006, International Conference, Glasgow, UK, May 8-11, 2006, Proceedings, Part III. Lecture Notes in Computer Science 3982, pp. 1024-1034.
[45] Zubcoff, J; Trujillo,J. Conceptual Modeling for Classification Mining in Data Warehouses. 8th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2006). Cracovia (Polonia). 4-8 Septiembre, 2006. Lecture Notes in Computer Science, Vol. 4081.
[46] Juan Trujillo and Il-Yeol Song . Special issue of Decision Support Systems on the Eighth International Workshop on Data Warehousing and OLAP (DOLAP'05). Decision Support Systems, In Press.
[47] Juan Trujillo and A. Min Tjoa. Data warehouse and knowledge discovery (DAWAK'05). Data & Knowledge Engineering, In Press.
[48] Eduardo Fernández-Medina, Juan Trujillo, Rodolfo Villarroel and Mario Piattini, Developing Secure Data Warehouses with a UML extension, Information Systems, In Press.
[49] Manuel Serrano, Juan Trujillo, Coral Calero and Mario Piattini. Metrics for data warehouse conceptual models understandability. Information and Software Technology, In Press.
[50] Jose Zubcoff and Juan Trujillo. A UML 2.0 profile to design Association Rule mining models in the multidimensional conceptual modeling of data warehouses. Data & Knowledge Engineering, In Press.
[51] Jose-Norberto Mazón, Juan Trujillo. An MDA Approach for the Development of Data Warehouses. Decision Support Systems. In Press.
[52] Proyecto METASIGN: Metodologías y herramientas avanzadas para la producción de software industrial de nueva generación. Memoria técnica del proyecto. Grupo IWAD. DLSI/UA. 2007.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

An Intelligent System for the Early Detection of Breast Cancer TIN2004-00986

Francisco Javier López Aligué *
Professor of Electronics. University of Extremadura.
Avda. de Elvas, s/n. 06007 - Badajoz - Spain

Resumen

With the help of a previous Grant TIC2001-0881, our team has developed software for the early diagnosis of breast cancer for the presence of microcalcifications. Our objective is to include the masses in the system, and to generate an elaborate interface for its use as a valuable medical system for the detection of all breast cancers. The final system will be made up of a pre-processing unit responsible for the ROI (The Regions Of Interest) generation followed by a topological pre-processor to eliminate the structural deformations of the images to be classified. Since the present Neural Network really behaves as a simple character classifier, the configurations of its inputs have to be clean of deformation and noise in order to guarantee a 0 % error rate in the final decision. The whole system will be accessible through a "friendly" interface adapted for use with habitual personal computers so that all medical specialists can make use of it. Nevertheless, the system can be re-dimensioned according to the particular needs of every application, so its design and development is being done on a personal computer (PC) with some enhanced characteristics, but allowable for every kind of medical installation. Finally, as are all the present systems, the unit will be connected to the scientific network with free use for the community of Extremadura, easily expandable to the rest of potential users.
Keywords: Breast Cancer, Microcalcifications, Medical CAD.

1. Objectives of the Project

The capacity of today's software development tools, together with the precision of modern medical instruments, has opened up new possibilities for the early diagnosis of numerous diseases. Image processing is one of the fields that has attracted most attention in recent years, and there are currently already many tools available that allow one to work with great effectiveness (in many of the usual senses of this term) on biomedical images. And present-day medicine is devoting a major effort, both economic and human, to the early diagnosis of the different ailments that appear with the greatest incidence in the population. In particular, breast cancer is one of the most frequent reasons for death in our society, and its incidence could

*Email: aligue@unex.es

be reduced enormously if there were some method to detect it when medication or surgery still have the possibility of saving the patient's life. All breast cancer detection involves preventive analysis. There are attempts to avoid the negative aspects produced by a radiological exploration, replacing it by thermal or ultrasound scans - amongst others - but still today radiographical inspection is the safest, most reliable, and above all most sensitive. The new technologies of digital exploration, together with digitizers for conventional plates, provide us with very detailed information of whatever is taking place in the breast. Consequently, the last few years have been an ideal time for the application of the techniques of Artificial Intelligence and Image Processing to aid radiologists in attempting to palliate this true pandemic. The basic starting idea is that on the mammograms the typical lesions of malignancies (nodules and microcalcifications) can appear as too excessively small and hidden among the other details for the human eye. Nevertheless, electronic equipment and processing algorithms can detect them in their most minimal manifestations. This can enhance the effectiveness of preventive exploration in detecting possible malignancies relative to its present capacity.

In particular, the objectives of this Project can be summarized in the five following points:

1. To develop the algorithms for pre-processing the mammogram images, and those for extracting the ROIs (Regions-of-Interest) for the case of the nodules or condensations, from what was achieved with the microcalcifications.

2. To incorporate the classes of malignancies corresponding to the aforementioned ROIs into the operation of the adaptive neuronal network, resizing it for them to be treated in a single procedural context.

3. To check the operation of the system both on the cluster and on a simpler (but equally effective) version that can be run on personal computers.

4. To generate an interface that is "friendly" and useful for the radiologist, and that will in the end be found to be a true medical tool in aiding diagnosis.

5. To manage to attract companies of the sector (AGFA, Philips, etc.) to validate our results and, if the outcome is positive, to incorporate the method into early detection protocols.

Given the objectives indicated, we can identify the following Tasks:

Task T.1. Mammogram Processing Algorithms. Development for the detection of condensations and generation of their ROIs

Objectives: To develop the mammogram processing functions for the detection of possible condensations and their extraction from the overall image for their subsequent processing. To this end, a variant of the present pre-processing is used, the reason being brevity in the time needed and security in execution. Duration: It starts with the Project and covers the first complete year. Personnel: Being a key point in the Project, the greater part of the personnel participate in this task, i.e., LA, AS, and PC on the technical side, and GU and PP on the medical side, these latter shared with the following Task. Person responsible: LA.

Task T.2. Adaptation of the Associative Neural Network to new categories of malignancies

Objectives: Here we will proceed to to modify the entire programming of our Adaptive Network for it to include the new prototypes that identify the different forms taken by this type of malignancy. Duration: This will be developed throughout the Project, since it will have to be operative at any given time to accept whatever new identification. Nevertheless, it is anticipated that the greatest work load will occur simultaneously with the previous task. Personnel: In this case, the rest of the technical personnel, i.e., GV and GO, with the sporadic intervention of the medical members, GU and PP. Person responsible: LA, GV

Task T.3. Integration of the two units and their application to the Database (DDSM) and to our own mammograms

Objectives: Once the objectives corresponding to the previous tasks have been attained, the system will be run and tested together. This is still an experimental product, and therefore one can only make an approximation to its development. But it already has proven functional bases and should give positive results if tasks T1 and T2 have ended correctly. Duration: It will occupy the entire second year and, as a final result, there will be a methodological framework that is open for the inclusion of new lesions or symptoms, above all with a view to the future. Personnel: Now the personnel of Task T.1 will pass to this one. Person responsible: LA, GO.

Task T.4. Generation of the user software

Objectives: This task is that devoted to the generation of an appropriate interface once the system of recognition and classification is fully operative. Here there should appear the specific needs of the medical specialists and of the practical possibilities that must be required of the system when it comes into real operation. Duration: This task will occupy this entire last year of the Project. In fact, it will remain finally open (as in the previous case), since the fact that the expected results were covered completely will mean that the unit constructed is indeed operative, and therefore dependent on the needs and requirements of users. Hence, it is anticipated that it will occupy the entire third year. Personnel: The whole team participates in this task, since it will include the revisions of the functioning and implementation of the different versions that might be of interest in view of the different applications that will be covered. The only exception is of the members engaged in in the following task (LA, PC, and PP). Person responsible: LA, PP/GU.

Task T.5. Offer for testing in practical professional clinical equipment

Objectives: At the present stage of the development that we have now been doing for almost three years, the main objective is the possibility of the application of our work in commercial clinical diagnosis equipment. For that reason, the last semester of the third year will be used to try to implement its operation within the process of a commercial scanner, and, if the result is positive, opening up its use to whichever industrial companies might be interested. Duration: In order to achieve our objective, we will initiate this activity during the last semester. Obviously, this task is included as part of the Project since it should be its main objective, but it is expected that its transcendence will spread outside the time limits and will serve to

open up new possibilities and applications. Personnel: This task must involve trying to cover practical needs deriving from the habitual medical use of these systems. For that reason, we are thinking in terms of a commission, formed by and LA and PP or GU. Person responsible: LA, PP/GU.

Task T.6 Documentation

Objectives: This is the aspect that is always reserved for the end, and its mission is obviously the generation of the information necessary for the use of the system by medical personnel and technical maintenance services. Duration: In order to attain our objective, we will initiate this activity during the last semester. Personnel: This task involves covering the practical needs deriving from the habitual medical use of these systems. For that reason, we are thinking in terms of a commission comprising all the members of the group, who will participate in one form or another. Person responsible: LA, PP/GU.
The Table presents the timetable summarizing the time distribution of the Project.

Timetable of the Project

Activities / Tasks	First Year (04/01/2005 - 03/31/2006)											
T.1. Processing Algorithms	X	X	X	X	X	X	X	X	X	X	X	X
T.2. Adaptation of the Associative Network	X	X	X	X	X	X	X	X	X	X	X	X
T.6. Documentation										X	X	X
	Second Year (04/01/2006 - 03/31/2007)											
T.2. Adaptation of the Associative Network	X	X	X	X	X	X	X	X	X	X	X	X
T.3. Analysis and Integration	X	X	X	X	X	X	X	X	X	X	X	X
T.6. Documentation										X	X	X
	Third Year (04/01/2007 - 03/31/2008)											
T.2. Adaptation of the Associative Network	X	X	X	X	X	X	X	X	X	X	X	X
T.4. User Software	X	X	X	X	X	X	X	X	X	X	X	X
T.5. Testing in Clinical Environments	X	X	X	X	X	X	X	X	X	X	X	X
T.6. Documentation							X	X	X	X	X	X

2. Level of success achieved in the Project

Reality, as so often occurs, surpasses the initial expectations, and imposes its own pathways for development. In our case, we shall describe the level of success achieved in the Project, bearing in mind that at the present time it is close to reaching 2/3 of its execution, i.e., still before finalizing the second year, and therefore in the expectation of initiating the third and last.
In any case, to give a specific and exhaustive description of the work and its achievements, we can say that we have attained a System for the Early Detection of Breast Cancer that is totally functional, and that has even allowed us to make the jump to its industrial exploitation. In the following sections, we shall present the most outstanding aspects by way of synthesizing the accomplishment of the different tasks.

2.1. Tasks

As is detailed in the above timetable, the first two years are foreseen to obtain the mammogram processing algorithm and to be in the process of adapting the Associative Neural Network to that algorithm. In this regard, it must be made clear from the outset that the initial request included the aid of a support technician for most of the software development tasks, and even for some cases of assembling the hardware units. Nevertheless, in the award of the grant, this collaboration was not approved, and the responsibility for this work fell to the researchers themselves. Even in this part of the work, collaboration has been unequal (as can be verified from the References), having to make recourse to other forms of attaining the objectives.
Nevertheless, even before finalizing these first two years, we can state that the tasks have been accomplished in their totality, and that at this present time we have available the following configurational elements of the system:

Level of Accomplishment of the Tasks

T.1. Processing Algorithms	Completed
T.2. Adaptation of the Associative Network	Sections 1 and 2 completed
T.3. Analysis and Integration	On schedule
T.4. User Software	Under development
T.5. Testing in Clinical Environment	In progress at the Mérida Hospital
T.6. Documentation	On schedule

3. Indicators of results

Aside from the conventional indicators, publications, etc., which are included in the following section, the element that is the clearest indicator of the development of the Project is its practical repercussion: the constitution of a Spin-Off company incubated by the University of Extremadura and constituted under the sponsorship of the Sociedad Gestora de Inversiones Innovadoras (SGII), a participation of diverse industrial and commercial organizations, with the shareholding intervention of the Servicio Extremeño de Salud (Extremadura Health Service, SES) through its enterprise trust "FundeSalud".
The end result is the constitution of the company Tecnodiag Médica, S.L., whose Web site can be consulted at www.tecnodiag.com. The purpose of the company is the exploitation of the results of the series of Research Projects that our group has been carrying out in this field for nine years now.
Figure 1 shows the aspect of the system that was finally constructed under the name of "DipreCam" with the Tecnodiag company, and currently in a period of tests in the Hospital of Mérida.

As can be seen, the system consists of a portable "DipreCam" unit (on the ground under the table) and two screens for viewing the images: the first, black & white and high resolution (2 Mpixels or 5 Mpixels), and the second, colour, also of high quality (but less than the first). The first is used to display the mammograms and the marks indicating the detected ROIs, and second for the specific analysis of each ROI in particular.
Although its specific use may not be of interest in this report, but by way of a simple example,

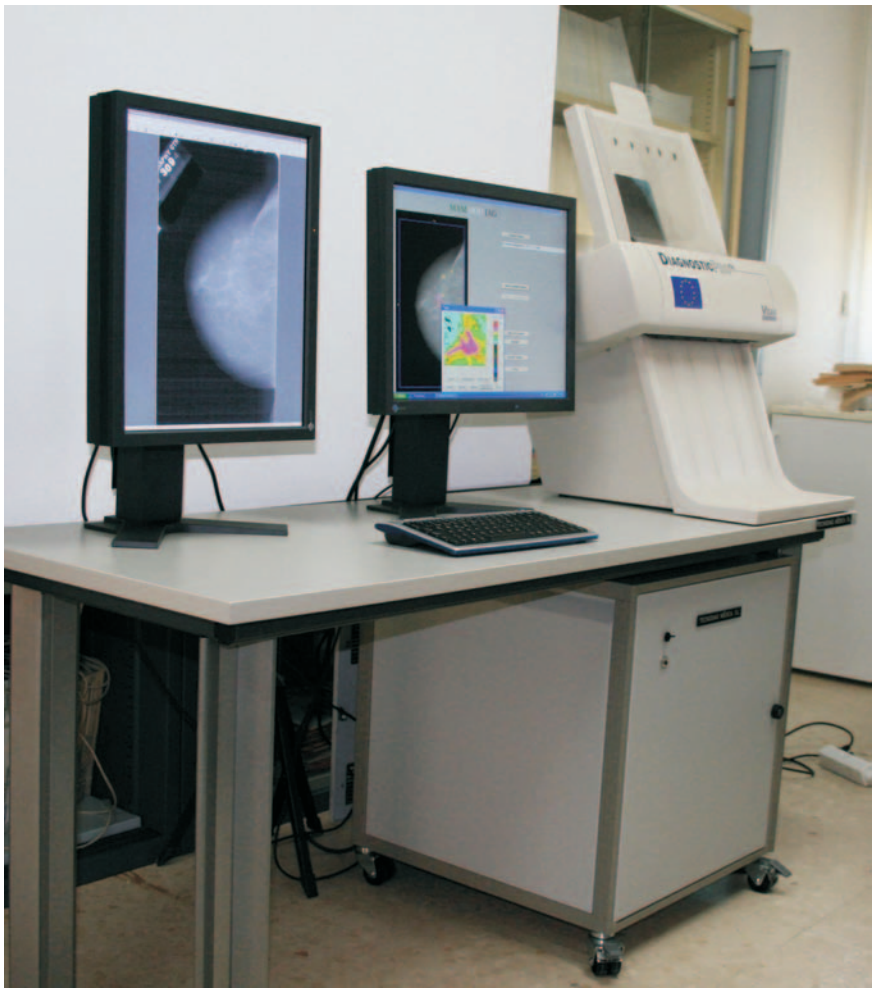


Figura 1: Overall aspect of DipreCam, the practical implementation of the results of the IACA-MA Project (TIN2004-00986). The system as shown is ready to work in clinical environments.

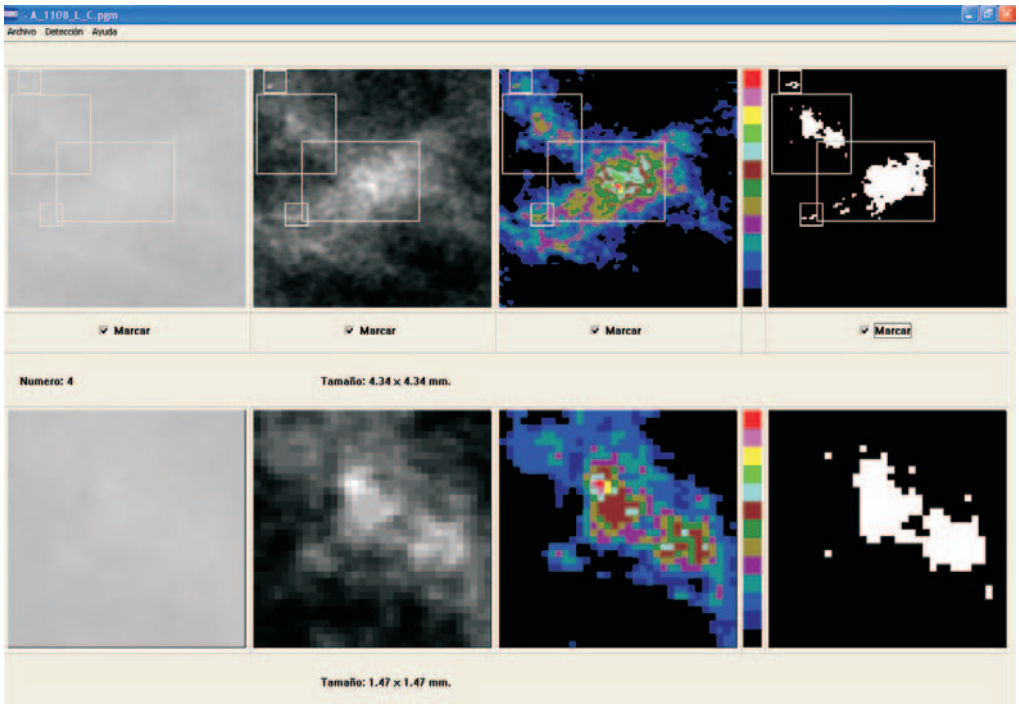


Figura 2: The information offered by DipreCam to the radiologist. The upper screens correspond to the detected ROI (number 4), and the lower to a calcification in that ROI. See the different sizes of each image: 4.34 x 4.34 mm the ROI and 1.47 x 1.47 mm the selected calcification.

Fig. 2 shows the aspect of the analysis screen (in colour) following the screening of a mammogram. In this image there appear the data relative to the ROI, such as its number within the total of the mammogram (here it is number 4), the size, and other information. The figure shows four possible displays: original image, enhanced, in false colour, and finally with only the detected calcifications.
The lower part of the figure shows the same type of detail, but now referred to one calcification in particular, here that located in the upper left quadrant. Among other things, there appears its size and the same four displays - original, enhanced, false colour, and calcification.

In this case, the calcification is of a size comparable to that of the ROI, for which reason it might give the impression that the second row of images clarifies little with respect to that of the ROI. Nevertheless, when numerous calcifications are detected in an ROI (which is quite usually the case), their size of these goes below 1 mm on a side. The second row then allows details to be to appreciated that would be impossible in any other way.

The following section lists the publications and works presented to congresses that were obtained as results of the work in this Project. For further detail, see the Web site www.tecnodiag.com.

4. Published works

Relation of published papers or works presented at some International Congresses, from 2004 up to 2006.

- MICROCALCIFICATIONS DETECTION IN DIGITAL MAMMOGRAMS. Francisco J. López Aligué, M. Isabel Acevedo Sotoca, Carlos J. García Orellana. *Proceedings of the 26th International Conference of the IEEE EMBS*. San Francisco (USA) september 2004.
- MAMMODIAG. Ponencia invitada. López Aligué, F. J., García Urra, F., Poveda Pierola, A., *XIV Certamen Internacional de Cine Médico, Salud y Telemedicina*. Videomed 2004. Badajoz, november 2004.
- SÍNTESIS DE MEMORIAS ASOCIATIVAS CON REDES NEURONALES. Francisco J. López Aligué, Isabel Acevedo Sotoca. *SICO 2005*. Granada, september 2005.
- REDES NEURONAIS E RADIOLOGIA. Invited session in the *II Congreso Ibérico de Telemedicina*. Viseu (Portugal) , october 2005.
- INDEPENDENT COMPONENT ANALYSIS APPLIED TO BREAST CANCER DETECTION ON DIGITIZED MAMMOGRAMS. *ICS Elsevier. Vol. 1281. pp. 1052 - 1057*, 2005. Ramón García Caballero, Carlos García Orellana, Francisco J. López Aligué.
- TECNODIAG MÉDICA, S.L. Available at *www.tecnodiag.com*. 2006.
- MICROCALCIFICATIONS DETECTION FOR EARLY BREAST CANCER DIAGNOSTIC. *Computer Aided Radiology and Surgery*, Francisco Javier López Aligué. Alberto Poveda Pierola. Isabel Acevedo Sotoca and Fernando García Urra. Berlin, June 2007.
- DETECTION OF MICROCALCIFICATIONS IN DIGITAL MAMMOGRAMS. *IEEE Engineerign in Medicine and Biology EMBC*, Francisco Javier López Aligué. Alberto Poveda Pierola. Isabel Acevedo Sotoca and Fernando García Urra. Lyon, August 2007.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Mejora y nuevos desarrollos de un sistema inteligente para la extracción y GEStión del CONocimiento en bases de Datos Ambientales (GESCONDA-II) TIN2004-01368

Miquel Sánchez-Marrè, *Karina Gibert*, Mario Martín, *Ramon Nonell*[*]
Knowledge Engineering & Machine Learning Group (KEMLG) / *Dept. of Statistics and Operations Research*
Technical University of Catalonia

Ignasi Rodríguez-Roda, Joaquim Comas[**]
Chemical and Environmental Engineering Laboratory (LEQUIA)
University of Girona

Abstract

The main goal of the project is to complete and to broaden the existing prototype (GESCONDA, TIC2000-1011) to improve, design and develop new components for a tool for intelligent data analysis and implicit knowledge management of data bases, with special focus on environmental data bases. Currently, that software prototype is composed by several statistical data filtering analysis methods, such as one-way and two-way descriptive statistics, missing data analysis, clustering, relationship between variables, hybrid Artificial Intelligence and Statistical methods, as well as several machine learning techniques, coming from Artificial Intelligence, such as conceptual clustering methods, decision tree induction and classification rule induction. The project purpose is to extend these intelligent system with some new agents and computational modules, such as case-based reasoning techniques, soft computing methods, support vector machines approaches, statistical models, dynamical analysis techniques, and hybrid methods integrating Artificial Intelligence approaches and Statistical ones.

Keywords: Intelligent Data Analysis, Knowledge Discovery and Management, Environmental Data Bases.

1 Project Goals

Major issue of the current project is to complete and to extend the prototype of intelligent data analysis tool resulting from the previous research project "Development of an intelligent system for knowledge management in Environmental data bases" (GESCONDA, TIC2000-1011).

[*] Email: {miquel|mmartin}@lsi.upc.edu, karina.gibert@upc.edu, ramon.nonell@upc.edu
[**] Email: {ignasi|quim}@lequia.udg.es

The concrete objectives of the project are the necessary research and development of hybrid, techniques and methods originated in the fields of Statistics and Artificial intelligence, specially in the area of inductive machine learning, statistical modelling, soft computing methods like artificial neuronal networks or support vector machines, case-based reasoning, meta-knowledge and knowledge patterns management, *etc.* All these techniques will be integrated within the previous prototype of an intelligent data analysis and knowledge management tool. All these techniques will be implemented within several modules or agents, who constitute the concrete objectives of the project, as depicted in figure 1, which shows the proposed architecture of GESCONDA-II. Due to the huge reduction for the personal funding, and also to the evolution of the research and implementation tasks, it has been changed as it will be explained in the next section (section 2).

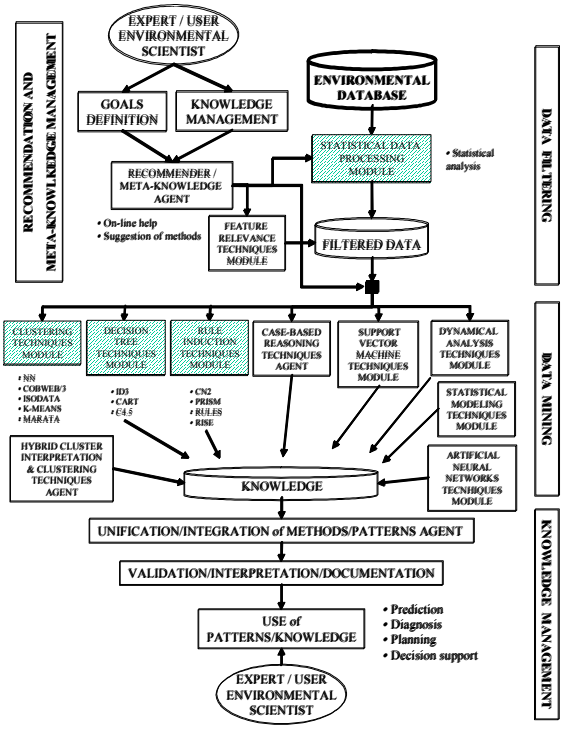


Fig. 1. Initial architecture proposed for GESCONDA-II

To achieve the issues of the project we had a major serious problem from the beginning of the previous project (TIC2000-1011), when no funding was assigned for labour tasks. This means, that nobody else than ourselves could be contracted to be involved in the project. This feature caused already in the previous project that some goals and tasks were cut. Having done this remark, we can state that we have the tools necessary for the development of the project to a determined extent. Basically, the personal directly involved in the project (4.5 EDPs), which is taking the research leading goal, some students which are collaborating without any economic compensation and the computing and research infrastructure, mainly computers, printer consumables and some books as a reference works, to undertake the project.

The chronogram of the tasks and subtasks is depicted in figure 2. There is an estimation of the amount of work done, and also, there are the tasks which are stopped, and the new tasks emerging from the project.

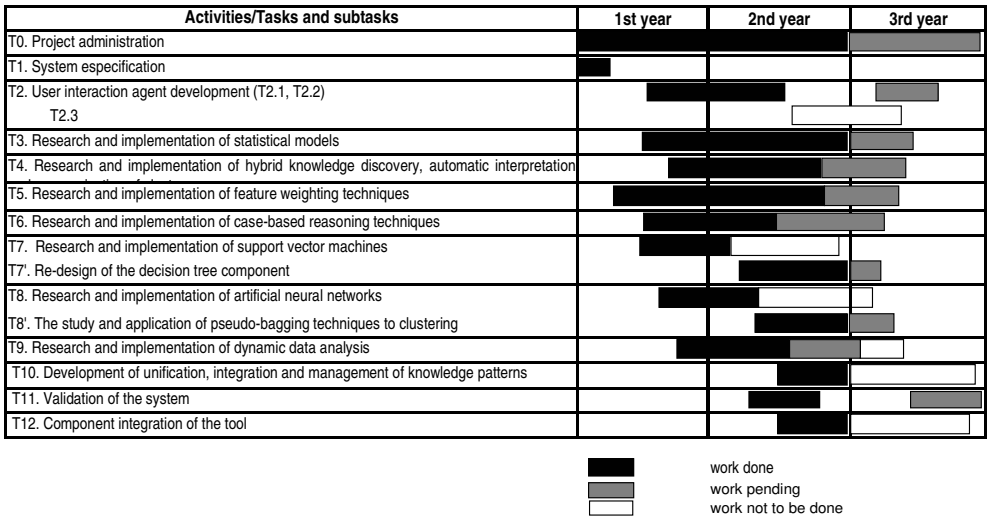


Fig 2. Estimated task chronogram for the project

2 Development and success of the project

In general, and taking into account the remark made in section 1, the development of the project follows the planned schedule, with some reductions, additions and changes in the scope of several tasks and subtasks. The activities made to accomplish the sub-goals of the project in each one of the subtasks are the following:

T0. Project administration: Management tasks and scientific coordination is advancing as it was planned. Relating to results diffusion, the following facts should be stressed: a new course within the interuniversity (UB-UPC-URV) Master's degree in Artificial Intelligence has been launched during academic year 2006/2007, about Intelligent Decision Support Systems, where the GESCONDA prototype is being used and validated; several international publications have been made during last year 2006 (see reference section); The web page for the project[1] have been continuously improved. Project outcomes and utility of the tool have been presented in several international events such as in iEMSs'2004 and iEMSs'2006 conferences [2, 5-7, 14, 23, 26-27, 34], both from the theoretical point of view and from the environmental application side, and two international workshop were organised in 2004 (BESAI'2004) [29] and 2006 (BESAI'2006) [3] within ECAI conferences, where the some of the results project were presented and discussed.
T1. System specification: In first months of the project, the initial architecture for the GESCONDA-II system was deeply analyzed and discussed. Some changes were decided upon the knowledge of what has been explained in section 1, and with some incompatibility problems

[1] http://www.lsi.upc.edu/~webia/KEMLG/projects/GESCONDA

among the different software modules implemented in parallel. Thus, the final architecture, the final software components and the techniques to be implemented were finally decided by the project team, and described in some publications as the project was progressing [14, 23, 34].

T2. User interaction agent development: The subtasks 2.1 and 2.2 have been mostly finished. The goals and the meta-knowledge agent have been designed to manage knowledge about the attributes, about the examples and about the domain knowledge available. This last item has originated a very probable Master thesis to be developed in year 2007. The subtask about the recommender will not be implemented in the GESCONDA-II prototype due to its complexity and its dependence on new methods not yet available, even though some research on recommendation algorithms has been done [17, 22]. It will be faced in the next GESCONDA-III phase.

T3. Research and implementation of statistical models: The statistical methods to be implemented in the prototype were decided. By now, the multiple linear regression method is already implemented. The analysis of variance of factors has been recently implemented. During the lat year of the project, the logistic regression and the multivariate analysis techniques will be at least studied, and if possible will be implemented. If not, they will be implemented in the next phase of the project.

T4. Research and implementation of hybrid knowledge discovery, automatic interpretation and summarization of clusters: All techniques within the subtasks have been deeply studied and analysed, as several publications show [1, 8, 11-13, 15, 18-19, 21, 32]. Implementation of the techniques will be undertaken in the last year of the project.

T5. Research and implementation of feature weighting techniques: The study and research of feature weighting assignment techniques has already been done [4]. The unsupervised methods implemented are the gradient method, UEB-1 and UEB-2 methods [9, 31]. In the next months the supervised methods will be implemented. Thus, it is a little bit delayed, but no serious problems are envisioned. The task 5.3 about the implementation of feature weighting adjustment methods will not be undertaken. Instead, some experimentation about feature weighting algorithms recommendation has been done [17, 22].

T6. Research and implementation of case-based reasoning techniques: Although the project team have undertaken much research on case-based reasoning [20, 24-25, 28, 33], the task is a little bit delayed, due to some more urgent new tasks that must been done. The subtask 6.1 is completed, and we have to implement the methods in the prototype and the validation procedures in 2007.

T7. Research and implementation of support vector machines: As it was already decided during the first months of the project, and stated in the last report, both this task and task T7 have reduced their scope. Subtask 7.1 has been accomplished, where the most suitable SVM methods were evaluated, but they will be implemented in the next step of GESCONDA development. Instead, a re-design task of the decision tree component has been started (T7'), and it will be completed in the next months. Re-design was necessary as the software design was not homogeneous with the other modules, and some problems of integration must be solved.

T8. Research and implementation of artificial neural networks: Also T8 has reduced its scope. After a revision and selection of the suitable ANN models to be integrated in the future, the task has been stopped. Nevertheless, another task (T8') has emerged as a more interesting to complete the clustering component: the study and application of pseudo-bagging techniques to clustering processes, in K-means and nearest neighbour methods [10]. The bagging techniques have been integrated into the clustering component.

T9. Research and implementation of dynamic data analysis: Some research has been done in dynamic data analysis, such as in transition networks, and in dynamic CBR techniques [7, 20, 24,

30]. During the last year of the project, these techniques will be implemented. Micro dynamic analysis techniques and time series will not be studied, due to the reasons stated before.
T10. Development of unification, integration and management of knowledge patterns: Within this task, some research has been done in the automatic interpretation of results, especially in clustering processes [1, 8, 11, 19] and in the near future a knowledge pattern manager will be designed and implemented, probably within a new Master's thesis.
T11. Validation of the system: The validation of the different components developed and implemented is progressing as expected, and in the last year of the project the some global validation of the system will be done.
T12. Component integration of the tool: As previous validation tests showed, some major problems in the integration of the different modules of the tool were detected (format of database files, etc.). To that end, this new task not planned in the original schedule, has been originated, and it is successfully progressing.

3 Results

The results of the project are good and relevant to some extent as several indicators show. In this period, two special issues on Environmental Sciences and Artificial Intelligence have been edited in AI Communications journal [16] and in Environmental Modelling and Software journal [35]. Also, we have organised 2 BESAI workshops within ECAI main conferences, many research publications were made, some students were formed, the project team participates in several international projects and relationships with some international and national groups were made. In the next subsections all this information is summarised.

3.1 Graduate and undergraduate students formation

Several people were formed within the context of the project. Two students developed the graduate project under our supervision. Very probably, 1-2 students will undertake the Master's thesis with our team. Two students obtained the Advanced Study Diploma (DEA): one in the Ph.D. programme on Artificial Intelligence and one in the Ph. D. programme of Statistics under the supervision of our team, and another will obtain it during 2007. One student did his PH.D under our supervision and three are currently working in her/his Ph. D. thesis, which will last until 2007-2008.

3.2 Publications

International publications made during the project development can be summarised in two special issues on journals, 2 Workshop notes edition, 3 chapters of book, 7 journal papers, 18 international conferences, 2 national conferences, 2 research reports. They are listed in the references' section.

3.3 Technological transference

There are some perspectives about transferring the software developed both to universities and research centres for scientific purposes, and to some companies that could be interested in it, such as "Depuradores d'Osona S.L." and "Sanejament Intel·ligent S.L." which expressed their interest in

the software system as a knowledge discovery and management tool for their application in environmental domains.

3.4 Participation in international projects

Members of the project team participate in several international projects and support action related with the project to some extent, such as in European projects: Argumentation Service Platform with Integrated Components (ASPIC, IST-2004-002307). 2004-2007 (using case-based reasoning for argumentation), Enabling and Supporting Provenance in Grids for Complex Problems (PROVENANCE, IST-511085) 2004-2006 (an architecture based on open data model), Widening IST Networking Development Support - Latin America (WINDS, IST-45108 SSA)
Also, two proposal for European international projects were recently submitted in past October 2006: eContent for advanced environmental education: distributed web-based repository, multi-media toolkit and teaching tools (eCAMPUS) submitted to IST track of European Union and Summer School on Environmental Sciences and Information technologies (SESSIT) submitted to the European Science Foundation (ESF). In both projects the validation of GESCONDA-II could be done.

3.5 Collaboration with national and international groups

The project development has generated new collaborations through the attendance of several major conferences where the advances of the project were presented (iEMSs'06, BESAI'06, ECAI'06, ICCBR'05, iEMSs'04, BESAI'04, ECAI'04), and some other relationships have been strengthen, which will cause an improvement in the output of the project. Collaboration with European, North and South American, Oceanian, Mexican and Chinese groups have been fostered: Centre for Scientific and Technological Research, Istituto Trentino di Cultura (IRST-ITC), Italy; Institute for Software technology, Technical University of Graz, Austria; Environmental Software & Services, Austria; Faculty of Sciences and Technology, New University of Lisbon, Portugal; Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA), Switzerland; Faculty of Civil and Geodetic Engineering, University of Ljubljana, and Dept. of Intelligent Systems, Josef Stefan Institute, Slovenia; Dept. of Environmental Sciences, University of Kuopio, Finland; Dept. of Computer Science, Technical University of Munich, Germany; RWTH Aachen University, Germany; Dept. of mechanical Engineering, Aristotle University of Thessaloniki, Greece; Dept. of Engineering Science, Oxford University, England, UK; Human Computer Studies Laboratory, University of Amsterdam (UvA), The Netherlands; Department of Informatics, University Petroleum-Gas from Ploiesti, Romania; Laboratory of Environmental Biotechnology (LBE), INRA Narbonne, France; Ecology Dept., Montana State University, USA; Computing Science Dept., Federal University of Amazonas, Brazil; Centre for Resources and Environmental Studies (CRES), and Integrated Catchment Assessment and Management Centre (iCAM), Australian National University, Australia; Dept. of Computer Science, University of Waikato, New Zealand: Depto. de Sistemas Computacionales, Universidad de Las Américas (UDLA), Puebla, México; Dept. of Computer Science and Technology, Nanjing University, P.R. China.
Also, through some national conferences (CCIA'06, TAMIDA'05, CCIA'2005, CAEPIA'2003, and CCIA'2003) and the following Spanish and Catalan thematic networks, some new links have been created: "Data mining and Machine Learning" (TIC2002-11124-E, TIN2006-27675-E); "Integral

management of river basins" (2003XT 00045). Dirección General de Investigación, Generalitat de Catalunya; "Monitoring and modelling as a support tools for improving water quality in Catalonia" (1999XT 0037, 2001XT 0030, 2003XT 0069). Dirección General de Investigación, Generalitat de Catalunya.

References

[1] K. Gibert & A. Pérez-Bonilla y G. Rodríguez-Silva. *A comparative analysis of different classes-interpretation support techniques*. 9th Catalan Conference on Artificial Intelligence (CCIA'2006). Frontiers in Artificial Intelligence and Applications Series, Vol. 146, pp. 37-46. IOSS Press. ISBN 1-58603-663-7. Octubre de 2006. Perpinyà, Francia.

[2] T. Rendón, M. Sànchez-Marrè, M. Aulinas y J. Comas. *Designing a Multi-Agent System to Simulate Scenarios for Decision Making in River Basin Systems*. 9th Catalan Conference on Artificial Intelligence (CCIA'2006). Frontiers in Artificial Intelligence and Applications Series, Vol. 146, pp. 291-298. IOSS Press. ISBN 1-58603-663-7. Octubre de 2006. Perpinyà, Francia.

[3] M. Oprea, M. Sànchez-Marrè y F. Wotawa. *Workshop Notes on 5th ECAI'2006 Workshop on Binding Environmental Sciences and Artificial Intelligence (BESAI'2006)*. 17th European Conference on Artificial Intelligence (ECAI'2006). Riva del Garda, Italia, Agosto de 2006.

[4] K. Gibert, M. Sànchez-Marrè y J. Comas. *The Impact of Feature Weighting in Environmental Class Discovery using GESCONDA*. 5th ECAI Workshop on Binding Environmental Sciences and Artificial Intelligence (BESAI'2006), within ECAI'2006. BESAI'2006 Workshop Notes, pp. 5:1-5:10. Riva del Garda, Italia. 28 de Agosto-1 de Septiembre de 2006.

[5] T. Rendón, M. Sànchez-Marrè, F. Devesa y M. Poch. *Simulating scenarios for decision-making in river basin systems through a Multi-Agent System*. 4th International Environmental Modelling and Software Society Conference (iEMSs'2006). iEMSs'2006 Procceedings, pp. 253-258. ISBN 1-4243-0852-6. Burlington, Vermont, USA. Julio de 2006

[6] J.M. Spate, K. Gibert, M. Sànchez-Marrè, E. Frank, J. Comas, I. Athanasiadis y R. Letcher. *Data Mining as a Tool for Environmental Scientists*. 4th International Environmental Modelling and Software Society Conference (iEMSs'2006). iEMSs'2006 Procceedings, pp. 1270-1291. ISBN 1-4243-0852-6. Burlington, Vermont, USA. Julio de 2006

[7] M. Sànchez-Marrè, K. Gibert, R. Sojda, J. P. Steyer, P. Struss and I. Rodríguez-Roda. *Uncertainty Management, Spatial and Temporal Reasoning and Validation of Intelligent Environmental Decision Support Systems*. 4th International Environmental Modelling and Software Society Conference (iEMSs'2006). iEMSs'2006 Procceedings, pp. 1352-1377. ISBN 1-4243-0852-6. Burlington, Vermont, USA. Julio de 2006

[8] K. Gibert & A. Pérez-Bonilla. *Revised Box-plot based discretization as the kernel of automatic interpretation of classes using numerical variables*. Data Science and classification, Springer, pp. 229-237. Julio de 2006

[9] M. Sànchez-Marrè, S. Gómez, F. Teixidó y K. Gibert (2006). *Técnicas de Feature Weighting para casos no supervisados. Implementación en GESCONDA* (en catalán). Documento de trabajo LSI-06-03-T. Depto. de Llenguatges i Sistemes Informàtics (UPC). Julio de 2006.

[10] K. Gibert, L.J. Oliva, I. Piñol y M. Sànchez-Marrè. *El bagging en casos no supervisados. Implementación en GESCONDA para algoritmos de clustering* (en catalán). Documento de trabajo LSI-06-02-T. Depto. de Llenguatges i Sistemes Informàtics (UPC). Julio de 2006.

[11] K. Gibert & G. Rodríguez-Silva & G. Sánchez. *Análisis comparativo de distintas técnicas de apoyo a la interpretación de clusters en un caso real.* Contribuciones a la Estadística y a la Investigación Operativa. Edita SEIO, Universidad de La Laguna, pp. 1394-1408, Mayo 2006.

[12] A. Pérez-Bonilla & K. Gibert. *El papel del box-plot based discretization en la interpretación conceptual de una classificación jerárquica*. Contribuciones a la Estadística y a la Investigación Operativa. Edita SEIO, Universidad de La Laguna, pp. 479-493, Mayo 2006.

[13] K. Gibert & A. Pérez-Bonilla. *Towards automatic generation of interpretation as a tool for modelling decisions*. 3rd Modelling Decision for Artificial Intelligence (MDAI). Publicación electrónica en CD-ROM, Tarragona, Abril de 2006.

[14] K. Gibert, M. Sànchez-Marrè y I. Rodríguez-Roda. *GESCONDA: An Intelligent Data Analysis System for Knowledge Discovery and Management in Environmental Data Bases*. Environmental Modelling & Software 21(1):115-121. Enero de 2006.

[15] K. Gibert, M. Sànchez-Marrè y X. Flores. *Cluster discovery in environmental databases using GESCONDA: The added value of comparisons*. AI Communications 18(4):319-331. Diciembre 2005.

[16] M. Oprea, M. Sànchez-Marrè y Franz Wotawa. *Guest editorial Binding Environmental Sciences and Artificial Intelligence*. AI Communications 18(4):243-245. Diciembre 2005.

[17] H. Núñez y M. Sànchez-Marrè. *A Case-Based Methodology for Feature Weighting Algorithm Recommendation*. 8th Catalan Conference on Artificial Intelligence (CCIA'2005). Frontiers in Artificial Intelligence and Applications Series, Vol. 131, pp. 223-230. ISBN 1-58603-560-6. Octubre de 2005. L'Alguer, Cerdeña, Italia.

[18] K. Gibert, X. Flores y M. Sànchez-Marrè. *Comparación de clasificaciones en bases e datos ambientales utilizando GESCONDA*. III Taller Nacional de Minería de Datos y Aprendizaje (TAMIDA'2005), pp. 299-308. Edición electrónica en CD. 13-16 de Septiembre de 2005, Granada.

[19] K. Gibert y A Pérez-Bonilla. *Ventajas de la estructura jerárquica del clustering en la interpretación automática de clasificaciones*. III Taller Nacional de Minería de Datos y Aprendizaje (TAMIDA 2005). pp. 67-76. Edición electrónica en CD. 13-16 de Septiembre de 2005, Granada.

[20] M. Sànchez-Marrè, U. Cortés, M. Martínez, J. Comas y I. Rodríguez-Roda. *An Approach for Temporal Case-Based Reasoning: Episode-Based Reasoning*. 6th International Conference on Case-Based Reasoning (ICCBR'2005). Springer-Verlag, Lecture Notes in Artificial Intelligence 3620, pp. 465-476. Chicago, USA. ISBN 3-540-28174-6. Agosto de 2005.

[21] K. Gibert, A. Pérez-Bonilla. *Fuzzy Box-plot based induction rules. Towards automatic generation of classes-interpretations*. En Procc. of EUSFLAT 2005, pp 524-529. Septiembre de 2005, Barcelona.

[22] H. Núñez, M. Sànchez-Marrè. *Una metodología de recomendación de algoritmos de asignación de pesos basada en casos*. En el libro Tendencias de la Minería de Datos en España (Eds. R. Giráldez, J. C. Riquelme, J. S. Aguilar-Ruiz). Ed. Digital3, Sevilla. Diciembre de 2004. ISBN 84-688-8442-1

[23] M. Sànchez-Marrè, K. Gibert and I. Rodríguez-Roda. *GESCONDA: A Tool for Knowledge Discovery and Data Mining in Environmental Databases*. En el libro e-Environment: Progress and Challenge. Research on Computing Science, Vol. 11 pp. 348-364. CIC, Instituto Politécnico Nacional, México D.F., México. Noviembre de 2004. ISBN 970-36-0215-0.

[24] M. Martínez, M. Sànchez-Marrè, J. Comas y I. Rodríguez-Roda. *Case-Based Reasoning, a Promising Tool to Face Solids Separation Problems in the Activated Sludge Process*. 6th International Symposium on Systems Analysis and Integrated Assessment in Water Management (WATERMATEX' 2004). Proc. of Watermatex 2004, pp. 213-220. Beijing, China. 3-5 de Noviembre, 2004

[25] H. Núñez, M. Sànchez-Marrè, U. Cortés, J. Comas, M. Martínez, I. Rodríguez-Roda y M. Poch. *A Comparative Study on the Use of Similarity Measures in Case-Based Reasoning to improve the*

classification of Environmental System Situations. Environmental Modelling & Software 19(9):809-820. Septiembre de 2004.

[26] M. Poch, J. Comas, I. Rodríguez-Roda, M. Sànchez-Marrè y U. Cortés. *Designing and Building real Environmental Decision Support Systems*. Environmental Modelling & Software 19(9):857-873. Septiembre de 2004.

[27] L. Ceccaroni, U. Cortés y M. Sànchez-Marrè. *OntoWEDSS: augmenting environmental decision-support systems with ontologies*. Environmental Modelling & Software 19(9):785-798. Septiembre de 2004.

[28] M. Sànchez-Marrè, Martinez M., Rodríguez-Roda I., J. Alemany y C. Cortés. *Using CBR to improve intelligent supervision and management of wastewater treatment plants: the atl_EDAR system*. 7th European Conference on Case-Based Reasoning (ECCBR'2004) (Industrial day). Proc. of Industrial day, ECCBR'2004, Madrid, pp. 79-91. 30 Agosto – 2 Septiembre 2004.

[29] Mihaela Oprea y M. Sànchez-Marrè. *Workshop Notes on 4th ECAI'2004 Workshop on Binding Environmental Sciences and Artificial Intelligence (BESAI'2004)*. València, Agosto de 2004.

[30] J.M. Gimeno, J. Béjar, y M. Sànchez-Marrè. *Nearest-Neighbours for Time Series*. Applied Intelligence 20(1):21-35. Enero de 2004.

[31] H. Núñez y M. Sànchez-Marrè. *Instance-based Learning Techniques of Unsupervised Feature Weighting Do not Perform So Badly!* 16th European Conference on Artificial Intelligence (ECAI'2004). ECAI'2004 Proceedings, pp. 102-106. IOS Press ISBN 1-58603-452-9. València. 22-27 de Agosto de 2004.

[32] K. Gibert, X. Flores, I. Rodríguez-Roda y M. Sànchez-Marrè. *Comparison of Classifications in Environmental Databases using GESCONDA*. 4th ECAI Workshop on Binding Environmental Sciences and Artificial Intelligence (BESAI'2004), within ECAI'2004. BESAI'2004 Workshop Notes, pp. 13:1-13:10. València. 22-27 de Agosto de 2004.

[33] M. Martínez, M. Sànchez-Marrè, J. Comas y I. Rodríguez-Roda. *Case-Based Reasoning, a promising Tool to Face Solids Separation Problems in the Activated Sludge Process*. 4th ECAI Workshop on Binding Environmental Sciences and Artificial Intelligence (BESAI'2004), within ECAI'2004. BESAI'2004 Workshop Notes, pp. 11:1-11:5. València. 22-27 de Agosto de 2004.

[34] K. Gibert, X. Flores, I. Rodríguez-Roda y M. Sànchez-Marrè. *Knowledge Discovery in Environmental Data Bases using GESCONDA*. 2nd International Environmental Modelling & Software Society Conference (iEMSs'2004). IEMSs'2004 Proceedings, pp. 35-40. ISBN 88-900787-1-5. Osnabrück, Germany, Junio de 2004.

[35] M. Sànchez-Marrè, U. Cortés and Joaquim Comas. *Guest editorial: Environmental Sciences and Artificial Intelligence*. Environmental Modelling & Software 19(9):761-762 Septiembre 2004.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

SINTA-CC: Adaptive Intelligent Systems for Modelling, Prediction and Dynamic Optimization in Clusters of Computers TIN2004-01419

Ignacio Rojas Ruiz*
Department of Computer Architecture and Technology, University of Granada,
E.T.S.I.I.T, 18071 Granada, Spain

Abstract

The project undertakes the study and development of parallel intelligent procedures and systems for modelling, prediction, and dynamic optimization in high performance computing platforms, analyzing the synergy between diverse paradigms like neuronal networks, evolutionary computation, fuzzy logic and matrix transformations, that allow problems concerning learning, adaptation to changing conditions and presence of uncertainty to be tackled properly. Under this scope, the project also tries to continue the work of research made in the development of tools which take advantage of parallelism (using toolboxes which we have developed for the parallel processing in Matlab™), and to improve the benefits of communication in clusters of computers and the dynamic management of protocols in computation environments of GRID type. Therefore, the project tries to make contributions in: (a) the development of new procedures of modelling, prediction, and optimization that take advantage of the softcomputing techniques and the calculation capabilities that clusters do offer; and (b) the use of the new procedures in realistic applications for which there exists available data provided by diverse companies; (c) the improvement of the communications and the conditions of use of clusters of computers; and (d) to allow the access to the developed procedures and platforms through Internet.
Keywords: High Performance Computation, Intelligent Network Interface Cards, Soft Computing, Hybridization of Metaheuristic

1 Research Project Goals

This project is a continuation of the research effort carried out through the following projects belonging to the National Plan of Scientific Research TIC2000-1348 and DPI2001-3219 all of them accomplished by members of the CASIP (Circuits and System for Information Processing) research group, for the last ten years, and all been favourably evaluated under the corresponding annual reports. The project TIN2004-1419 meant the unification of different research lines within the CASIP research group in order to reach more ambitious goals. The initial hypothesis for our

* Email: irojas@atc.ugr.es

work is the productive synergy existing within the soft-computing field when integrating in just one unique platform, paradigms such as neural techniques, fuzzy logic, co-operative and/or co-evolutionary algorithms in order to solve, on parallel platforms, computational intensive and complex problems related to function approximation, classification, forecast and optimization.

This global goal can be divided into three main objectives, with different task:

Objetive 1.- Development of new soft-computing-based parallel methodologies based for adaptive systems in modelling, prediction and optimization problems. This objective involves the following three stages:

a) Design and development of new intelligent architectures for complex system modelling with self-learning and self-organization capabilities in real time. The task for this first stage are:

- *T.1.1.1. Algorithms to compute the degree of importance of input variables using hybrid statistical and soft-computing-related techniques*
- *T.1.1.2. Design and development of new neural and/or fuzzy architectures for the efficient exploitation of complex multi-dimensional problems*
- *T.1.1.3. Parameter optimization of the proposed architectures in multi-dimensional ill-defined spaces*
- *T.1.1.4. Use of parallel platforms for the global optimization of the proposed system*
- *T.1.1.5. Conception and implementation of intelligent control systems with self-learning and self-organization capabilities in real time*

b) Development of hybrid methodologies for time series modelling, characterization and prediction

- *T.1.2.1. Time series prediction and modelling by hybridizing classic statistical techniques with soft-computing systems*
- *T.1.2.2. Advanced prediction systems using soft-computing techniques*
- *T.1.2.3. Parallelization of hybrid techniques for time series modelling, characterization and prediction*

c) Hybrid methodology for parallel and distributed dynamic optimization

- *T.1.3.1. Development of a new multi-objective dynamic optimization procedure based on evolutionary computation*
- *T.1.3.2. Multi-agent implementation of the multi-objective dynamic optimization procedure developed*
- *T.1.3.3. Integration of the modelling and prediction procedures designed in tasks T.1.1.3 and T.1.2.3 within the developed distributed dynamic optimization scheme.*

Objetive 2.- Implementation, analysis and design of efficient parallel platforms for the simulation of intelligent systems. In this objective, the research group has been working in four specific research lines:

a) Optimization of the Communication Operative System support. The main task was:

- *T.2.1 Optimization of the CLIC protocol for Gigabit Ethernet*

b) Effective implementation of communication protocols in intelligent NICs to alleviate the CPU of communication-related tasks

- *T.2.2 Development of a PCI card with re-configurable hardware (FPGAs) to implement communication protocols*

c) Dynamic management of different communication protocols

- *T.2.3.1 Implementation of TCP/IP and CLIC protocols in the designed card with the possibility of dynamically select the protocol to use by means of a message*
- *T.2.3.2 Assessment of the communication performance with a protocol dynamic management scheme*

d) Development in MatlabTM of specific software for the integration of the proposed methodologies. The following four task has been carried out:

- *T.2.4.1 Extension of the developed system to MPI 2.0*
- *T.2.4.2 Improvement of the proposed system adding multi-threading support*
- *T.2.4.3 Translation of the designed system into Octave*

Objetive 3.- Application and dissemination of the proposed intelligent methodologies and architectures in real problems: modelling, classification, time series prediction and real time control. The dissemination and real application of the results and parallel procedures investigated through the afore-mentioned objectives is of great importance to provide the accomplished work with truly practical utility, and analyze its viability and effectiveness. Our aim in this objective is to work in four specific research lines:

a) Modelling and prediction of the water and electricity consumption pattern in the city of Granada. Application to the robust and on-line intelligent control of the water-supply network

T.3.1.1 Modelling and prediction of the sectorized and the global water consumption pattern in Granada

T.3.1.2 Adaptive real-time control of the more relevant elements in the system

b) Bio-medical applications: classification of several skin pathologies and digital mamographies

T.3.2.1 Entropic detection/classification using intelligent systems from digital mamography

T.3.2.2 Diagnosis of skin diseases from fluorescence techniques

c) Management of inventory chains of supply systems: application to a book supplier.

T.3.3 Application to a book supplier

d) Implementation of a web page for the project: Interface for the access to parallel resources through the Internet

T.3.4 Implementation of a web page for the project that allows us the access and execution of modelling, prediction and dynamic optimization applications in the developed platforms

Finally, the estimated schedule of the task proposed in the project is the presented in Table 1.

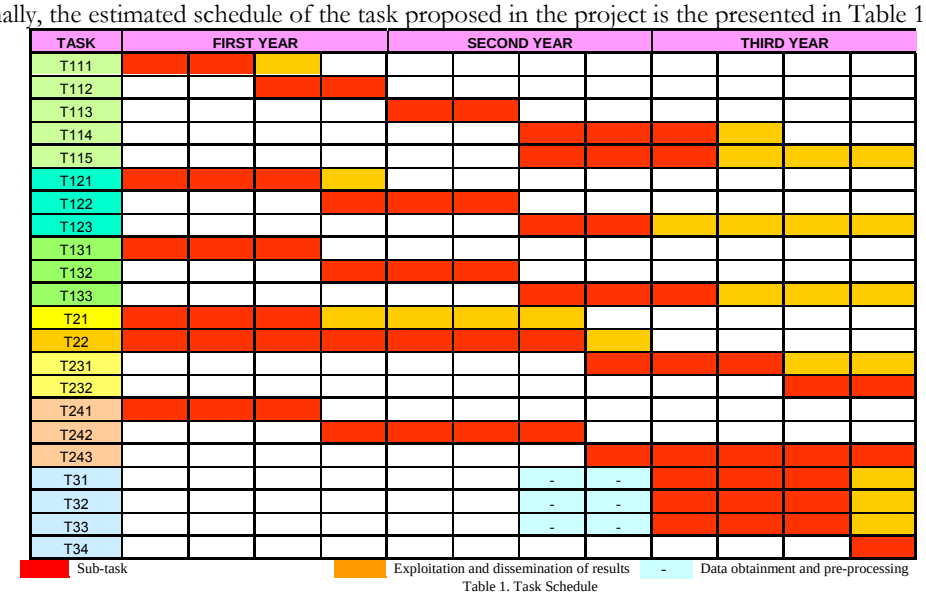


Table 1. Task Schedule

2 Successful achievements of the Project TIN2004-1419

The level of success of the research developed until now is very satisfactory and relevant. We have solved the nearly all the tasks described in the previous section (some tasks of the objective 3 are not completely finished). The next three subsections are devoted to describe the results of the main

objectives. For each of them, we highlight the most relevant scientific and technological results, together with the publications generated.

2.1 Objetive 1: Development of new soft-computing-based parallel methodologies based for adaptive systems in modelling, prediction and optimization problems

a) Advanced neuro-fuzzy models for function approximation and time series prediction problems

Takagi-Sugeno-Kang (TSK) neuro-fuzzy models are widely used to solve modelling problems, due is its representative power, capable of describing a highly complex nonlinear system using a small number of simple rules. However, the optimization of a TSK model normally leads to a good general performance of the whole TSK fuzzy system give the idea of what it does, but the sub-models given by each rule in the TSK fuzzy system can give no interpretable information by themselves. An intensive research has been performed to refine the treatment of the two main goals in the training and design of neurofuzzy models. A preliminary result was obtained in [1] that uses a modified TSK neuro-fuzzy model, using a grid partitioning of the input space, to obtain a set of fully interpretable rules without decreasing the general performance in the modelling of the I/O data. The interpretability of the local models was obtained by using the Taylor Series Expansion (TSE) concept, since each rule consequent provides the behaviour of the global around the rule centre (the TSE of the global model around the rule centre). This research was extended and published in [2], obtaining a final formulation for the model (TaSe model), and presenting interesting results, that also showed up that using high order rule consequents in the TSK model can provide a better generalization capability of the model, but without loosing interpretability of the local models, since their TSE of the higher order is obtained. An statistical ANOVA test was performed and presented in [3][4], that in fact confirmed that using higher order rules in a general TSK model is able to provide a better generalization capability for function approximation problems. The TaSe model was also satisfactorily applied for time series prediction problems as was published in [5][6].

Normally TSK neuro-fuzzy models present a grid partitioning of the input space, however, they can also present a clustering-partitioning of the input space, which is more associated with the equivalent Radial Basis Function Network (RBFN) model. An important research was also performed for this type of models. First a preliminary work to extract interpretable rules using also the TSE concept from RBFNs was presented in [7][14], which was extended and compared with grid-based TSK systems in [8]. A second phase of work, to improve the initialization of the rule centres before the complete optimization of the model was presented in EUNITE Competition, which showed to bring excellent results. The methodology is based on the initialization of the centres such that the error surface along the input space is flatted. The idea is also supported by the fact that the generalization of the model is expected to improve by homogenizing the error along the input domain.

An important drawback of neuro-fuzzy models in general is the curse of dimensionality. This is specially crucial in grid based fuzzy models since the number of obtained rules is exponential in the number of input dimensions of the problem. However, grid-based fuzzy systems are important since they perform a full coverage of the input space. It is also essential to obtain rules with a low number of antecedents in order to obtain better interpretable models. On those purposes, a modified grid-based fuzzy approach was presented in [9] (MultiGrid Fuzzy System -MGFS-), that performs a subdivision of the input space and obtains a grid-based TSK model for each subdivision, joining the all the rules output as is done for general TSK fuzzy models by

weighted average. A complete learning algorithm was designed for the MGFS model, which also is able to perform variable selection. The application of the learning methodology was first applied for time series prediction problems, in which the input space has to be selected too to perform the modelling. A first work was presented in [10], and the MGFS model was applied to the CATS Benchmark proposed in the IJCNN conference in 2004, getting a good position in the time series competition [11], and being selected the work for publication in the CATS Neurocomputing Special Issue. The extended paper with the complete learning methodology exposition was accepted for publication [12].

It is also important to note the research carried out using multiobjetivo evolutionary algorithms, cooperative-coevolutionary optimization, and hybrid methods for function approximation and time series forecast [49]-[55], and its parallelization using multi-objective metaheuristics [69][70].

b) LS-SVM and Neuro-Fuzzy Models and long term Time Series Prediction

Time series forecasting is a challenge in many fields. It is a complex problem, which in general has several points in common with function approximation problems. The analysis of time series is based on the assumption that successive values in the data file represent consecutive measurements taken at equally spaced time intervals. There are two main goals of time series analysis: (a) identifying the nature of the phenomenon represented by the sequence of observations, and (b) forecasting (predicting future values of the time series variable). Those goals are normally joined by approaching the problem as a modelling I/O data task, which on the one hand, builds up a model that tries to identify the internal structure of the observed data, in order to, on the other hand, predict novel values of the series using that predictive model. Long term predictions require a more careful study, since long term data dependencies are more difficult to model. Two approaches can be taken to tackle this problem: direct prediction and recursive prediction. Several methodologies have been applied to the problem of time series prediction. Our research has been centred in the improvements of long term time series prediction using different strategies and techniques. In [13] a preliminary work using SVM and LS-SVM was performed to predict the product quality in a glass melting process (Eunite competition 2003). A modified strategy was used to perform the prediction, and the results exceeded those obtained in the official competition. A novel technique to improve recursive prediction was presented in [6], which showed up to improve the long term prediction using recursive models. The technique was checked using two completely different approaches as are the LS-SVM model and TSK models, and the results obtained confirmed its effectiveness. LS-SVMs showed to bring an excellent effectiveness while they are not interpretable. TSK models showed to bring a good performance, providing an interpretable set of rules.

Further works in [15][16] have been performed for long term time series prediction using LS-SVM and other traditional statistics techniques, showing the importance of including the seasonality information of the time series in the model. These works have been presented in the first ESTSP conference to also participate in a new benchmark competition. With respect to the problem of variable selection in modelling problems, a preliminary work has been presented in [17], that makes use of the concept of Mutual Information. It is a successful adaptation of a previous work for classification problems. Furthermore, time series prediction using the MGFS model and including variable selection, were studied in [10][11][12].

c) Radial Basis Function Neural Networks: Clustering and optimization algorithms in parallel platforms

The research was oriented to the development of new clustering algorithms that perform a more adequate initialization of the centers, concretely for function approximation problems. The research started from the basis established by the CFA algorithm [18]. This algorithm presented a new clustering technique oriented to design RBFNN using the concept of expected output of a center. This output represents the position of a center in the output axis. The CFA algorithm creates several partitions of the input vectors, one for each center, where an input vector only belongs to one partition. This kind of partition, know as a hard partition, was presented in the Hard C-means algorithm and later, was improved by the Fuzzy C-means algorithm which allows the sharing of the input vectors between the centers of the clusters. Several aspects of the CFA were improved, the first one was the use of a fuzzy partition instead of a hard partition [21][22], later, other aspects such as the convergence and efficiency were revised in [23]. Once the algorithm was refined, several applications were presented where the new developed algorithm obtained a satisfactory performance [24][25]. A further study on how the input vectors were distributed among the cluster was carried out considering other kind of partitions such as possibilistic and hybrid approaches, the results showed that if a mixed approach considering a fuzzy partition and a possibilistic partition is used in the initialization of the centers, the results can be improved [26][27][28].

One of the main drawbacks of genetic algorithms is their excessive computational time. The research then focused in optimizing the previously developed algorithm so the computational effort was decreased. Due to the intrinsic parallelism of the genetic algorithms, the algorithm was designed to be executed in parallel platforms. Parallel computing is a discipline that can be the solution for many problems that require a huge computational effort. The algorithm was divided into smaller algorithms executed in different processors, and periodically, a communication step allowed the isolated populations on each processor communicate with the others, this parallel paradigm is know as the island model[29][30]. One new aspect introduced during the research was the specialization of the different islands in the evolution of different parameters of the RBFNNs, this approach increases the exploration and exploitation capabilities of the algorithm, designing better RBFNNs as was demonstrated in [31][32]

Many aspects of the parallel algorithm could be studied in future research, as example, efficiency and speed-up must be calculated in order to determine the scalability of the algorithm, this aspect is not trivial since the algorithm is subdivided into 4 algorithms. Depending on the previous results, a load balancing algorithm could be applied in case one of the subalgorithms could require more computation, even the granularity level of the parallel design could be incremented to be able to make a proportionate load distribution. Besides the parallel algorithm, the design of the RBFNNs still has some aspects that should be studied: how to select the input variables and if a distribution of these variables among the RBFs could provide better results than feeding all the RBFs with the same input vector.

d) Dynamic Multi-objective Optimization Problems

The benefits that can be obtained from parallel processing of dynamic multi-objective optimization problems could the same that in static multi-objective optimization (lower processing times and bigger efficiency and diversity in the population), but also the possibility of improving the algorithm reaction to changes. This reduces the processing time required, and allows a set of non-dominated solutions near to the Pareto front to be reached earlier. Thus, dynamic optimization problems, where the change rate is faster, could also be tackled.

The results of our approach to the possibilities of parallel processing for dynamic multi-objective optimization are quite remarkable. First of all, we have developed a parallel procedure to reach solution sets quite near to the changing Pareto fronts in multi-objective problems [42][43]. This procedure is an adaptation to dynamic environments of the PSFGA algorithm for multi-objective optimization [46][47]. It uses a master process to distribute the population of solutions among the processors that evolve their corresponding subpopulations for genpar iterations. Then, the master collects the (partial) Pareto fronts independently determined by the worker processors, builds a whole Pareto front from the partial ones, executes genser iterations of the evolutionary algorithm, and distributes the obtained population of solutions again. In the master, there is a crowding mechanism for keeping the diversity and the distribution of the solutions on the Pareto front founded. So, after reaching a number of solutions, equal or above to a given percent of the population size, only the non-dominated solutions that are far enough of the other ones are chosen.

Thus, our procedure allows a continuous transition between a master-worker operation model, when genpar is set to 0 (the workers only compute the fitness of their subpopulations) and an island model (genpar>0) where the processors communicate through the master. The speedup results obtained by our procedure allow a reduction in the convergence times, and hence, the ability to satisfy stronger time restrictions in the dynamic problem. We consider that the super-linear speedups that have been observed in some cases show the usefulness of parallel processing in keeping up the diversity of the population, in the improvement of the reaction capability and in the algorithm adaptability. It is clear that there are many things to do yet. On one side, we think that many algorithm characteristics and parameters should be analyzed and optimized, both in the sequential and parallel versions of the algorithm. Thus, we plan to study the scalability and performance behaviour for different versions of the algorithm in which the worker and master processes run asynchronously, with different communication schemes (including the ability of direct communications between workers), and genser/genpar rates. We also plan to consider other more flexible schemes where, for example, more than one process acts as a master at a given time. Furthermore, it is also necessary to evaluate the performance of the procedure with a broader set of benchmarks and some real world applications. Up to now, we have considered the use of evolutionary computation in dynamic scheduling problems [45].

2.2 Objetive 2.- Implementation, analysis and design of efficient parallel platforms for the simulation of intelligent systems.

a) Efficient network interfaces

After considering the advantages of drawbacks of different alternatives for developing intelligent Network Interface Cards (NICs), we consider that the NICs based on network processors are the most adequate ones. This choice has been motivated by their programmability, the availability of libraries and programming and debugging tools, and by the kind of parallelism that network processor microarchitectures usually implement.

Network processors (NP) are programmable circuits that provide fast and flexible resources for high-speed communication functions processing. They are composed of multiple cores that are usually organized either into a parallel model in which each packet goes to a different NP core that processes it, or a pipeline model where each packet is pipelined through a subset of NP cores that together implement the code for packet processing. The use of these network processors has been driven by the increasing demand for high throughput and the flexibility required for packet

processing. As higher bandwidths are available simple Network Interfaces Cards (NICs) should be substituted by Intelligent NICs (INICs). Nevertheless, instead of using application specific processors in these INICs, the need for adaptability, differentiation, and short time to market brought the so called network processors (NP), application-specific instruction set processors that can be programmed quickly and easily to get the required network application. Among the different alternatives we have chosen the Intel IXP network processors. The Intel NP of the IXP2XXXseries implement highly parallel microarchitectures that include several programmable processors: one Intel XScale general-purpose processor (RISC processor compliant with ARM architecture) and 8 or 16 coprocessors optimized for packet processing, called microengines. This parallelism allows processing of multiple packets simultaneously, which can greatly increase the throughput of the processor. Each microengine has eight threads of execution with zero-overhead context switch (the microengine has registers and program counters for each thread in order to reach a quickly switch from executing one thread to another thread in the machine). This feature is very important to hide latencies and it is necessary to take advantage of it in order to achieve high performance. Moreover, the IXP2XXX processors offer flexibility through upgradeability, programmability, libraries and developing tools.

Thus, our research work is dealt with the programming and efficient use of the multi-core microarchitectures that network processors implement. Nowadays, as the technology trends make the multi-core microprocessor as the microarchitecture of choice, this research line is very relevant. We have been working on offloaded TCP/UDP protocols. Thus, we have some TCP and UDP implementations on a NIC based on the IXP2400 network processor (using the Radisys ENP2611 board) [37]. They allow us to demonstrate the correct operation of the offloaded protocols. Although we have reached a moderate latency reduction, the bandwidth improvements achieved are not significant. The reasons are that these first implementations do not use multiple microengines and their poor optimized use of multithreading. Model LAWS and the Moore's law can aid us to understand this situation, as we are offloading part of the communication work to a processor (the network processor) with lower clock frequency that the CPU and we are not using the parallelism provided by the offloading engine.

Nevertheless, some recent works provide experimental results to argue that protocol offloading, in particular TCP offloading, does not clearly benefit the communication performance of the applications. On the one hand, the reasons for this scepticism are the difficulties in the implementation, debugging, quality assurance and management of the offloaded protocols. The communication between the NIC (with the offloaded protocol) and the CPU and the API could be as complex as the protocol to be offloaded. Protocol offloading requires the coordination between the NIC and the OS for a correct management of resources such as the buffers, the port numbers, etc. In case of protocols such as TCP, the control of the buffers is complicated and could hamper the offloading benefits (for example, the TCP buffers must be held until acknowledged or pending reassembly). Moreover, the inefficiency of short TCP connections is due to the overhead of processing the events that are visible to the application and cannot be avoided by protocol offloading. Probably, these are not definitive arguments with respect to the offloading usefulness but they counterbalance the possible benefits. On the other hand, there are fundamental reasons that affect the possible offloading advantages. One of them is the ratio of host CPU speed to NIC processing speed. The CPU speed is usually higher than the processors in the NIC and, moreover, the increment in the CPU speeds according the Moore's law tends to maintain or even to increase this ratio in the case of the specific purpose processors in the NIC. Thus, the part of the protocol that is offloaded would require more execution time in the NIC than in the CPU, and the NIC could appear as the communication bottleneck. The use of general-purpose processors in the NIC

(with speeds similar to the CPU) could represent a bad compromise between performance and cost. Moreover, the limitations in the resources (memory) available in the NIC could imply restrictions in the system scalability (for example, limitations in the size of the IP routing table).

Some papers have recently appeared to understand the fundamental principles under the experimental results through models such as LAWS and EMO. Both models provide a good starting point to get accurate models and descriptions about the conditions where offloading could be advantageous. Nevertheless, they should be extended to include wider sets of application domains and benchmarks and they also require a more detailed experimental validation with these applications and benchmarks. We have been working in these problems by using simulation [38][39][40]. In these papers we propose the use of Simics to analyze protocol offloading. Although Simics presents some limitations and it is possible to use other simulators for our purposes, the resources provided by Simics for device modelling and the debugging facilities, make Simics an appropriate tool. Moreover, it has allowed us a relative fast simulation of our models.

The simulation results obtained shows the improvement provided by offloading heavy protocols like TCP, not only in the ideal case, in which we use ideal buses memory, but also in more realistic situations, in which memory latencies and non-ideal buses are modelled. Thanks to the Simics model, it is possible to analyse the important parameters and the conditions in which offloading determines greater improvements in the overall communication performance.

The results obtained in the experiments we have done show throughput improvements in all the cases considered by using a host processor and NIC processor with similar speeds. Moreover it is shown that offloading releases the 40% of the system CPU cycles in applications with intensive processor utilization On the other side, we also present results that show how the technology of the processor included in the NIC affects to the overall communication performance. The behaviour we have observed in our experiments coincides with the analyses and conclusions provided in other previous works. This situation constitutes an evidence of the correctness of our Simic model for protocol offloading. One of our papers,[39], can be downloaded from the web site of Virtutech, the Simics manufacturer [41].

2.3 Objetive 3.- Application and dissemination of the proposed intelligent methodologies and architectures in real problems: modelling, classification, time series prediction and real time control.

a) Intelligent control: Application to the Control of the Temperature of a Dynamic Room in Real Time, application to Water distribution networks scheduling in Granada and Mineral extraction factory in Cuba .

Substantial developments in optimizing control methods for different purposes have been made in the field of intelligent control in the bibliography. However, most of them are based on a known system model, whereas in practice such models are not usually available due to the complexity of the plant to be controlled. In this paper, we present a novel approach to achieve global adaptation in fuzzy controllers. The algorithm does not need a mathematical model of the plant or its approximation by means of a Jacobian matrix. Qualitative knowledge and auxiliary fuzzy controllers help the main controller to accomplish its task in real time [59][60][61]. One application was the Control of the Temperature of a Dynamic Room in Real Time [48]. Another important problem is the water supply system of Granada (Spain). We have developed a general procedure that can be applied to problems related to the scheduling of distribution networks. This procedure offers a very simple way of making good use of the computation capability of a cluster of computers to explore feasible control trajectories and avoids a complex modelling of the system

thanks to the incorporation of a prediction method we have developed, and which is use to forecast an approximate future consumption level. In the next figure we describe the system upon which the developed scheduling procedure has been applied. Briefly, the problem is to obtain the optimum water flow distribution given by the water treatment plant (E.T.A.P.), which can be seen inside the box, for the water tanks marked. The implemented system allows us to have a centralized and real time control over the Water Integral Cycle in the city of Granada. The central control room is located in the headquarters of the company EMASAGRA, from where we can establish our communications via radio (SPIRAL system by ICR developed by researched of our group) with several remote stations controlled by PLCs (Siemens Simatic S5) and other stations based on the distributed tele-control system SPIRAL-5000. In some places of the network close to the radio-linked remote stations, it has been utilized a field bus (Siemens Sinec L1/L2). The application software was developed from the software package SCADA INTOUCH.

Another research line has been carried out, to present a system that can be used to control and optimize the mineral extraction from source materials [62]. One of the modules that builds the system is in charge of predicting the final amount of extracted mineral from empirical data obtained previously. The module consists in an RBFNN, that is able to predict quite precisely the real output of material, and in a genetic algorithm that trains the network within the time frame required by the system. A multiobjective genetic algorithm that designs the RBFNNs for the prediction module was presented, obtaining a very good performance when it was compared against other techniques for the design of RBFNNs.

b) Biomedical application: classification of several cancer pathologies, digital mamographies and electro cardiogram signals

In this project we have considered the viability of a skin lesion diagnosis scheme based on fluorescence images. Three kinds of skin lesions are considered: actinitic keratosis, basal cell carcinoma and psoriasis. A wide and diverse set of simple parameters have been extracted and their discrimination potential is evaluated through an automatic diagnosis scheme based on the intelligent algorithm developed in previous section. We use a sequential scanning technique that automatically selects the most relevant parameters for the addressed problem. [19][20][71]

The new intelligent neural model presented in this project, has also been applied in electro cardiogram signals. The objective of the approximation is to be able to classify a heart beat in order to determine if it suffers pathology [19][24][47].

Another important application is image processing [66][68], specifically digital mammography. Digital mammography has been introduced in several countries over the last few years. The new technology requires new optimising methods considering for instance the increased possibility of changing the absorbed dose and the effect it has on the image processing and noise level. There is no agreement on the noise level that may be accepted and the absorbed doses needed. We have proposed a newly algorithms based on an entropic measure in recognizing cancer regions on mammograms, which is very robust against the presence of noise, low contrast or blurring in the image. This property makes it very suitable in digital mammograms because previous filter is not needed. Experiments show the performance of the proposed procedure using several digital mammography's. It can be shown that the algorithm is robust, effective and presents better results than other method presented in the literature [58].

Finally, it also important to note the application to automatic classification of prostate cancer using intelligent system [56][57]. We have proposed an automatic procedure for prostate cancer light micrograph based on soft-computing technique, for image interpretation, with increased accuracy.

We have proposed a feature subset selection algorithm that selects the most important features, used by a pseudo-gaussian radial basis function neural networks to classify the prostate cancer light micrograph. A high classification rate has been achieved which will reduce the subjective human invention and will increase the diagnostic speed.

3 Indices to measure the quality of the obtained results

The main indicators that show the results of the research carried out in the project TIN2004-1419 are classified in the following three categories: formation of students, publication and national and international collaboration.

3.1 Formation of students

Regarding the staff under formation, it is important to note that the project team members are advising the eleven different PhD dissertations in the project topics. Six doctoral thesis have been defended and five PhD students have presented their research projects as a previous step to present during this an next year their doctoral thesis. In particular :

1.- D. Mohammed Awad, PhD student in the Department of Computer Architecture and Technology (Granada), has finished his doctoral thesis titled: "Model complex system by hierarchical structured based on intelligent system" with Drs. Ignacio Rojas and Drs. Héctor Pomares as advisors. He presented it by June 2005

2.- D. Mouncef Filali Bouami, PhD student in the Department of Computer Architecture and Technology (Granada), has finished his doctoral thesis titled: "Development and Optimization of New Radial Basis Function Neural Model", with Drs. Ignacio Rojas and Drs. Carlos García-Puntonet as advisors. He presented it by July 2005.

3.- D. Carlos Castro Serrato, PhD student in the Department of Computer Architecture and Technology (Granada) in collaboration with SIEMENS (Munich, Germany), has finished his doctoral thesis titled: "Development of a high accuracy analogue-to-digital converter system: application in data logging units", with Drs. Ignacio Rojas, Drs. Alberto Prieto and Dr- Ing and Doctor Honoris Karl Goser as advisors. He presented it by May 2006 (European doctor)

4.- Dª. Manuela Alba Bueno, PhD student in the Department of Computer Architecture and Technology (Granada) in collaboration with SIEMENS (Munich, Germany), has finished her doctoral thesis titled: "Study and Characterization With Scanning Probe Methods of Employable Nanomaterials in new Architectures for Molecular Memories" with Drs. Ignacio Rojas, Drs. Alberto Prieto and Dr- Ing and Doctor Honoris Karl Goser as advisors. She presented it by May 2006 (European doctor)

5.- D. Suhail Musa Issa Odeh, PhD student in the Department of Computer Architecture and Technology (Granada), has finished his doctoral thesis titled: "Development of intelligent system for diagnosis and classification in medicine", with Drs. Eduardo Ros and Drs. Ignacio Rojas as advisors. He presented it by July 2006.

6.- D. Javier Díaz Alonso, PhD student in the Department of Computer Architecture and Technology (Granada), has finished his doctoral thesis titled: "Bio-inspired vision system. Architecture for movement processing and stereo-vision for high performance" with Drs. Alberto Prieto Espinosa and Drs. Eduardo Ros as advisors. He presented it by July 2006.

6.- D. Christián Agustín Morillas Gutiérrez, PhD student in the Department of Computer Architecture and Technology (Granada), has finished his doctoral thesis titled: "Model and Intelligent tools for visual rehabilitation" with Drs. Alberto Prieto Espinosa and Drs. Francisco José Pelayo Valle as advisors. He presented it by September 2006.

7.- D.Francisco Fernandez, PhD student in the Department of Computer Architecture and Technology (Granada), has presented his research project titled: "New intelligent system using Multiobjective RBFNNs Designer: Application for Mineral Reduction", with Drs. Julio Ortega Lopera and Drs. Ignacio Rojas as advisors. He has presented a first version of his doctoral thesis, and he will present it during next months.

8.- D. Alberto Guillen, PhD student in the Department of Computer Architecture and Technology (Granada), and actually assistant professor at the University of Jaen. He has presented in 2005 his research project titled: "Intelligent computation in parallel platforms: application to functional approximation and classification", with Drs. Ignacio Rojas, Drs. Héctor Pomares and Drs. Jesús González as advisors. He has presented a first version of his doctoral thesis, and he will present it during next months.

9.- D. Luis Javier Herrera, PhD student in the Department of Computer Architecture and Technology (Granada), and actually assistant professor at the University of Granada. He has presented in 2005 his research project titled: "Adaptive and intelligent systems for function approximation and prediction using high performance computation" with Drs. Ignacio Rojas and Drs. Héctor Pomares as advisors. He has presented a first version of his doctoral thesis, and he will present it during next months.

10.- D. Gines Rubio, PhD student in the Department of Computer Architecture and Technology (Granada). He has presented in 2005 his research project titled: "Implementation and application of kernel methods on high performance computation" with Drs. Ignacio Rojas and Drs. Héctor Pomares as advisors. He will present his doctoral thesis the next year.

11.- D. Pablo Cascón, PhD student in the Department of Computer Architecture and Technology (Granada). He has presented in 2006 his research project titled: "High performance computation and communication in clusters " with Drs. Julio Ortega as advisors. He will present his doctoral thesis the next year.

3.2 Publication

In summary, a total number of 93 publications have been developed during this project, distributed mainly in international journal and international conferences (78), and 15 in different national conference. All these publications will be available in the web page of the Research Group, to which the project team members belong (http://atc.ugr.es).

3.3 National and international collaboration

3.3.1 Joint activities organized:

The following events have been co-organized:

- The 8th International Work-Conference on Artificial Neural Networks (IWANN'2005) (Computational Intelligence and Bioinspired Systems), in Vilanova i la Geltrú (Barcelona, Spain) June 8-10, 2005.
- First International Workshop on Genetic Fuzzy Systems, March 17-19, Granada, Spain, in cooperation with: Genetic Fuzzy Systems Task Force, Fuzzy Systems Technical Committee, IEEE Computational Intelligence Society
- I Congreso Español de Informática (CEDI'2005), Granada, September 13-16, 2005
- I Simposio de Inteligencia Computacional, SICO'2007 (IEEE Computational Intelligence Society, SC), Granada, September 13-16, 2005

To be organized during this year:

- II Congreso Español de Informática (CEDI'2007), Zaragoza, September 11-14, 2007
- II Simposio de Inteligencia Computacional, SICO'2007 (IEEE Computational Intelligence Society, SC), Zaragoza, September 11-14, 2007
- 9th International Work-Conference on Artificial Neural Networks (IWANN'2007) (Computational and Ambient Intelligence) will take place in San Sebastián (Spain) June 20-22, 2007,

It is also important to note the organization of different invited sessions in international an national conferences (ESANN, GFS, SICO, CEDI, etc.), and the edition of two special issues in international journal (*Neurocomputing:* Alberto Prieto, Joan Cabestany and Francisco Sandoval (Eds.) Selected paper from IWANN 05, and IEEE Transaction on Fuzzy Systems, Special Issue: "Genetic Fuzzy Systems: What's Next?", O.Cordón, I.Rojas, R.Alcalá, (Eds.)).

3.3.2 National and International projects with different enterprices

Member of our group have participated with other Spanish and International research groups and enterprises in the following projects:

- Title: Module consumptions of the Cooling Area Network Systems (CANS). Reference: OTRI 2006-2324. Entity: CIATESA (CIATESA is a company dedicated to the design, production and commercialization of air conditioning teams, refrigeration, treatment of air and thermal exchanges.). Keywords: Intelligent system, Telecontrol, Web technology.

- Title: Module tariffication of the Cooling Area Network Systems (CANS).Reference: OTRI 2006-2325.Entity: CIATESA (CIATESA is a company dedicated to the design, production and commercialization of air conditioning teams, refrigeration, treatment of air and thermal exchanges).Keywords: Intelligent system, Telecontrol, Web technology.

- Title: Module Fuzzy of the Cooling Area Network Systems (CANS).Reference: OTRI 2006_2326. Entity: CIATESA (CIATESA is a company dedicated to the design, production and commercialization of air conditioning teams, refrigeration, treatment of air and thermal exchanges).Keywords: Intelligent system, Telecontrol, Web technology.

- Title: Advanced Tele-asistance, domotic and multimedia (TELEADM: Teleasistencia avanzada domótica y multimedia).Reference: CTA 06/0053. Entity: Coorporación Tecnologica de Andalucia (CTA). Consorcio de empresas Telefónica I+D y Sadiel.

- Title: eQuirófano: Multimedia Collaborative Envirotment (eQECM).Reference: CTA06. Entity: Coorporación Tecnológica de Andalcía (CTA), and Consergería de Innovación, Ciencia y Empresa.

- Title: Advanced teleasistance platform (PLATELA). Reference: FIT-350301-2006-10 TRACTOR-PRO.Entity: Ministerio de Industría, Turismo y Comercio, y consorcio de empresas. Programa Nacional de Tecnologías de Servicios de la Sociedad de la Información. Acción estratégica e-inclusión y e-asistencia. MITyC. Note: Consortium formed for: Telefónica I+D, Instituto Tecnológico de Aragón, Gotor Comunicaciones, MAPFRE Quavitae, Bioingeniería Aragonesa, y Universidad de Granada (Grupo CASIP - ATC). Date: 01/11/2006 to 31/12/2007

- Title: Development and Integration of new Technologies for supervision and control on industrial environment". Reference: F2284-01. Entity: Fundación Empresa Universidad de Granada and ICR (01-10-2004 to 31-09-2005). Keywords: Comunication networks, Distribuited control

- Title: Escalability and high computation in clusters of computers. Reference: P06-TIC-01935. Entity: Proyecto de Investigación de Excelencia de la Junta de Andalucía. Date: 01/01/2007 to 31/12/ 2009

- Title: Technological platforms for a remote unit terminal (UTR). Reference: FIT-330100-2006-60 PROFIT. Entity: Ministerio de Industría, Turismo y Comercio, y consorcio de empresas. Note: Consortium formed for: Telvent Energía y Medio Ambiente S.A., Asoc. de Investigación y Coop. Industrial de Andalucia, Universidad de Granada (CASIP) y CSIC (Centro Nacional de Microelectrónica. Date: 15/09/2006 to 31/12/2007

- Title: Intelligent and distributed control in agricultural hothouses. Reference: To be confirmed by Junta de Andalucia. Entity: Junta de Andalucia and Eurocastell Sociedad Agraria de Transformación

European Project:

- Title: Learning to emulate perception action cycles in a driving school scenario (DRIVSCO). Reference: IST-016276-2 (VI Programa Marc). Date: 01/02/2006 to

31/07/2009 . Entity: European Union. Participants: Universidad de Granada (España), University of Genoa (Italia), Katholieke Universiteit Leuven (Bélgica), University College London (Reino Unido), University of Göttingen (Alemania), Münster University (Alemania), Aalborg University (Dinamarca), Vytautas Magnus University (Lituania) y Hella KG Hueck & Co (Alemania).

In collaboration with Telefónica Investigación y Desarrollo, member of our reserach group also have presente dan CENIT Project, entitle: "Personal Digital Enviroment for the health and well-being", Reference: AmIVital. If is also important to note the collaboration with the European Commission's Research Infrastructuresactivity of the Structuring European Research Area programme, contract number RII3-CT-2003-506079 (HPC-Europa).

4 References

[1] Herrera, L.J., Pomares, H., Rojas, I., Valenzuela, O., Prieto, A., "TaSe, a Taylor Series Based Fuzzy System Model that Combines Interpretability and Accuracy" Fuzzy Sets and Systems, 153 (2005) 403–427

[2] Herrera, L.J., Pomares, H., Rojas, I., González, J., Valenzuela, O., "Function Approximation through Fuzzy Systems Using Taylor Series Expansion-Based Rules, Interpretability and Parameter Tuning". Lecture Notes In Computer Science Vol. 2972, ISBN 3-540-21459-3, pp.508-516, 2004.

[3] Luis Javier Herrera, Héctor Pomares, Ignacio Rojas, Alberto Guillén, Mohammed Awad, "Analysis of the TaSe-II TSK-type Fuzzy System for function approximation", Lecture Notes in Computer Science, ECSQARU'2005, pp 613-624

[4] L.J. Herrera, H. Pomares, I. Rojas, A. Guillén, J. González, Análisis estadístico del uso de reglas Takagi-Sugeno-Kang de orden alto en el diseño del proceso de inferencia difusa, Actas del XII Congreso Español sobre Tecnologías y Lógica Fuzzy, ESTYLF'2004, ISBN 84-609-2160-3, Jaén Septiembre 2004, pp. 363-368

[5] Luis Javier Herrera, Héctor Pomares, Ignacio Rojas, and Alberto Guillén, "TaSe model for Long Term Time Series Forecasting", Lecture Notes in Computer Science, IWANN'2005, pp 1027-1034

[6] L. J. Herrera, H. Pomares, I. Rojas, A. Guillén, A. Prieto, O. Valenzuela: "Recursive Prediction for Long Term Time Series Forecasting Using Advanced Models", Neurocomputing, 2006, In Press

[7] Luis Javier Herrera, Héctor Pomares, Ignacio Rojas, Alberto Guillén, Jesús González, and Mohammed Awad, "Clustering-Based TSK Neuro-Fuzzy Model for Function Approximation with Interpretable Sub-models", Lecture Notes in Computer Science, IWANN'2005, pp. 399-406

[8] L. J. Herrera, H. Pomares, I. Rojas, A. Guillén, M. Awad, and J. González, "Interpretable Rule Extraction and Function Approximation from numerical Input/Output Data Using the modified Fuzzy TSK model: TaSe model", Lecture Notes in Computer Science, RSFDGrC'2005,

[9] Herrera, L.J., Pomares, H., Rojas, I., Valenzuela, O., Awad, M., "MultiGrid-Based Fuzzy Systems for Function Approximation". Lecture Notes In Computer Science Vol. 2972, ISBN 3-540-21459-3, pp.252-261, April 2004

[10] L.J.Herrera, H.Pomares, I.Rojas, O.Valenzuela, J.Gonzalez, M.Awad, Multigrid-Based Fuzzy Systems for Time Series Forecasting: Overcoming the curse of dimensionality, European Symposium on Artificial Neural Networks. Proceedings of the 12th. European Symposium on Artificial Neural Networks (ESANN'2004), D-Facto public. ISBN 2-930307-04-8, pp. 127-132. Brujas (Bélgica), April 28-30, 2004

[11] L. J. Herrera, Hector Pomares, Ignacio Rojas, Jesús Gonzalez and Mohammed Awad, MultiGrid-Based Fuzzy Systems for Time Series, IJCNN'04, Forecasting: CATS Benchmark IJCNN Competition (special session Sf) , Budapest July, 2004. paginas, ISBN 0-7803-8360-5

[12] L. J. Herrera, H. Pomares, I. Rojas, A. Guillén, J. González, M. Awad, A. Herrera, "MultiGrid-Based Fuzzy Systems for Time Series Prediction: CATS Competition", Neurocomputing, In Press, 2007

[13] Tobias Jung and Luis J. Herrera, "Long Term Prediction of Product Quality in a Glass Manufacturing Process Using a Kernel Based Approach", Lecture Notes in Computer Science, IWANN'2005, pp. 960-967

[14] L. Herrera Maldonado, H. Pomares, I. Rojas, A. Guillén-Perales, "New Clustering Technique for Inicialization of Centres in TSK Clustering-Based Fuzzy Systems", Lecture Notes in Computer Science, ECSQARU'2005, pp, 980-991

[15] L. J. Herrera, H. Pomares, I. Rojas, A.Guillén, G. Rubio, "Incorporating Seasonal Information on Direct and Recursive Predictors Using LS-SVM", ESTSP'2007, Accepted

[16] L. J. Herrera, H. Pomares, I. Rojas, G. Rubio, A.Guillén, "Removing seasonality on time series, a practical case", ESTSP2007, Accepted

[17] L.J. Herrera, H. Pomares, I. Rojas, M. Verleysen, and A. Guillén, "Effective Input Variable Selection for Function Approximation", Lecture Notes in Computer Science, ICANN'2006, pp 41-50

[18] J. Gonzalez, I. Rojas, H. Pomares, J. Ortega, and A. Prieto. *"A new Clustering Technique for Function Aproximation"*. IEEE Transactions on Neural Networks, 13(1):132-142,

[19] E.Ros, S.Mota, F.J.Fernandez, F.J. Toro, J.L.Benier, "ECG Characterization of Paroxysmal Atrial Fibrillation: Parameter Extraction and Automatic Diagnosis Algorithm", Computer in Biology and Medicine, vol.34, no.8, pp.679-696, 2004.

[20] S.M. Odeh, E. Ros, I. Rojas, "Título: Exploración de características para el diagnóstico de lesiones de piel basado en imágenes de la fluorescencia" I Congreso Español de Informática (CEDI2005). SICO2005. pp. 431-438, Thomson, ISBN: 84-9732-444-7. Septiembre 2005

[21] A.Guillén, I.Rojas, J.González, H. Pomares, L.J. Herrera Maldonado, *Clustering Based Algorithm for RBF Centres Initialization*. Fuzzy Economic Review, Vol X(2): 27-45, November 2005

[22] A.Guillén, I. Rojas, J. González, H. Pomares, L. J. Herrera and A. Prieto, *Using a New Clustering Algorithm to Design RBF Networks for Functional Approximation Problems*. Proceedings of the Learning'04 International Conference, ISBN 84-688-8453-7, Elche, Spain, 2004, pp. 19-24

[23] A. Guillén, I. Rojas, J. González, H. Pomares and L. J. Herrera, O. Valenzuela and A. Prieto, "*Improving Clustering Technique for Functional Approximation Problem using Fuzzy Logic: ICFA algorithm*", Lecture Notes in Computer Science, IWANN'2005, pp. 272-279

[24] A. Guillén, I. Rojas, E. Ros, L. J. Herrera: "*Using Fuzzy Clustering Technique for Function Approximation to Approximate ECG Signals*". Lecture Notes in Computer Science (2) 2005: 538-547

[25] A. Guillén, I. Rojas, J. González, H. Pomares and L. J. Herrera, O. Valenzuela and A. Prieto "*Using Fuzzy Logic to Improve a Clustering Technique for Function Approximation"* Neurocomputing (NEUCOM-D-05-00326R1)

[26] A. Guillén, I. Rojas, J. González, H. Pomares, L.J. Herrera, O. Valenzuela, and A. Prieto, "*A Possibilitic Approach to RBFN Centers Initialization*", RSFDGrC 2005,LNAI 3642,pp.174 –183,2005

[27] A. Guillén, I. Rojas, J. González, H. Pomares, L.J. Herrera, and A. Prieto: *"A Fuzzy-Possibilistic Fuzzy Ruled Clustering Algorithm for RBFNNs Design"* RSCTC 2006, LNAI 4259, pp.n 647-656

[28] A. Guillén H. Pomares I. Rojas J. González L.J. Herrera F. Rojas *"Studying Possibility in a Clustering Algorithm for RBFNN Design for Function Approximation"*. Neural Computing & Applications journal, 2006

[29] J. Gonzalez, I. Rojas, J. Ortega, H. Pomares, F.J. Fernandez, and A. Diaz. "*Multi-objective evolutionary optimization of the size, shape, and position parameters of radial basis function networks for function approximation"*. IEEE Transactions on Neural Networks, 14(6):1478{1495, November 2004

[30] A. Guillén, I. Rojas, J. González, H. Pomares, L.J. Herrera and Francisco Fernández: "*Multiobjective RBFNNs Designer for Function Approximation: An Application for Mineral Reduction"* Lecture Notes in Computer Science, 2006

[31] A. Guillén, I. Rojas, J. González, H. Pomares, L. J. Herrera, B. Paechter *"Improving the Performance of Multi-objective Genetic Algorithm for Function Approximation through Parallel Islands Specialisation"* Lecture Notes in Computer Science, 2006

[32] A. Guillén, I. Rojas, J. González, H. Pomares, L. J. Herrera, B. Paechter *"Boosting the performance of a multiobjective algorithm to design RBFNNs through parallelization"* Lecture Notes in Computer Science, 2006

[33] G. Rubio, H. Pomares. *"A Basic Approach to Reduce the Complexity of a Self-generated Fuzzy Rule-Table for Function Approximation by Use of Symbolic Interpolation"* Lecture Notes on Computer Science (LNCS) Vol. 3512 pag. 34-41, 2005

[34] G. Rubio, H. Pomares, I. Rojas, A. Guillén. *"A Basic Approach to Reduce the Complexity of a Self-generated Fuzzy Rule-Table for Function Approximation by Use of Symbolic Regression in 1D and 2D Cases"* Lecture Notes on Computer Science Vol. 3562 pag. 143-152, 2005.

[35] G. Rubio, H. Pomares, I. Rojas. *"Aplicación de la programación genética a la obtención de modelos analíticos no lineales para predicción de series temporales"* Actas del Simposio de Inteligencia Computacional, SICO2005 (IEEE Computacional Inteligence society, SC), pp. 3-10, Thomson, ISBN84-9732-444-7, I Congreso Español de Informática (CEDI2005) Granada, 13-16 Septiembre 2005

[36] Ginés Rubio, Luis J. Herrera, Héctor Pomares. *"Gaussian Process regression applied to time series prediction: a particular case study in ESTSP 2007 time series prediction competition"*. Neurocomputing Journal. Elsevier. *Accepted but not yet published.*

[37] Cascón, P.; Ortega, J., Díaz, A.F.; Rojas, I.:"Externalización de protocolos en procesadores de red". XVII Jornadas de Paralelismo, pp.205-210, Albacete. Septiembre, 2006.

[38] Ortíz, A.; Ortega, J.; Díaz, A.F.; Prieto, A.:"Simulación de la externalización de protocolos de comunicación

mediante Simics". XVII Jornadas de Paralelismo, pp.103-108, Albacete. Septiembre, 2006.
[39] Ortíz, A.; Ortega, J.; Díaz, A. F.; Prieto, A.:"Protocol offload evaluation using Simics". IEEE Cluster Computing, Barcelona. Septiembre, 2006.
[40] Ortíz, A.; Ortega, J.; Díaz, A.F.; Prieto, A.:" Analyzing the benefits of protocol offload by full-system simulation". 15th Euromicro Conference on Parallel, Distributed and Network-based Processing, PDP 2007. (Aceptado).
[41] Virtutech web site www.virtutech.com. Selected publications at: http://www.virtutech.com/about/research/selected-pubs.html
[42] Cámara, M.; Toro, F.J.; Ortega, J.:"Procesamiento paralelo para optimización multiobjetivo en entornos dinámicos". XVII Jornadas de Paralelismo, pp.431-436, Albacete. Septiembre, 2006.
[43] Cámara, M.; Ortega, J.; Toro, F.J. :"Parallel processing for multi-objective optimization in dynamic environments". Int. Workshops on Nature Inspired Distributed Computing (NIDISC'07), 21th IEEE/ACM Int. Parallel and Distributed Processing (IPDPS 2007).
[44] Farina, M.; Deb, K.; Amato, P.:"Dynamic Multiobjective Optimization Problems: Test cases, Approximations, and Applications". IEEE Trans. on Evolutionary Computation, Vol.8, No.5, pp.425-342. Octubre, 2004.
[45] Pérez Gutiérrez, M. D.:"Computación evolutiva paralela en problemas de optimización dinámica. Aplicación a problemas de Planificación". Proyecto de Ingeniería Informática, E.T.S. Ingenierías Informática y de Telecomunicación, 2005.
[46] Toro, F.; Ortega, J.; Ros, E.; Mota, B.; Paechter, B.; Martín, J.M.:"PSFGA: Parallel processing and evolutionary computation for multiobjective optimization". Parallel Computing, Vol. 30, pp.721-739, 2004.
[47] Toro, F.; Ros, E.; Mota, S.; Ortega, J.: "Evolutionary Algorithms for Multiobjective and Multimodal Optimization of Diagnostic Schemes". IEEE Trans. on Biomedical Engineering, Vol.53, No.2, pp. 178-189. Febrero, 2006.
[48] I. Rojas, H. Pomares, J. Gonzalez, L.J. Herrera, A. Guillen, F. Rojas and O. Valenzuela, "Adaptive fuzzy controller: Application to the control of the temperature of a dynamic room in real time" Fuzzy Sets and Systems, 2006
[49] A.J. Rivera, I. Rojas, J. Ortega, M.J. del Jesús, "A New Hybrid Methodology for Cooperative-Coevolutionary Optimization of Radial Basis Function Networks", Soft-Computing, 2006
[50] Jesus Gonzalez, Ignacio Rojas, Hector Pomares, Luis J. Herrera, Alberto Guillen, and Jose M. Palomares, "Improving the Accuracy while preserving the Interpretability of Fuzzy Function Approximators by means of Multiobjective Evolutionary Algorithms", International Journal of Approximate Reasoning, 2006
[51] I. Rojas, F. Rojas, O. Valenzuela y A. Prieto, "The synergy of GA and Fuzzy System for multidimensional problem: Application toTime Series Prediction" Lecture Notes in Computer Science,3040/2004, pp.566-576
[52] I. Rojas, F. Rojas , H. Pomares, J. González, L. J. Herrera y O. Valenzuela, "The synergy between classical and soft-computing techniques for Time Series Prediction" Lecture Notes in Computer Science/LNAI, Vol. 2972, pp.30-39, 2005
[53] Antonio J. Rivera, Ignacio Rojas, Julio Ortega, "Application of ANOVA to a Cooperative-Coevolutionary Optimization of RBFNs" Lecture Notes in Computer Science, Volumen: 3512, pp.297-306,2005.
[54] Fernando Rojas Ruiz, Carlos G. Puntonet and Ignacio Rojas Ruiz, "Analysis of contrast functions in a genetic algorithm for post-nonlinear blind source separation" Proceedings of the 13th. European Symposium on Artificial Neural Networks,pp. 521-526, D-Facto public. ISBN 2-930307-05-6, April 28-30, 2005
[55] J.Gonzalez, I. Rojas, H.Pomares, L.J. Herrera. A. Guillén, "Multiobjective Evolutionary design of Fuzzy Systems" Proceedings of first International Workshop on Genetic Fuzzy Systems,ISBN:84-689-1117-8, pp. 13-21, 2005
[56] O. Valenzuela, I. Rojas, F. Rojas, L. Márquez, "Automatic classification of prostate cancer using pseudo-gaussian radial basis function neural network" Proceedings of the 13th. European Symposium on Artificial Neural Networks,pp. 145-150, D-Facto public. ISBN 2-930307-05-6, April 28-30, 2005
[57] O. Valenzuela, I.Rojas, A.Guillen, L.J. Herrera, H.Pomares, C. Atae-Allah, L.Marquez, "Classification of Prostate Cancer using Non-Linear Paradigms and Mutual Information", Proceedings of the Congress International MAMERN'05, pp.72-75, 2005
[58] O. Valenzuela, I.Rojas, L.Marquez, "Predicting and classification breast cancer in digital mammography using soft-computing techniques" Proceedings of the 17th IMACS World Congress. Scientific Computation, Applied Mathematics and Simulation, Paper T3-I-114-0597. ISBN: 2-915913-02-1. pp. 579-585, France, 2005.
[59] M.Teiner, I.Rojas, F.Rojas, O.Valenzuela,"Autonomous Car using a Modular Architecture" Proceedings of the 17th IMACS World Congress. Scientific Computation, Applied Mathematics and Simulation, Paper Paper T4-I-93-0600. ISBN: 2-915913-02-1. pp.600-605, France, 2005.
[60] I.Rojas, H.Pomares, J.Gonzalez, L.J.Herrera, A.Guillen, F.Rojas, "Soft-Computing techniques for the development of adaptive helicopter flight controller" Proceedings of the 9th IEEE International Workshop on

Advanced Motion Control. ISBN: 0-7803-9511-5. pp.709-710, March 27-29, 2006

[61] I.Rojas, H.Pomares, J.Gonzalez, L.J.Herrera, A.Guillen, O.Valenzuela, "Self-adaptive robot control using Fuzzy Logic" Proceedings of the 9th IEEE International Workshop on Advanced Motion Control. ISBN: 0-7803-9511-5. pp. 687-691, March 27-29, 2006

[62] Alberto Guillén, Ignacio Rojas, Jesús González, Héctor Pomares, L.J. Herrera, and Francisco Fernández "Multiobjective RBFNNs Designer for Function Approximation: An Application for Mineral Reduction" LNCS 4221, p. 511-520. 2006

[63] O.Valenzuela, I.Rojas, L.Marquez, M.Pasadas, "Feature Selection using Mutual Information and Neural Networks", Journées Zaragoza-Pau, Monografías del Seminario Matemático García de Galdeano 2006

[64] M.Anguita, J.M Martínez-Lechado; *MP3 Optimization Exploiting Processor Architecture and Using Better Algoritnhms*; IEEE Micro, Vol.25, No.3, pp. 81-92, ISSN 0272-1732, May-June 2005

[65] J. Fernández, M. Anguita, E. Ros, J.L. Bernier "SCE Toolboxes for the Development of High-Level Parallel Applications" Lecture Notes in Computer Science vol.3992 (ISBN: 3-540-34381-4), Vol.3992, pp.518-525, 2006.

[66] J. M. Palomares, J. Gonzalez and E. Ros"Designing a Fast Convolution Under the LIP Paradigm Applied to Edge Detection" Lecture Notes in Computer Science, Vol. 3687, pp.560-569, 2005:

[67] J. M. Gorriz, J. Ramirez, I. Turias, C. G. Puntonet, J. Gonzalez y E.W. Lang, "C-Means Clustering Applied to Speech Discrimination" Lecture Notes in Computer Science, Vol.3991, pp.649-656, 2006

[68] J. M. Palomares and J. Gonzalez and E. Ros and A. Prieto, "General Logarithmic Image Processing Convolution" IEEE Transactions on Image Processing, Vol. 15, pp.3602-3608, 2006.

[69] R.Baños, C. Gil, B. Paechter, J. Ortega: "Parallelization of Population-based Multi-Objective Metaheuristics: An Empirical Study". Applied Mathematical Modelling, 30/7 (2006) 578-592.

[70] R.Baños, C. Gil, B. Paechter, J. Ortega: "A Hybrid Meta-heuristic for Multi-objective optimization: MOSATS". Journal of Mathematical Modelling and Algorithms, 2006 (en prensa).

[71] C. Atae, A.Atae, I.Rojas, F.Rojas,O.Valenzuela, J.F.Gómez-Lopera "Region of Interest Detection using Jensen-Shannon Divergence for Fluorescence Image of Skin lesion" WSEAS Transactions on Mathematics, 2006, ISSN: 1109-2769.

[72] O. Valenzuela,L.Marquez,M.Posadas, I. Rojas, M.Rodriguez, F. Rojas, "Analysis of the functional block and operator involved in fuzzy system design" Lecture Notes in Computer Science,vol. 3040, pp. 157-166, 2005

[73] Mohammed Awad, Héctor Pomares, Ignacio Rojas, Luis J. Herrera, Alberto Prieto," Input Variable Selection in Hierarchical RBF Networks" Lecture Notes in Computer Science, Volumen: 3512, pp.280-288,2005.

[74] O. Valenzuela, L.Marquez, M.Pasadas,I.Rojas, F.Rojas, A.Guillen, L.J. Herrera, "Time series approximation using linear and non-linear method"Proceedings of the Congress International MAMERN'05, pp.72-75, 2005

[75] O. Valenzuela, I.Rojas, F.Rojas, L.Marquez, "Time series prediction using soft-computing techniques and Box-Jenkins Methodology" Proceedings of the 17th IMACS World Congress. Scientific Computation, Applied Mathematics and Simulation, Paper T3-I-114-0593. ISBN: 2-915913-02-1. pp. 593-598, France, 2005

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

HPSGRID: High Performance Storage for Grid Environments TIN2004-02156

Félix García Carballeira *
Universidad Carlos III de Madrid

Abstract

Data management is one of most important problem in grid environments. Most of the efforts in data management in grids have been focused in data replication. Data replication is a practical and effective method to achieve efficient data access in grids. However, all data replication schemes lack in to provide a grid file system. One important challenge facing grid computing is the design of a grid file system. The Open Grid Forum defines a Grid File System as a human-readable resource namespace for management of heterogeneous distributed data resources, that can span across multiple autonomous administrative domains. This project tries to investigate and to develop technologies that allow to use and integrate the resources and existing solutions to build a high performance, fault tolerance and secure grid file system for grids environment.
Keywords: parallel I/O, grid, replication, grid file system

1 Projet Objectives

The main goal of this project is the development of techniques for improving the performance in data access in grid environments. The specific objectives are the following:

1. Objective 1. To apply the parallelism used in parallel file systems for clusters to grid environments. With this aim, we propose the application of parallel I/O techniques to grid systems..
2. Objective 2. To develop techniques of location, resource management, caching, and load balancing.
3. Objective 3. To develop fault tolerance techniques different from the replication techniques used in grid environments. We will provide new fault tolerance models based on redundancy similar to the employed in RAIDs.
4. Objective 4. To develop a security model for the authentication, confidentiality, and data security.

*Email: felix.garcia@uc3m.es

2 Research Results

At the time of writing this report, the project has consumed 24 months out of its whole three year period. At present, a large part of the proposed goals have already been accomplished. Next, we show the level of success level reached in the former objectives:

- Objective 1. We have developed a parallel file system for grids using standard grid technologies. The last year of the project will be dedicated to implement new services in this parallel file system, as for example, a storage service using web services.
- Objective 2. The location is resolved using the grid file system implemented. Last year of the project will be dedicated to study caching and load balancing strategies.
- Objective 3. We have developed a fault tolerant model for parallel file system. This model has been implemented in the cluster version of Expand. The last year of the project will be dedicated to incorporate this model in the parallel file system for grid.
- Objective 4. Security in ensured in the parallel file system implemented because we use the grid security infrastructure (GSI) provided by the Globus Toolkit. This service provides authentitication and security in data transfers.

Next sections summarize the main research achievements reached until now.

2.1 A parallel file system for grids

Many applications in grids, require the access to large amounts of data. For this kind of applications, most of the efforts in data management in grids have been focused in data replication. Data replication is a practical and effective method to achieve efficient data access. File replicas are multiples copies of a file spread across the grid used for improving data access. However all data replication schemes provided in literature lack in to provide a global file system for accessing to files in data grids environments. One important challenge facing grid computing is a true and global file system for grid applications running in different administrative domains.

The Open Grid Forum [1] defines a Grid File System as a human-readable resource namespace for management of heterogeneous distributed data resources that can span across multiple autonomous administrative domains, that can include:

- A logical resource namespace across multiple administrative domains.
- Standard interfaces.
- A virtual namespace of WAN file system.
- Independence of physical data access/transport, authentication mechanisms.

Furthermore, a grid file system must to provide high performance data access. Parallel I/O has been the traditional technique for improving data access in clusters and multiprocessors. Parallelism in file systems is obtained by using several independent servers and striping data among these nodes to allow parallel access to files. There are many parallel file systems

for this kind of platforms, however very few parallel file systems has been provided for grid environments.

In this project we have developed a grid file system (objective 1) based on the Expand Parallel File System [2] designed by the research group. Expand is a parallel file system for clusters that uses standard servers and protocols for building parallel and distributed partitions where files are distributed. We have extended this parallel file system for providing a parallel file system for grids. The new version of Expand for grids uses standard grid technologies: GridFTP for data access, and the Resource Namespace Service for naming and data location (objective 2). This system allows the integration of existing heterogeneous servers span across multiple administrative domains in grids for providing parallel I/O services using standard interfaces like POSIX and MPI-IO. The new version of this parallel file system has been implemented using Globus [3], one of the most important middleware for building grid applications. The security is based on the grid security infrastructure (GSI) provided by the Globus Toolkit (objective 4).

Then, we described the main standard technologies used for providing the parallel file system for grids, and described the adaptation of Expand to grid environments.

GridFTP GridFTP [4] [3] is a data transfer protocol defined by the Global Grid Forum Recommendation, that provides secure and high performance data movement in grid systems. The Globus Toolkit provides the most commonly used implementation of that protocol, though others do exist (primarily tied to proprietary internal systems). This protocol extends the standard FTP protocol and includes the following features:

- Grid Security Infrastructure (GSI) support.
- Third-party control and data transfer.
- Parallel data transfer using multiple TCP streams.
- Striped data transfer using multiple servers.
- Partial file transfer and support for reliable and restartable data transfer.

The transfer data in our parallel file system is based on this protocol and the implementation provided by the Globus Toolkit. The access to this protocol in Globus is provided via two libraries: the Globus FTP control library, and the Globus FTP client library. The first library provides low-level services needed to implement FTP client and servers. The API provided by this library is protocol specific. The Globus FTP Client library provides a convenient way of accessing files on remote FTP servers.

Resource Namespace Service The Resource Namespace Service (RNS) is a specification of the Open File System Working Group (GFS-WG) of Open Grid Forum that allows the construction of a uniform, global, hierarchical namespace. It is a web service described by a RNS WSDL [5]

RNS is composed of two main components: *virtual directories* and *junctions*. Directories are virtual because they do not have any corresponding representation outside of the namespace. A junction is an RNS entry that interconnects a reference to an existing resource into the global

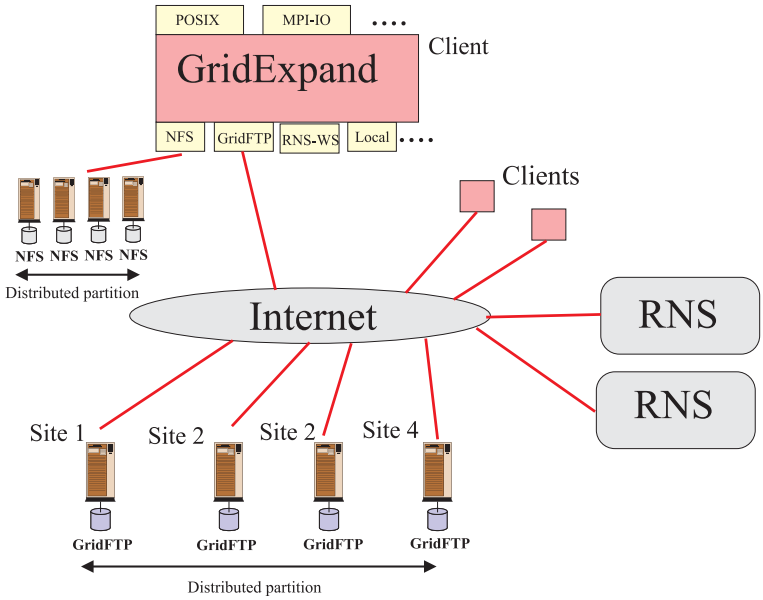


Figure 1: Expand Architecture for grid environments

namespace. There are two types of junction, referral junction and logical reference junction. The first is used to graft RNS namespaces. A logical reference junction contains a unique logical name. For this type of junction, the RNS service returns an endpoint reference (EPR). This endpoint is resolved using the Resource Endpoint Resolution Service or RNS resolver.

RNS is composed of several types of operations: operations for querying namespace entry information; operations for creating, removing, renaming, and updating entries; and operations for managing properties or status of an entry.

The RNS resolver is a service independent of RNS that has the mapping between logical names and endpoint references or address. For each logical name, the RNS resolver may store several endpoint references. The RNS resolver has operations for resolving names to endpoint references, and operations for creating, removing, and updating logical references.

We have implemented a prototype of web service for Globus that incorporates the RNS specification and that has been used in Expand for providing a global namespace in grids (objective 2).

2.1.1 Expand design for grids environments

Figure 1 shows the architecture of Expand for grid environments. This figure shows how Expand can be used for data management in clusters environments and how can be used to access to several sites using the GridFTP protocol. File data are striped by Expand among several servers using different protocols, using blocks of different sizes as stripping unit. Processes in clients use an Expand library to access to an Expand distributed partition. Expand offers an interface based on POSIX system calls. This interface, however, is not appropriate for parallel

applications using strided patterns with small access size [6]. For parallel applications, we use ROMIO [7] to support MPI-IO interface [8], implementing the appropriate Expand ADIO.

Expand combines several GridFTP servers (see Figure 1) in order to provide a generic distributed partition. The use of GridFTP allows to use servers allocated in different administrative domains. Each server provide one or more directories that are combined to build a distributed partition through the grid. All files in the system are striped across all GridFTP servers to facilitate parallel access, with each server storing conceptually a subfile of the parallel file.

Expand offers two different interfaces. The first interface is based on POSIX system call. This interface, however, is not appropriate for parallel applications using strided patterns with small access size [6]. Parallel applications can also used Expand with MPI-IO [8]. Expand have been integrated inside ROMIO [7] and can be used with MPICH. Portability in ROMIO is achieved using an *abstract-device interface for I/O* (ADIO).

Parallel access and authentication All file operations in Expand use a *virtual filehandle*. This virtual filehandle is the reference used in Expand to reference all operations. When Expand needs to access to a subfile, it uses the appropriated filehandle. For GridFTP Expand uses the appropriate handle managed by the GridFtp Client Library provided by Globus. To enhance I/O, user requests are split by the Expand library into parallel subrequests sent to the involved servers. When a request involves k GridFTP servers, Expand issues k requests in parallel to the servers, using threads to parallelize the operations. The same criteria is used in all Expand operations. A parallel operation to k servers is divided in k individual operations that are provided by the Globus GridFTP Client Library to access the corresponding subfile.

The access control and authentication is guaranteed in Expand for grid environments, because GridFTP uses the Grid Security Infrastructure provided by Globus. GSI uses public key cryptography as the basis for its functionality and provides secure communication (authenticated and perhaps confidential) between elements of a computational Grid, security across organizational boundaries, thus prohibiting a centrally-managed security system, and support *single sign-on* for users of the Grid, including delegation of credentials for computations that involve multiple resources and/or sites.

2.2 Branch Replication: A New Model for Data Replication in Large Scale Data Grids

Other contribution of this project has been the design of a new replication algorithm called *Branch Replication* [1] . This algorithm tries to bring more scalability through the parallel replication of a replica in several sites, where a replica is divided in subreplicas that reside in several nodes. The main features of this replication method are the following:

- In this algorithm, a storage node support the original file (root replica).
- When we want to create a new replica, n nodes have to be selected, where the subreplicas are going to be supported, but the union of all these subreplicas are the original replica.

[1] Algorithm described in the PhD Thesis *Performance Increase Mechanisms for data access in grids*, José María Pérez Menor

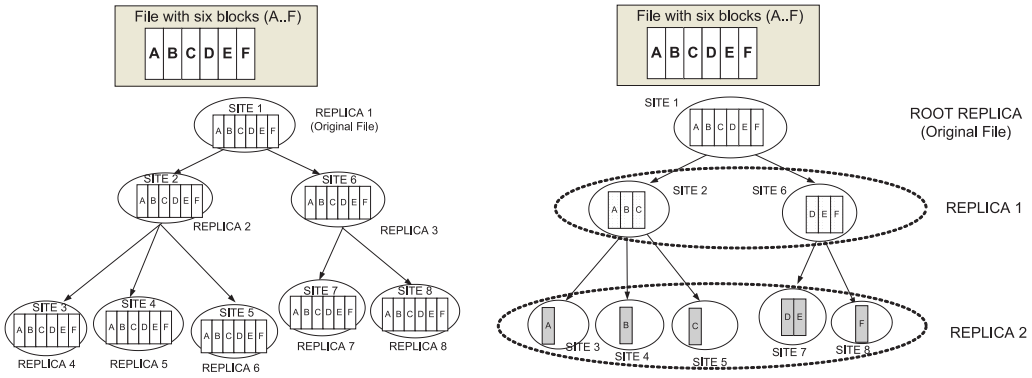


Figure 2: Differences between Hierarchical Replication (left) and Branch Replication (right)

We perform the stripping of the original replica. The subreplicas creation can be performed in parallel, so we can perform the replication in less time that the replica creation in one storage node.

- If we want to make another replica, the subreplicas can be split in several subsubreplicas, and so on.

This replication method allows to use this scheme in grid environments with many data storage resources with different sizes. Moreover, this approach allows parallel access to replicas and it requires less space per node to store the replicas. In traditional approaches, when an entire replica of a file with K GB is made we need K GB per node. Using the *branch replication* with n replicas we require more nodes, 2^{n-1} nodes. This is not a problem in large data grids, however, each node need less space to store data replicas. For an original file with K GB and n replicas, each storage node stores in average:

$$\frac{\sum_{i=1}^{N} \left(\frac{K}{2^{i-1}}\right)}{2^{n-1}} \quad GB$$

Figure 2 shows the differences between a Hierarchical Replication as the proposed in [9] and the branch replication designed in this project.

Replica Location Service Maintaining replicated data in grid environments requires a Replica Location Service (RLS) for providing a mechanism for registering new replicas and discovering them. To support the branch replication described above we must store for each subreplica the following information:

- FR: Replica of upper level.
- CR: Set of child subreplicas. It is stored the location of the files that support the subreplicas and the data replicated in each of them.
- BR: Set of subreplicas at the same level with a common upper level.

In order to provide a standard method for replica location we use the Resource Namespace Service (RNS), a specification of the Grid File System Working Group (GFS-WG) of Global Grid Forum that allows the construction of a uniform, global, hierarchical namespace. It is a web service described by a RNS WSDL [5].

Replica Consistency In order to maintain consistency among updates by clients we use the following mechanism. Clients only can modify the data located in the leaf nodes of the replication tree. So the location of the replica is reduced to the location of the subreplicas of more depth that support the range of data that the application wants to work with. The data update is performed from the child replicas to the parent using parallel and collective I/O techniques.

3 Result Indicators

3.1 PhD Thesis

- Alejandro Calderón Mateos. Fault tolerant techniques for parallel file systems in clusters. Universidad Carlos III de Madrid. 2005.
- José María Pérez Menor. Performance increase mechanisms for data access in grids. Universidad Carlos III de Madrid. 2006.

3.1.1 Thesis in progress

- Borja Bergua Guerra. Parallel file systems for grids and their use in jobs scheduling. Universidad Carlos III de Madrid.
- Margarita Reyes Rodríguez. New techniques for replication and scheduling in grid systems. Universidad Carlos III de Madrid.
- Luis Miguel Sanchez García. Scalability in parallel file systems for large scale clusters.
- Víctor Méndez Muñoz. Applying Artificial Intelligent techniques to replication services in grids. Universidad Carlos III de Madrid.

3.2 Publications

1. F. García, J. Carretero, A. Calderón, J. D. García, L. M. Sánchez. A global and parallel file system for Grids. Future Generation Computer Systems. Vol. 23, 2007. Pags. 116-122
2. M.S. Pérez, J. Carretero, F. García, J. M. Peña, V. Robles. MAPFS: A Flexible Multiagent Parallel File System for Clusters. Future Generation Computer Systems. Vol. 22, 2006. Pags. 620-632
3. David E. Singh, Florin Isaila, Ajelandro Calderón, Félix García, Jesús Carretero Multiple-Phase Collective I/O Technique for improving data access locality Aceptado para publicación en 15th Euromicro International Conference on Parallel, Distributed and Network-based Processing, PDP 2007. Napoles, Italia, 7-9 de febrero de 2007

4. F. Isaila, D. Singh, J. Carretero, F. García, G. Szeder, T. Moschny Integrating Logical and Physical File Models in the MPI-IO Implementation of Clusterfile 6th IEEE International Symposium on Cluster Computing and the Grid. CCGrid2006. Singapur 16-19 mayo de 2006. Pags: 462-469

5. F. Isaila, D. Singh, J. Carretero, F. García. On evaluating decentralized parallel I/O scheduling strategies for parallel file systems. Acepted for publication in Lecture Notes in Computer Science. Revisión posterior al congreso VEPCR'06, 7th International Meeting on High Performance Computing for Computacional Science. 2007.

6. V. Méndez, F. García PSO-Grid Data Replication Service Aceptado para publicación en Lecture Notes in Computer Science. Revisión posterior al congreso VEPCAR'06, 7th International Meeting on High Performance Computing for Computacional Science. A publicar en 2007

7. V. Méndez, F. García PSO vs. ACO, Data Grid Replication Services Performance Evaluation Lecture Notes in Computer Science. Vol 4331. Frontiers of High Performance Computing and Networking – ISPA 2006. International Workshop on Middleware Performance (WOMP). Sorrento, Italy, 4 -7 de diciembre de 2006. Pags. 833-843

8. M[a] B. Ibañez, F. García, J. Carretero A Profiling Approach for the Management of Writing in Irregular Applications Lecture Notes in Computer Science. Fourth International Symposium on Parallel and Distributed Processing and Applications (ISPA'2006) Sorrento, Italy ,4-6 de diciembrec. 2006. Pags. 251-259

9. L. M. Sanchez, F. Isaila, F. García, J. Carretero, R. Rabenseifner, P. Adamidis A New I/O Archirecture for Improving the Performance in Large Scale Clusters Lecture Notes in Computer Science. Vol 3984/2006. Computational Science and Its Applications - ICCSA 2006: International Conference, Glasgow, UK, May 8-11, 2006, Proceedings, Part V. Pags. 108-117

10. J. M. Pérez, F. García, J. Carretero, J. D. García, S. Escolar Data Driven Infrastructure and Policy Selection to Enhance Scientific Applications in Grid Lecture Notes in Computer Science. Vol 3458/2005. Scientific Applications of Grid Computing: First International Workshop, SAG 2004, Beijing, China, September 20-24, 2004, Revised Selected and Invited Papers. Pags. 59-74

11. M. B. Ibáñez, J. Carretero, F. Garcia. Symmetrical Data Sieving for Noncontiguous I/O Accesses in Molecular Dynamics Simulations. Lecture notes in Computer Science. Vol 3666/2005. Recent Advances in Parallel Virtual Machine and Message Passing Interface: 12th European PVM/MPI Users' Group Meeting Sorrento, Italy, September 18-21, 2005. Proceedings. Editors: Beniamino Di Martino, Dieter Kranzlmüller, Jack Dongarra. Pags. 441-448

12. M. S. Perez, J. Carretero, F. Garcia, J. M. Peña, V. Robles MAPFS-GRID: A Flexible Architecture for Data-Intensive Grid Applications Lecture Notes in Computer Science series. Vol. 2970/2004. Grid Computing: First European Across Grids Conference, Santiago de Compostela, Spain, February 13-14, 2003. Revised Papers. Pags. 111-118. Publicado en 2004

13. J. M. Perez, F. Garcia, J. Carretero, A. Calderón, J. Fernández A Parallel I/O Middleware to Integrate Heterogeneous Storage Resources on Grids Lecture Notes in Computer Science series. Vol. 2970/2004. Grid Computing: First European Across Grids Conference, Santiago de Compostela, Spain, February 13-14, 2004. Revised Papers. Pags. 124-131

14. D. E. Singh, F. Garcia, F. Carretero. I/O optimization for an air pollution model. Parallel Computing 2005 (Parco2005). 13-16 de septiembre de 2005 Malaga. Spain

15. J.M. Perez, L. M. Sánchez, , F. García, A. Calderón, J. Carretero. High Performance Java Input/Output for Heterogeneous Distributed Computing. 10th IEEE Symposium on Computers and Communications / ISCC 2005. Pp. 969-974. La Manga del Mar Menor, Cartagena, España. Junio 27-30, 2005

16. A. Calderón, F. García., J. Carretero, J. M. Pérez, L. M. Sánchez. A Fault Tolerant MPI-IO implementation using the Expand Parallel File System. 13th Euromicro Conference on Parallel, Distributed and Network-based Processing, PDP 2005. Lugano, Suiza. 9-11 February 2005. ISBN: 0-7695-2280-7. Pags. 274-281

17. A. Calderón, F. García, J. Carretero, L.M. Sánchez, J. M. Pérez. File schemes for fault tolerant in heterogeneous distributed storage systems New Trends in Distributed Data Access. Proceedings of the 2004 International Multiconferece in computer science and computer engineering (PDPTA'04), ISBN: 1-932415-24-6. Págs. 561-567

18. M.S. Pérez, A. Sánchez, J.M. Peña, V. Robles, J. Carretero, F. García. Storage Groups: A New Approach for Providing Dynamic reconfiguration in Data-Based Clusters. Proceedings of the IASTED International Conference on Parallel and Distributed Computing and Networks. February 17-19, 2004, Innsbruck, Austria. Pp: 70-75

19. A. Núñez, J. Carretero, F. García, L.M. Sánchez Sistema de ficheros paralelo basado en servicios web XVII Jornadas de Paralelismo. Albacete. 18-20 de septiembre de 2006. Pags. 191-186

20. D. E. Singh, Mª B. Ibáñez, F. Isaila, F. García-Carballeira, J. Carretero. Optimizaciones de entrada/salida para aplicaciones de dinámica molecular XVI Jornadas de Paralelismo. I Congreso Español de Informática. Granada. 13-16 de septiembre de 2005. Pags. 685-692

21. J. M. Pérez Menor, F. García Carballeira, J. Carretero , J. D. García , A. Calderón Incremento de prestaciones en el acceso a datos en Data Grids XVI Jornadas de Paralelismo. I Congreso Español de Informática. Granada. 13-16 de septiembre de 2005. Pags. 347-354

22. L. M. Sánchez, J. M. Pérez, F. García, A. Calderón, J. Carretero. Arquitectura escalable para E/S de altas prestaciones en sistemas heterogéneos XV Jornadas de Paralelismo. Almería. 15-17 de septiembre de 2004.

3.3 Cooperation with other groups

- Thematic net for the Coordination of Middleware Activities in Grids. This thematic net is coordinated by the Universidad Complutense de Madrid (http://asds.dacya.ucm.es/GridMiddleware/presentacion.php)
- GridiMADRID (http://www.gridimadrid.org), An initiative of grid in the Comunidad de Madrid.

References

[1] Open Grid Forum, Grid File System Working Group, http://www.opf.org

[2] F. Garcia, A. Calderon, J. Carretero, J.M. Perez, J. Fernandez, The design of the Expand parallel file system, International Journal of High Performance Computing Applications, Vol 17, N.1, 2003, pp. 21-37.

[3] The Globus Toolkit, http://www.globus.org

[4] B. Allock, J. Bester, J. Bresnahn, A. Chervenak, I. Foster, C. Keseelman, S. Meder, V. Nefedova, D. Quesnel, S. Tuecke, Secure, efficient data rransport and replica management for high performance data-intensive computing, In Proceedings of the Eighteenth IEEE Symposium on Mass Storage Systems and Technologies - Volume 00, April 17 - 20, 2001, pp. 13-28.

[5] M. Pereira, O. Tatebe, L. Luan, T. Anderson, J. Xu, Resource Namespace Service specification, nov. 2005. http://www.global.http://www.globalgridforum.net.

[6] N. Nieuwejaar, D. Kotz, The Galley parallel file system, In Proceedings of the 10th ACM International Conference on Supercomputing, May 1996.

[7] W. Gropp R. Takhur and E. Lusk, An abstract-device interface for implementing portable parallel-I/O interfaces, In Proceedings of the 6th Symposium on the Frontiers of Massively Parallel Computation, Oct. 1996, pp. 180-187.

[8] A. Calderon, F. Garcia, J. Carretero, J.M. Perez, J. Fernandez, An implementation of MPI-IO on Expand: A parallel file system based on NFS servers, In 9th PVM/MPI European User´s Group, Johannes Kepler University Linz, Austria. sep 29-oct 2, 2002. Pp. 306-313.

[9] H. Lamehamedi, Z. Shentu, B. Szymanski, E. Deelman. Simulation of Dynamic Data Replication Strategies in Data Grids. Proc. 12th Heterogeneous Computing Workshop (HCW2003) Nice, France, April 2003, IEEE Computer Science Press, Los Alamitos, CA, 2003.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Infraestructura tecnológica de servicios semánticos para la Web Semántica TIN2004-02660

Gómez Pérez, A.; Aguado de Cea, G.; Álvarez de Mon Rego, I.; Barrasa Rodríguez, J.; García Castro, R.; González Cabero, R.; López Cima, Á.; Pareja Lora, A.; Ramos Gargantilla, J.A. and Suarez-Figueroa, M.C.; Villazón Terrazas, B.
Universidad Politécnica de Madrid

Corcho, O.
University of Manchester

Fernández-López, M.
Universidad San Pablo CEU

Manuel Lama Penin
Universidad de Santiago de Compostela

Abstract

This project aims at creating a network of distributed interoperable semantic services for building more complex ones. These services will be available in semantic Web service libraries, so that they can be invoked by other systems (e.g., semantic portals, software agents, etc.). Thus, to accomplish this objective, the project proposes:

a) To create specific technology for developing and composing Semantic Web Services.
b) To migrate the WebODE ontology development workbench to this new distributed interoperable semantic service architecture.
c) To develop new semantic services (ontology learning, ontology mappings, incremental ontology evaluation, and ontology evolution).
d) To develop technological support that eases semantic portal interoperability, using Web services and Semantic Web Services.

The project results will be open source, so as to improve their technological transfer. The quality of these results is ensured by a benchmarking process.
Keywords: Ontologies and Semantic Web

1 Project objectives

This project aims at creating **a network of distributed interoperable semantic services and thus to build more complex ones. These services will be available in semantic Web service libraries, so that they can be invoked by other systems (e.g., semantic portals, software agents, etc.)**. In order to accomplish this objective, the project proposes to achieve the following sub-goals:

1. **Methodological and technological support for the creation of semantic Web services**. Specifically, we will have to work on the design, discovery, and (semi)automatic composition

of semantic Web services and in the management and building of distributed semantic Web service libraries.

2. **Interoperability and management of semantic portals**. We propose an architecture that facilitates the interoperability between semantic portals through Web services and semantic Web services.
3. **Basic research and development of an infrastructure in the ontology area**. We will develop new mapping and evolution services, extend the ontology learning services, and transform the evaluation service.
    a. **Creation of a mapping service** between heterogeneous knowledge models. It will be created a mapping specification language, tools for mapping development and management, and a method for automating mapping discovery.
    b. **Creation of the ontology evolution management service**, by identifying the components to be managed and providing software support.
    c. **Extension of the automatic ontology learning service resulting from the TIC-01-2745 project**. It will use standardized linguistic sources and information from distributed databases that will be accessed via ontology mappings. The extensions will consist on the improvement of concept and taxonomy learning, and on the learning of attributes, relations, and instances.
    d. **Transformation of the evaluation service into an incremental evaluation service** that analyses the changes performed in the ontologies and determines their consistency.
4. **Improvement of the METHONTOLOGY ontology development methodology and adaptation of this methodology to the new service paradigm.** To be precise, the methods to perform knowledge acquisition, integration, evaluation, configuration management, and maintenance will be defined in detail and METHONTOLOGY will be extended, improved, and adapted to the new distributed semantic service paradigm.
5. Decentralization of the current services of the WebODE platform and its transformation to semantic services that can be used by other platforms and/or applications. Migration of the WebODE platform to open source to favor the development of the Semantic Web and of semantic-based applications, and therefore to improve technology transfer to other national and international research groups as well as to companies and organizations. Concretely, we will redesign and reimplement some parts of the source code, obtain a continuous quality improvement through the establishment of a continuous benchmarking process that involves the development of supporting technology, and we will generate documentation, installation and user manuals, a list of most frequent errors, management of mailing lists, etc.

2 Main achievements of the project

2.1 Methodological and technological support for the creation of semantic Web services

At the beginning of the project, we had a conceptual model for describing semantic web and semantic grid services. We have specified, designed and developed an open source graphical user interface for annotating and designing semantic services for the Grid and for the Web that uses ontologies and Problem-Solving Methods for describing the features of service operations.
Choreography and orchestration of semantic services are represented by workflows that model the coordination between the services involved in the choreography and/or orchestration. In our approach, these workflows are described with High-Level Petri Nets (HLPN): each one of the workflow primitives (such as and/or-split, and/or-join, etc.) are modelled by a given HLPN.

As semantic services are described by means of ontologies, we have developed a workflow ontology (WO) to represent service choreography and orchestration. The core of the WO is an ontology of HLPN, which describes the semantics of both the dynamic and static models of the HLPN formalism: we have developed a concept taxonomy where the HLPN components are described and a set of axioms for constrain checking between the taxonomy concepts. This HLPN ontology is based on the ISO/IEC 15909-1 standard that compiles the shared knowledge of the Petri net community about HLPN.

2.2 Interoperability and management of semantic portals

The starting point at the beginning of the reporting period was the design of a conceptual architecture for semantic portal interoperability and data exchange in ODESeW. During this period, the ODESeW architecture has been refined.

The implementation efforts were devoted to the external gateway component, whose goal is to access an external resource that contains information according the domain ontologies of the portal and to annotate or translate this information according to the domain ontologies.

This activity makes use of results from other parts of the project. As ODESeW employs the WebODE ontology development platform for storing and managing multiple ontologies, data exchange in ODESeW strongly depends on WebODE import and export facilities, which are being benchmarked in this project as seen below.

2.3 Basic research and development of an infrastructure in the ontology area

Regarding the **creation of a mapping service between heterogeneous knowledge models**, the tasks that have been performed cover the following types of mappings: mappings between two ontologies and mappings between ontologies and databases.

In the case of mappings between two ontologies, the different casuistry has been analysed and the actions for carrying out the merge process have been established, though only for the limited case of concepts, attributes and relations. Research has been carried out in the use of linguistic tools for the automatic discovering of mappings between ontologies, as appears below in this paper. Furthermore, for the future mapping representation the existing representations have been analysed and a first version of a mapping representation ontology has been developed.

A WebODE Mapping service has been created for browsing and editing mappings. Currently, the service only shows mappings from the automatic discovery implemented.

In the case of the mappings between ontologies and databases, the group developed R_2O, an extensible and fully declarative language that describes mappings between relational database schemas and ontologies. Once mappings are defined in a R_2O mapping document, they are processed automatically by the ODEMapster mapping processor to populate an ontology (batch process) or to translate a query (on demand).

The work on the ODEMapster engine has been extended to fully cover the on-demand mapping scenario. A query defined in terms of the elements in the ontology (expressed in ODEMQL) is processed and translated in terms of the underlying database and the results of the dynamically generated relational query are translated into instances of the ontology (expressed in RDF). Compliance with standard query languages such as SPARQL is under development.

A module for implementing the Semantic Mapper algorithm for automatic database schema to ontology matching discovery has also been developed. Such module takes as input the relational schema of an existing database and an OWL ontology and delivers as output an skeleton of an R_2O mapping document containing the matches found.

We are also extending ODEMapster and enriching R_2O language to describe mappings between XML schemas and ontologies.

As for the **extension of the automatic ontology learning service resulting from the TIC-01-2745 project**, the tasks performed are not just related to the ontology learning but also to the ontology mapping discovery.
We have studied the use of the SUBCLASS-OF relation in texts. More than 60 linguistic patterns have been identified and systematized according to the main lemma, taking into account that each lemma can present different word-forms. In this case the patterns have been analyzed according to their word-forms.
Furthermore, 30 lexical collocations that provide semantic information related to concept extraction and 10 metalinguistic devices that help extract the SUBCLASS-OF relation have been identified; in addition, 25 linguistic pattern-based rules have been defined from the patterns identified for each lemma, and for each variant in these lemmas, in order to formalize them and achieve a (semi)automatic extraction of the concepts and relations in the texts found in the Web.
We have identified the localization of a new component in the architecture that can exploit the use of rules for knowledge extraction, and that can be used for ontology learning and for discovering mappings between several ontologies. We have started the implementation of this information extraction component in ODELinger. We must add that the OntoTagger-ODELinger communication was outdated. In the last few months, this communication has been updated and the ODELinger database model has been extended for covering semantic annotation information.
Meanwhile, an automatic mapping discovery tool has been implemented into WebODE that uses EuroWordNet to create mappings between concepts with the synonymy semantic relation.

2.4 Improvement of the METHONTOLOGY ontology development methodology and adaptation of this methodology to the new service paradigm

The work performed is focused on the notion of networked ontologies, analyzing different dependencies between the activities of the development process of several ontologies. Based on such analyses, the development process proposed in METHONTOLOGY has been modified.
The activities involved in the development of networked ontologies (conceptualization, merging, documentation, evaluation, etc.) are being defined within the new notion. We have created a consensuated glossary of activities to be used in the improved development process.
Research has also been carried out on the use of formal ontologies in the design, evaluation and integration of ontologies and in the analysis of the potential of Web 2.0 technologies to build ontologies collaboratively.

2.5 Migration of the WebODE platform to open source

The development of WebODE OS has started. WebODE OS is an ontology development and management platform extensible, scalable and open source. The WebODE OS platform is being developed using J2SE 1.5 over JBoss, an application server that follows the J2EE 1.4 standard specification. The software is being developed platform-independent (Linux and Windows) and database independent (MySQL and Oracle). The knowledge model of WebODE OS has been developed and implemented in EJB 3.0; the ontology management API has also been implemented. During this process, unit tests have been performed.

Simultaneously, the FPI grant holder has started the benchmarking of the WebODE platform.

Taking into account the current State of the Art of software benchmarking, we have defined the process to follow for benchmarking WebODE, considering the different existing methodologies on software and software process improvement of the Benchmarking, Experimentation in Software Engineering and Measurement in Software Engineering areas. For each of these areas, we have identified the main definitions, classifications proposed in the literature, and the tasks of the methodologies.
From the State of the Art previously cited, we have defined the process and the criteria for benchmarking the WebODE platform and its services. One of the main factors taken into account when defining the process has been the continuous improvement of the WebODE platform.
This benchmarking has been planned with the goal of assessing and improving the interoperability of WebODE with the main ontology development tools and ontology repositories. The benchmarking will evaluate the degree of interoperability in terms of information addition or loss that occurs when two tools interchange an ontology by translating it to an interchange language. The interchange languages considered are the languages recommended by the W3C, namely RDF(S) and OWL.
Furthermore, different software has been developed to support the execution of the experimentation in the benchmarking and the analysis of the results obtained in the benchmarking.

3 Result indicators

3.1 Personnel being trained

Two grants have been assigned and charged to the project:

- A FPI grant (Beca Predoctoral de Formación de Personal Investigador) has been assigned to Raúl García Castro for 48 months (BES-2005-8024).
- A project grant has been assigned to José Ángel Ramos Gargantilla for 9 months.

A research stay related to the project:

- Mariano Fernández López had a research stay at UPM from 1st January 2006 to 1st February 2006 to work in the creation of a mapping service between heterogeneous knowledge models and in the improvement of the METHONTOLOGY ontology development methodology.

3.2 Dissemination

A summary of the dissemination activities performed in this project is the following: 2 papers in international journals, 14 papers in international conferences, 2 posters in international conferences, 3 papers in national conferences, 5 papers in international workshops, 2 papers in national workshops, 7 book chapters, and 1 PhD thesis. One of these posters was awarded as Best Poster in the WWW2006 international conference.
Furthermore, technology demonstrations have been carried out for the following companies or socio-economic agents: Intelligent Software Components, Atos Origin, TPI, Telefónica Sistemas, Gabinete de Investigación Militar Operativa (GIMO), Instituto Geográfico Nacional, and Parquesoft (Colombia).

3.3 Technology transfer

The semantic portal technology is being used to develop semantic portals in the European projects such as Agentlink III, NeOn and OntoGrid, in the Network of Excellence Knowledge Web, and in the Spanish network Red Temática de Web Semántica. It has also been used to develop the

semantic portal of this project. The Knowledge Web, OntoGrid and NeOn portals include a functionality to help project management and to create progress reports. With this idea, we have proposed the creation of a spin-off company for managing EU projects based on this functionality built with ODESeW. The proposal won the "XIII Programa de Creación de Empresas de Base Tecnológica UPM", a competition at UPM.
There also exist with the group on Sheffield University lead by Prof. Dr. Fabio Ciravegna a contract for using the ODESeW technology in the management of European projects.
The R2O technology is being used in the SEEMP (Single European Employment Market-Place) European project and in the NeOn European project.
A collaboration contract has been signed with Spanish Instituto Geográfico Nacional (IGN) to develop an ontology by refining and extending BCN25, and to develop an algorithm for the automatic identification of mappings between the ontology, nomenclators and the IGN cartographic databases.

3.4 Participation in international projects

The research group participates in the following projects of the Sixth European Framework Programme:

- NeOn: Lifecycle Support for Networked Ontologies (FP6-027595).
  Type of participation: consortium member. From 01-03-2006 to 28-02-2010.
- Single European Market Place SEEMP (FP6-027347)
  Type of participation: consortium member. From 01-01-2006 to 30-06-2008.
- OntoGrid: Paving the way for Knowledgeable Grid Services and Systems (FP6-511513)
  Type of participation: project coordinator. From 01-09-2004 to 31-08-2007.
- Knowledge Web (FP6-507482) Network of Excellence
  Type of participation: consortium member. From 01-01-2004 to 31-12-2007.

Additionally, it must be highlighted the participation of the research group in the Spanish thematic network that follows below, whose leader is the project coordinator and in which 24 Spanish universities are participating with 180 researchers:

- Red Temática de Web Semántica (TSI2006-26928-E)
  Type of participation: project coordinator. From 01-01-2007 to 31-12-2007.

3.5 Collaboration with research groups

The Ontology Engineering Group collaborates with all the universities and companies that participate in the projects presented above. These partners are from Austria (1), Belgium (2), France (4), Germany (7), Greece (3), Ireland (1), Italy (8), The Netherlands (3), Poland (1), Slovenia (1), Spain (28), Switzerland (1), and United Kingdom (5).
Besides these collaborations, this project has provided the following collaborations with other non-academic entities:

- Several demonstrations to non-academic entities have been performed. Also, the group has applied for three CENIT projects with companies that have received technology demonstrations.
- The idea of networked ontologies has led to the creation of an IP European project called NeOn (FP6-027595).

- The proposed architecture for semantic portal interoperability will be used to interchange content between the NeOn Project portals. These portals are now being developed by the Atos Origin company and by the UPM.
- The WebODE ontology development platform is used by the Instituto Geográfico Nacional to develop ontologies in the domain of Geographic Information Systems.

This project has provided the following collaborations with other research groups:

- Collaboration with the Instituto de Lingüística Aplicada (IULA) from Universidad Pompeu Fabra.
- Contact with teachers from Bristol University in the BAAL Conference to present the work results.
- Meeting with Dr. Alejandro Curado from Universidad de Extremadura and a group of teachers from different universities (Valladolid, Granada, Valencia) to present the results of the linguistic section.
- Collaboration with the research center Centrum voor Vaktaal en Communicatie (CVC), led by Dra. Rita Temmerman, from Erasmushogeschool in Brussels.
- Collaboration with the research group coordinated by Dr. Ricardo Mairal, professor at the UNED and with Dr. Francisco Ruiz de Mendoza, professor at the University of La Rioja.
- Collaboration with Dr. Enrique Alcaraz Varó, professor and director of Instituto Universitario e Lenguas Modernas Aplicadas (IULMA) Alicante.
- Contact with Dr. Carlos Subirats, Professor at the Universidad Autónoma de Barcelona.
- The group coordinated by Dra. Alcina Caudet, Universitat Jaume I, Castellón.
- Antonio Pareja was invited to deliver a seminar at the Universitat Jaume I, in the summer course "Terminología e Ingeniería Lingüística", from 10 to 13 July 2006.

During 2006 the NLP group has received two visiting scholars: Dra. Alcina who stayed three months, and Dña. Rosario Bautista, from the University of Malaga, who stayed two months.

4 References

Creation of semantic Web services

[1] SGSDesigner, a Graphical Interface for Annotating and Designing Semantic Grid and Web Services. Gómez-Pérez, A. González Cabero, R. Poster in the 15th Internacional World Wide Web Conference. Edimburgo, Escocia. May 2006. *Best poster award.*

Interoperability and management of semantic portals

[2] ODESeW for the creation of R&D projects' Intranets and Extranets. Corcho, O.; Gómez-Pérez, A.; López-Cima, A. IntraWebs Workshop en la 15th Internacional World Wide Web Conference May 2006.

[3] The ODESeW 2.0 Semantic Web application. . Corcho, O.; Gómez-Pérez, A.; López-Cima, A. Póster en la 15th Internacional World Wide Web Conference en mayo de 2006.

[4] A platform for the development of Semantic Web portals. Corcho, O; López-Cima, A.; Gómez-Pérez, A. 6th International Conference on Web Engineering (ICWE2006). July 2006.

[5] The ODESeW platform as a tool for managing EU projects: the KnowledgeWeb case study. Corcho, O.; Gómez-Pérez, A.; López-Cima, A.; Suárez-Figueroa, M.C. 15th International Conference on Knowledge Engineering and Knowledge Management Managing Knowledge in a World of Networks (EKAW006). July 2006.

Mappings between heterogeneous knowledge models

[6] OEGMerge: un modelo de mezcla de ontologías basado en casuísticas. De Diego R, Fernández-López M, Gómez-Pérez A, Ramos JA. CAEPIA 2005: Conferencia de la Asociación Española para la Inteligencia Artificial. Santiago de Compostela, España.16-18 de noviembre de 2005. Editado por la Asociación Española para la Inteligencia Artificial. ISBN: 84-96474-13-5. Vol I. Páginas: 283-292. (Seleccionado para ser publicado en la revista Iberoamericana de IA.).
[7] OEGMerge: a case-based model for merging ontologies. Fernández-López M, Gómez-Pérez A, Ramos JA. MDAI 2006: Modeling Decisions for Artificial Intelligence. Edited by Torra V and Narukawa Y. April 2006. Tarragona, Spain.
[8] OEGMerge: modelo de mezcla de ontologías basado en casuísticas. Fernández-López M, Gómez-Pérez A, Ramos JA. Revista Iberoamericana de Inteligencia Artificial. 2006.
[9] Semantic Upgrade and Publication of legacy data. Jesús Barrasa. Chapter in book: Ontologies for Software Engineering and Software Technology. Calero, Coral; Ruiz, Francisco; Piattini, Mario (Eds.) ISBN: 978-3-540-34517-6.
[10] A Semantic Portal for Fund Finding in the EU: Semantic Upgrade, Integration and Publication of heterogeneous legacy data. Jesús Barrasa Rodrígez, Óscar Corcho, Asunción Gómez-Pérez. 10th International Conference on Knowledge-Based & Intelligent Information & Engineering Systems KES2006. Bournemouth, UK. 9-11 October 2006.
[11] Upgrading Relational Legacy Data to the Semantic Web. Jesús Barrasa Rodrígez, Asunción Gómez-Pérez. International World Wide Web Conference. Edinburgh. UK. 23-26 May 2006.
[12] Modelo para la definición automática de correspondencias semánticas entre ontologías y modelos relacionales. December 2006. Jesús Barrasa Rodríguez. PHD Thesis. Departamento de Inteligencia Artificial. Facultad de Informática. UPM.

Extension of the automatic ontology learning service

[13] An Overview of Methods and Tools for Ontology Learning from Text. Gómez-Pérez, A.; Manzano-Macho, D. The Knowledge Engineering Review. Cambridge University Press (Reino Unido).2005
[14] Reflections on Language Use in the Academic Context. Durán Escribano, P.; Aguado de Cea, G.; Roldán Riejos, A. (editoras). Introduction to Reflections on Language Use in the Academic Context. Editores: Fundación Conde del Valle de Salazar. UPM Páginas161. Madrid 2005.
[15] Conocimiento lingüístico para la Web semántica: Los nombres de persona en el dominio del cine. Aguado de Cea, Guadalupe; Alvarez de Mon, Inmaculada; Pareja-Lora, Antonio. Perspectivas interdisciplinares de la lingüística aplicada . Mª Luisa Carrió Pastor (editora). Editorial: AESLA: UPV Páginas 369 a 380 Valencia 2005. Nº ISBN: 84- 609-6062-5.
[16] Contenido, Forma y Función en textos del dominio del cine: La reseña y la crítica cinematográfica. Aguado de Cea, Guadalupe; Alvarez de Mon, Inmaculada; Pareja-Lora, Antonio. XXIV Congreso de la Asociación Española de Lingüística Aplicada, Palma de Mallorca, España 10-14 Marzo 2005 En prensa.
[17] Fraseología y neología en la sociedad de la información. Aguado, G. Ciclo de conferencias sobre Nuevas tendencias en las lenguas de especialidad.Universidad Alfonso X España 2 Marzo, 2005
[18] Phraseology in LSP. Aguado de Cea, Guadalupe. Workshop on New Issues on Applied Linguistics. Universitat Jaume I, Castelló, España 22 Abril, 2005.
[19] El lenguaje de las nuevas tecnologías. Aguado de Cea, Guadalupe. Mesa Redonda: La terminología tecnológica. Instituto de la Ingeniería de España 11 Mayo 2005.
[20] A contrastive study of encapsulation and prospection in written scientific text. Alvarez de Mon y Rego, Inmaculada. XV European Symposium on Language for Special Purposes. University of Bergamo 29 Agosto a 2 Septiembre 2005. En prensa.

[21] The language of film reviews in Spanish: The interaction of purpose, mode, and ideology in the construction of texts. Aguado, G.; Alvarez de Mon, I.; Pareja-Lora, A. British Association of Applied Linguistics- BAAL 2005. Bristol, Inglaterra 15 a 17de Septiembre 2005. En prensa.
[22] Aguado, G. & Alvarez de Mon "Corpus linguistics applied to knowledge extraction for ontological engineering: the phraseology of classification from English into Spanish", to be published by Equinox.
[23] Aguado, G. Alvarez de Mon, I. & Pareja-Lora, A. (forthcoming) "Getting to Grips with the Complex field of Semantic Annotation", edited by Amparo Alcina.
[24] Aguado, G & Alvarez de Mon 2006. "Corpus linguistics applied to knowledge extraction for ontological engineering: the phraseology of classification in Spanish", paper presented in Cork (Ireland) at the BAAL Conference, September 2006.
[25] Aguado, G & Alvarez de Mon 2006. "Estructuras de clasificación en español. Terminología y adquisición de conocimiento explícito para la Web semántica", V Internacional Conference of the AELFE Association. Zaragoza, September, 2006, ISBN- 84-7733-846-9.
[26] Aguado de Cea, G. 2006. "De bits y bugs a blogs y webs: aspectos interdisciplinares, socioculturales y lingüísticos de la terminología informática" in CORCILLVM. Estudios de traducción, lingüística y filología dedicados a Valentín García Yebra. Gonzalo C. y Hernúñez, P. (eds.), Madrid: Arco Libros, 693- 720, ISBN: 84-7635-648-X.
[27] Aguado de Cea, G. 2006. "La fraseología en las lenguas de especialidad", Invited plenary lecture at the I Internacional Conference on "Lenguas Modernas Aplicadas" (Applied Modern Languages), May, Alicante, 2006.

Improvement of the METHONTOLOGY ontology development methodology

[28] O. Corcho, M. Fernández-López, A. Gómez-Pérez. Ontological Engineering: principles, methods, tools and languages. Ontologies for Software Engineering and Technology. Lecture Notes in Computer Science. Springer 2006. Eds Francisco Ruiz, Coral Calero and Mario Piattini.

Benchmarking ontology development tools

[29] García-Castro R.; Gómez-Pérez A. "Guidelines for Evaluating the Performance of Ontology Management APIs". Proceedings of the 4th International Semantic Web Conference (ISWC 2005). Galway, November 2005.
[30] García-Castro R.; Gómez-Pérez A. "A Method for Performing an Exhaustive Evaluation of RDF(S) Importers". Proceedings of the International Workshop on Scalable Semantic Web Knowledge Based Systems (SSWS2005). New York, November 20th 2005.
[31] García-Castro R.; Gómez-Pérez A. "Benchmark Suites for Improving the RDF(S) Importers and Exporters of Ontology Development Tools" Proceedings of the 3rd European Semantic Web Conference (ESWC 2006). Budva, Montenegro. June 2006.
[32] García-Castro, R.; Gómez-Pérez, A. "Interoperability of Protégé using RDF(S) as Interchange Language". 9th International Protégé Conference (Protégé2006). July, 2006. Stanford, USA.
[33] David, S., García-Castro, R., Gómez-Pérez, A. "Defining a Benchmark Suite for Evaluating the Import of OWL Lite Ontologies" OWL: Experiences and Directions 2006 (OWLED2006) Athens, GA, USA, 10-11 November de 2006.
[34] García-Castro, R. "Keynote: Towards the improvement of the Semantic Web technology". Second International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS2006) collocated with the 5th International Semantic Web Conference (ISWC2006) Athens, GA, USA, 5 de noviembre de 2006.
[35] García-Castro, R. "Keynote: Benchmarking como herramienta de transferencia tecnológica" III Encuentro Internacional de Investigadores en Informática y Computación Parquesoft Popayán, Colombia, 4 octubre 2006.

[36] García-Castro, R. "Keynote: Tecnologías de la Web Semántica: cómo funcionan y cómo interoperan" IV Seminario Internacional Tecnologías Internet (SITI 2006). Popayán, Colombia, 5-7 octubre 2006.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Automatic Textual Annotation on Image Databases based on Visual Descriptors combining Spatio-Chromatic Information TIN2004-02970

M. Vanrell, R. Baldrich, X. Otazu, A. Párraga,
R. Benavente, A. Salvatella, F. Tous, S. Álvarez

Abstract

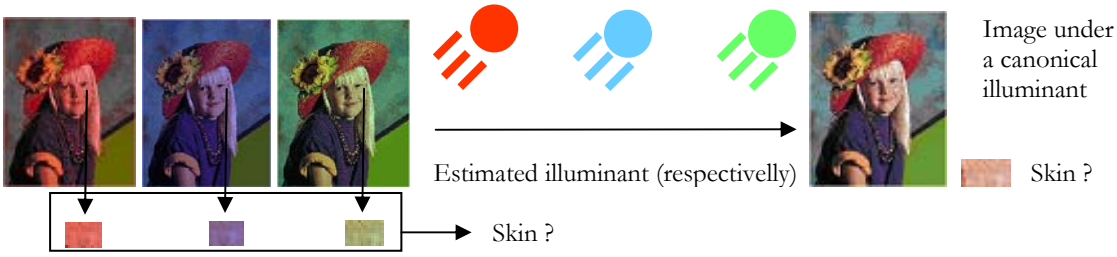
The big volume of multimedia content that is being generated currently, take us to pay think on the automatisation of annotation tasks. A textual annotation improves the retrieval process since queries can be formulated in natural language terms. In the last years a great effort has been made to define visual descriptors allowing specifying the contents of pictures. However, it is accepted that still exists a *semantic gap*, referring to the broad distance between the numerical visual descriptors and the interpretation given by a human operator on the picture content. In this project we plan to bridge on this gap by proposing the definition of new visual descriptors that allow to derive textual annotators. Given a particular image, a textual annotator will return a textual annotation in natural language terms about the image content. MPEG-7 standard gives 4 types of image visual descriptors: colour, texture, shape and motion. In this project we will focus on the first two visual cues, texture and colour, and we propose to define three new visual descriptors: a colour descriptor based on the name of colours, a texture descriptor based on a decomposition of texture in their basic structural components, an illuminant descriptor.
Keywords:

1 Project Goals

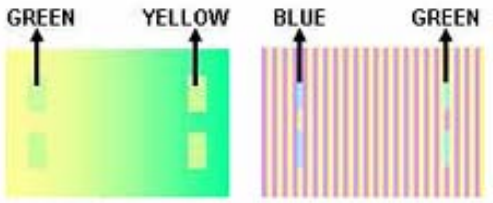
The global goal of this project is to go deeply on the problem of automatic image annotation. It is accepted that still exists a *semantic gap*, referring to the broad distance between the numerical visual descriptors and the interpretation given by a human operator on the picture content. Given a particular image, a textual annotator will return a textual annotation in natural language terms about the image content. MPEG-7 standard gives 4 types of image visual descriptors: colour, texture, shape and motion. Considering the previous know-how of the group, we do not try to bridge on the semantic gap globally; we will focus on the first two visual cues: **texture and colour**.

Towards this goal, we divided the general goals of the project in seven subgoals:

Goal 1. Illuminant estimation. Given a image we need to estimate the most likely illuminant under which the colour image has been acquired. This goal has been referred in colour vision as the colour constancy problem, according to the ability of the human visual system to deal with this task. In the following figure we show the need to solve this problem if we want to annotate a colour image.



Goal 2. Perceptual representation of colour and texture. Colour perception in human visual system depends on texture properties, hence a given colour is perceived differently depending on the colour of its neighbourhood. This can be seen in the following images, where the same stimulus is perceived as green on the left side and as yellow on the right side, on the left figure. On the other hand, figure on the right presents the same stimulus that is perceived as blue or as green on the left or on the right respectively. In both cases it depends on the near surround and on the spatial frequency of the background. To annotate correctly in both figures, we propose to build the *perceived image*, that is, an image where colours are represented as are they seen by the human visual system instead of directly working on the image given by the acquisition device.



Goal 3. Unsupervised image segmentation. On the perceived image we need to segment image regions considering colour and texture. To this end, we need to work in two steps, firstly detecting image blobs, followed by a a grouping step to get the image segments sharing blob properties. In order to get the results we show in the following figures.

Goal 4. Definition of illuminant descriptors and annotators. The results derived from goal 1, must be interpreted in the direction is giving an illuminant descriptor, such as, *"indoor", "outdoor", "sunny", "cloudy", "halogen"*, etc.

Goal 5. Definition of colour descriptors and annotators. Describing the colour of the image blobs imply to define a colour naming model, where for a given colour a basic colour name is assigned, these are: *"white", "black", "red", "green", "blue", "yellow", "pink", "brown", "orange", "purple", "grey"*.

Goal 6. Definition of texture descriptors and annotators. To describe spatio-chromatic relations in the image, we consider a texture as a group of blobs sharing its local properties, hence a texture will be described as set of blobs with different attributes, such as, *"elongated"*, or *"circular"* for the shape, *"X-oriented"* in a specific direction (denoted by X), *"small"*, or *"large"* depending on its size, and *"random-distribution"*, *"structured"*, *"stripped*", *"chequered"*, depending on how they are spatially organized.

Goal 7. Annotation server development. The algorithms derived from the previous goals must be organised on a server, ready to be applied on any given image. This goal was essentially foreseen to be made along this year, is for this reason we will not show any result on it in the next section.

2 Degree of achievement

Considering the project goals, in the next subsections we will explain the results we have got on each one. All of them are presented in the frame of the four research lines.

2.1 Colour Constancy

To deal with **goal 1**, we have proposed a new colour constancy algorithm to describe the image illuminant. As most of the computer vision problems, colour constancy is an ill-posed problem, since multiple solutions can be derived from a given image, the essential contribution of our work has been to give a new algorithm to introduce some knowledge from likely colour names in the image to constraint the problem. Hence we have posed the *Nameability assumption*, as a high level constraint to weight all the solutions within the feasible set. This have made us to propose a new algorithm of colour constancy based on nameability that has been made in the frame of a PhD thesis, presented by F. Tous on July 2006 [4]. Partial results of this work have been published in [1], [2] and still in reviewing process we have [3].

The last part of this work that is still not finished are the tasks involved in **goal 4**. Some steps have already been done in this direction but are not finished yet. To discover the most feasible illuminants has been dealt by computing the ridges of the weighted feasible set. The algorithm of ridges detection on weighted distributions has been proven in [27,28] as a powerful technique to reduce colour information. This has provoked the initiation in the frame of this project of a new research line that will be the basis of the Master thesis work of E. Vázquez [29], who has been incorporated to this research project.

2.2 Perceptual representation of colour

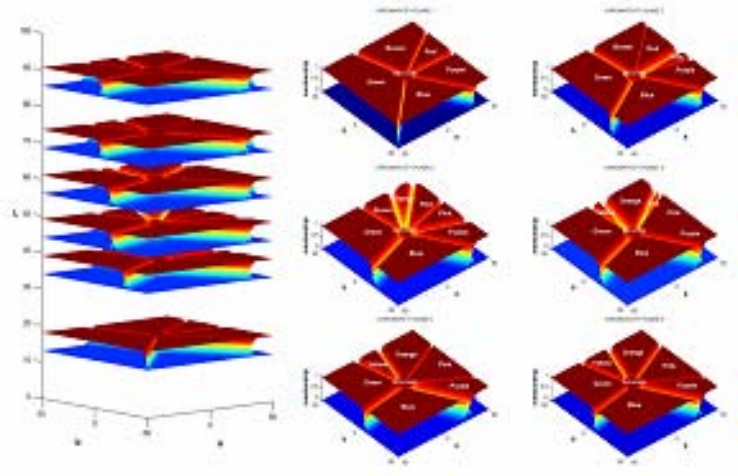
Goal 2 has made us to continue with the research line of building the perceived image that is a key goal of our research group [5-8]. The essential contribution in this goal has been to propose a complete multiresolution wavelet model that is predicting the behaviour of the human visual system induction mechanisms [9-12]. The development of this model has implied more time than

was previewed since it is based in an important background knowledge coming from Psychophysics.

This obligates us to go deeper in this new discipline that was far from the knowledge background of this group. To solve this problem and thanks to this project, we have been able to join a new member to the group from this expertise area. The new member is a specialist in psychophysics but with an strong background on computer science. We have received the help from the Juan de la Cierva program to incorporate Dr. Alejandro Párraga [13] to our group. He has made us to make interesting progress on this problem, being able to demonstrate the model from real psychophysical data but, results of this work are still under revision [14]. Thanks to Dr. Párraga, we are being able to build a psychophysics laboratory in order to be able to plan specific experiments we need for our research. This will allow increasing the level of multidisciplinary quality of the group research.

2.3 Colour Naming

The problem of reaching an automatic colour naming task was initiated in the group before to start with this project [15-17]. We have continued in this line, **goal 5**, finishing a full colour naming model that, under standard conditions, allow to automatically assigning a fuzzy colour name to any colour represented by its three-dimensional coordinates. The essential contribution has been to propose a parametric model based on a triple sigmoid function fitted on psychophysical data [18]. The final model is still under a revision process [19], and has been done in the work of the PhD Thesis of Robert Benavente [19], whose presentation is foreseen in March-2007. In the following figure we show the obtained model by means of the membership function of the naming sets. We hope it can have an important impact in colour research either from a computational or a psychophysical point of view.



2.4 Texture Description and Grouping

Texture description is the hardest goal of this project; it was summed up in **goal 6**. Some work was already done before to begin this project by our group [21,22]; we proposed a texture description based on the subtexture components. By subtexture we mean, those sets of image blobs sharing an attribute. Hence, a texture description can be given by a set of subtexture descriptions, each one is easily described in terms of its blob attributes. Hence, a texture could be explained by two subtexture descriptions, given for instance by *"large elongated 45º-oriented blobs"* and *"small round blobs"*. These subtexture descriptions can be the basis of any high-level description of any texture.

In the frame of this project we have proposed a perceptual blob space that allows defining a computational algorithm to detect subtexture descriptions, this has been done in the frame of the PhD thesis of A. Salvatella [23], whose presentation is foreseen in June-2007. The work of extending this description to colour blobs and its application to an image indexing problem with good results has been compiled in a paper that is still under revision [24].

The extension of colour blobs an a grouping algorithm of blobs is the basis to deal with **goal 3**. A first algorithm to group image blobs has been proposed in [25], where the inter-feature distance map is used as the basic feature for an unsupervised clustering. This last work is being done in the frame of the PhD thesis of S. Álvarez [24], whose presentation is foreseen in December-2007.

3 Results indicators

As an essential indicator of the results in this project we can see the consolidation of the colour-texture research group by joining some new researchers:

- The project has allowed incorporating a Post-Doc Researcher thanks to the Juan de la Cierva Program, A. Párraga. He introduces a multidisciplinary value to the group due to his background on psychophysics. He is leading the organisation of a psychophysics lab for our group.

- Two new PhD students have also been joined to the group, E. Vázquez and J. Vázquez, who are Assistant professors of the Computer Science Department of the Universitat Autònoma de Barcelona.

The dissemination of the project results are most of them under a revision process and have been explained and cited in previous sections. They can be summed up in the following list:

- 1 Accepted paper in a JCR journal [18].
- 3 Papers to JCR journals under revision [3, 14, 19]
- 1 CVC Technical report [10]
- 2 Contributions in international conferences with abstract in JCR journal [11,12].
- 1 Phd Thesis [4] (with European mention)
- 3 Phd Thesis whose finalisation is foreseen in this year [20, 23, 26]
- 1 Master Thesis whose finalisation is foreseen in this year [29]

- 2 book chapters derived from two conference presentations [25, 27]
- 1 submission for an international conference [28]

We have established new international relationships with other groups:

- University of Amsterdam, Prof. Gevers.
- University of Bristol, Prof. Troscianko.

This relationship will allow the collaboration within the CVC group in the VIDI-Video project (*Interactive semantic video search with a large thesaurus of machine-learned audio-visual concepts*) funded by the 7th European Framework program, and leadered by the University of Amsterdam. The collaboration of our group in this project will essentially be based on the automatisation of the colour naming tasks.

We also have maintained the usual collaboration with Prof. Graham Finlayson of the Norwich University, who has received the PhD student, F. Tous, in two different research stays.

As technological transfer activities we want to highlight two. On one hand a meeting with an industrial sector of wood and furniture (with different representative of this sectorial area), who are interested in the transfer of the model of perceptual representation of texture and colour. On the second hand, we want to see the transfer of the know-how on this background from our personnel, since Dr. F. Tous, will initiate a professional stay at the spin-off of the University of Norwich, leadered by Prof. G. Finlayson. This start-up is devoted to develop colour software for image enhancement.

4 References

[1] F. Tous, M. Vanrell, and R. Baldrich. *Exploring Colour Constancy Solutions*. Proceedings of the Second European Conference on Colour in Graphics, Imaging, and Vision (CGIV'2004), 1:24-29, Aachen (Germany), April 5-8, 2004.

[2] F. Tous, M. Vanrell, and R. Baldrich. *Relaxed Grey-World: Computational Colour Constancy by Surface Matching*. Second Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'2005), Lecture Notes in Computer Science 3522:192-199, Estoril (Portugal), June 7-9, 2005.

[3] F. Tous, M. Vanrell, and R. Baldrich. *Computational white point interpretation model based on nameability assumption*. Image and Vision Computing, 2006. (En proceso de revisión).

[4] F. Tous. *Computational framework for the white point interpretation based on colour nameability*. Phd Thesis. Universitat Autònoma de Barcelona, 2006. Directores: M. Vanrell y R. Baldrich.

[5] R. Baldrich. Perceptual approach to a computational colour texture represetnation for surface inspection. Phd Thesis. Universitat Autònoma de Barcelona, 2001. Directora: M. Vanrell.

[6] M. Vanrell, R. Baldrich, A. Salvatella, R. Benavente, and F. Tous. *Induction operators for a computational colour-texture representation*. Computer Vision and Image Understanding, 94(1-3):92-114, 2004.

[7] R. Baldrich, M. Vanrell, R. Benavente, and A. Salvatella. *Color Enhancement based on perceptual sharpening*. Proceedings of the IEEE International Conference on Image Processing (ICIP'2003), 3:401-404, Barcelona, September 14-17, 2003.

[8] X. Otazu and M. Vanrell. *Building perceived colour images*. Proceedings of the Second European Conference on Color in Graphics, Imaging, and Vision-CGIV'04, 1:140-145, Aachen (Germany), April 2004.

[9] X. Otazu and M. Vanrell. *Perceptual Representation of Textured Images*. Journal of the Imaging Science and Technology, 49(3):262-271, 2005.

[10] X. Otazu and M. Vanrell. Decimated multiresolution framework with surround-induction function to unify assimilation and contrast mechanisms. Computer Vision Center Technical Report #99, 2005.

[11] X. Otazu and M. Vanrell. A surround-induction function to unify assimilation and contrast in a computational model of color apearance. Perception, 34-suppl s:215, 2005.

[12] X. Otazu and M. Vanrell. Several lightness induction effects with a computational multiresolution wavelet framework. Perception, 35-suppl s:56, 2006.

[13] P. G. Lovell, D.J. Tolhurst, C.A. Párraga, R. Baddeley, U. Leonards, J. Troscianko, et al. *On the contributions of the "red-green" and "blue-yellow" color-opponent systems to encoding the information in natural scenes*. Journal of the Optical Society of America A, 22(10):2060-2071, 2005.

[14] X.Otazu, M. Vanrell, and A. Parraga. A multiresolution wavelet framework for the account of several brightness induction effects. Vision Research, 2006. (En proceso de revisión).

[15] R. Benavente, F. Tous, R. Baldrich, and M. Vanrell. *Statistical Modelling of a Colour Naming Space*. Proceedings of the First European Conference on Color in Graphics, Imaging and Vision (CGIV'02), 1:406 - 411, Poitiers (France), April 2-5, 2002.

[16] R. Benavente and M. Vanrell. *Fuzzy colour naming based on sigmoid membership functions*. Proc. 2nd Euro. Conference on Color in Graphics, Imaging and Vision-CGIV'04, 1:135-139, Aachen (Germany), April 2004.

[17] R. Benavente, M. Vanrell, and R. Baldrich. *Estimation of Fuzzy Sets for Computational Colour Categorization*. Color Research and Application, 29(5): 342-353, 2004.

[18] R. Benavente, M. Vanrell, and R. Baldrich. *A Data Set for Fuzzy Colour Naming*. Color Research and Application, 31(1):48-56, 2006.

[19] R. Benavente, M. Vanrell, and R. Baldrich. *Parametric Fuzzy Sets for Automatic Colour Naming*. Pattern Recognition, 2006. (En proceso de revisión).

[20] R. Benavente. *A Computational Model of Color Naming for Digital Images*. Phd Thesis. Universitat Autònoma de Barcelona. (Fecha prevista de finalización: marzo 2007). Directora: M. Vanrell.

[21] A. Salvatella, M. Vanrell, and J.J. Villanueva. *Texture Description based on Subtexture Components*. Proc. of the 3rd International Workshop in Texture Analysis and Synthesis, 1:77-82, Nice (France), October 17, 2003.

[22] A. Salvatella, M. Vanrell, and R. Baldrich. *Subtexture Components for Texture Description*. 1st Iberian Conf. on Pattern Recognition and Image Analysis-IbPRIA'2003, LNCS 2652:884-892, Mallorca, June 4-6, 2003.

[23] A. Salvatella. *Texture description based on a blob-emergent pattern approach*. Phd Thesis. Universitat Autònoma de Barcelona, (Fecha prevista de finalización: junio 2007). Directora: M. Vanrell.

[24] A. Salvatella, M. Vanrell, S. Álvarez, and X. Otazu. *Image similarity based on blob detection and grouping*. IEEE Transactions on Pattern Analysis and Machine Intelligence (En proceso de revisión).

[25] S. Alvarez, X. Otazu, and M. Vanrell. *Image Segmentation Based on Inter-Feature Distance Map*. Frontiers in Artificial Intelligence and Applications, IOS Press, 131:75-82, 2005.

[26] S. Alvarez. *Perceptual grouping of colour blobs*. Phd Thesis. Universitat Autònoma de Barcelona. (Fecha prevista de finalización: diciembre 2007). Directores: M. Vanrell y X. Otazu.

[27] E. Vazquez, F. Tous, R. Baldrich, and M. Vanrell. *N-dimensional distribution reduction preserving its structure*. Frontiers in Artificial Intelligence and Applications, IOS Press, 146:167-175, 2006.

[28] E. Vazquez and R. Baldrich. *Colour image segmentation using ridge extraction*. Third Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'2007). (En proceso de revisión).

[29] E. Vazquez. *Extraction of main structures in n-dimensional distributions*. Application to colour image segmentation. Master Thesis. Universitat Autònoma de Barcelona. (Fecha prevista de finalización: febrero 2007). Director: R. Baldrich.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Microarchitectures and Compilers for Future Nanotechnologies TIN2004-03072

Antonio González*
Intel-UPC Barcelona Research Center
Universitat Politècnica de Catalunya

Abstract

The increasing performance of processors during the last decades has been mainly determined by two factors: technological advances in the manufacturing of microprocessors and the use of new or more efficient microarchitectural and compiler techniques. However, recent studies show that these advances will face new challenges in future generation microprocessors: high power consumption, heat dissipation, wire delays, the complexity in the design and the limited instruction-level parallelism. The main objective of this project is the research in the design of microprocessors for the next decade, taking into account the key characteristics of future process technologies. With that goal, new mechanisms will be proposed from the point of view of the microarchitecture as well as the necessary support from the compiler. This project will deal with four areas that we consider fundamental for future processors: (i) reduction in power consumption and better approaches for heat dissipation, (ii) multithreaded processors with the aim of using more efficiently the resources in the system, (iii) speculative multithreaded processors in order to accelerate the execution of irregular codes, and (iv) reliable processors that can recover from faults during its operation.

Keywords: Future processors, thread-level parallelism, power-efficiency, reliable systems.

1 Project Goals

The main objective of the project consists of proposing microarchitectural solutions for future generation microprocessors. As it was pointed out in the project proposal, four main research areas have been identified as crucial for future processors: (i) power consumption, (ii) multithreaded processors, (iii) speculative thread-level parallelism, and, (iv) reliability. The specific objectives for each of them are enumerated in this section.

1.1. Reduction in Power Dissipation and Control of the Temperature

- Investigate new adaptive algorithms for the active units in clustered processors with the aim of minimizing the energy consumed by the processor.
- Reduce power dissipation through new schemes for cluster frequency and voltage regulation combined with intelligent policies for distributing instructions in order to minimize performance loss.
- Reduce the static power consumption by controlling the temperature and the design of circuits with multiple threshold voltages and/or dynamic threshold voltage regulation.

* Email: antonio@ac.upc.edu

- Propose techniques to control the maximum temperature and reduce the average temperature of the processor.

1.2. Clustered Multithreaded Processors

- Design power-effective clustered multithreaded architectures.
- Perform a bottleneck analysis and propose hardware solutions to reduce their impact.
- Propose new compilation techniques to improve performance in the architecture.

1.3. Speculative Multithreaded Processors

- Implement a simulator of a speculative multithreaded architecture and propose new designs for the register file and the memory hierarchy capable of dealing with multiple independent versions for each control thread.
- Design novel compiler approaches for partitioning a source code into speculative threads and evaluate alternative partitioning heuristics.
- Performance evaluation of the various architectures and compiler techniques developed in this project.

1.4. Reliability

- Develop a tool to simulate the injection of errors at a very fine grain during the operation of a processor.
- Propose new mechanisms for fault detection.
- Propose new techniques to recover the correct state of the processor and evaluate them in terms of performance and power.

2 Level of Success Reached in the Project

Most of the objectives proposed have already been accomplished during these two first years after the beginning of this research project. Some others are still ongoing and some others have been changed since the partial results suggested low potential and were replaced by new ones. The individual level of success for each of the different research areas is presented below.

2.1. Reduction in Power Dissipation and Control of the Temperature

In the power and thermals area, we have been working on both single and multi core architectures. In single core, we have proposed a clustered multiclock domain processor. The microarchitecture is divided in three different clock domains (each one of them having its own voltage regulator): front-end, back-end and L2 cache. Then, the frequency/voltage of each domain is dynamically changed in order to minimize energy-delay^2 product [21]. Regarding temperature, we have also split some of the hot structures of the front-end, in order to distribute the activity and therefore reduce power density. In particular, the rename table and the reorder buffer were split into two different structures, each one of them feeding one of the backend clusters. In addition, the trace cache was divided into different banks, and a bank hoping technique was developed. The temperature reduction achieved in each of these banks is around 30%, with just 2% performance loss [46].

Cache memories dissipate high power due to their high activity (dynamic power) and large area (leakage power). Based on the observation that all cache ways in conventional set-associative caches have the same size but their usage is different, we have proposed the heterogeneous way size (HWS) cache. HWS caches fit much better program requirements requiring lower cache capacity, thus saving both dynamic and leakage power[24].

A very large part of the dynamic power of conventional cores is devoted to the scheduling logic, which comprises the scheduler, the register file and the load/store queue. We have proposed new techniques to reduce the power consumption of those structures. A novel compiler-based technique has been developed to decrease the number of in-flight instructions so that the activity in the scheduler and the register file decreases and significant power is saved[19][45]. To further reduce the energy consumption of the register file we have proposed several compiler-based techniques to release registers as soon as possible[19][34]. Finally, a new design for load/store queues has been proposed based on the observation that in-flight memory instructions access few cache lines, and thus, instructions can be grouped to reduce the number of address comparisons and to reduce the activity in the L1 data cache and the data translation look-aside buffer[28].

In the multi-core area, we have proposed a technique to reduce the power dissipated in the bus by using narrow busses and encoding the most significant bits of the addresses that go through the bus, by taking advantage of the address locality. Most significant bits are encoded using less information and each core is augmented with a table used to translate this new encoding.

Furthermore, we have investigated the combination of thread migration and dynamic voltage scaling to reduce the performance penalties due to the fact that any of the cores has reached the thermal emergency limit and it has to be shut down. Thread migration is implemented by means of a novel scheme that, at every time interval, sorts the cores according to their respective temperature and sorts the current threads according to the power dissipated in the last interval. Then, high power threads are sent to cold cores and cold threads are sent to hot cores. This algorithm is combined with per-core dynamic voltage and frequency scaling.

2.2. Clustered Multithreaded Processors

In this research area, we have proposed a clustered microarchitecture and we have focused on proposing techniques to enhance it through novel schemes to steer instructions to the clusters and improving the memory bandwidth of the processor. Different steering techniques have been explored for multithreaded workloads and new schemes have been developed that improve the state of the art[16]. Besides, to reduce the complexity of the steering logic may become complex, a hybrid software/hardware scheme has also been proposed to perform instruction steering for single threaded workloads. The proposed scheme achieves a performance very close to the best hardware-only proposal existing in the literature. Finally, memory bandwidth has been enhanced by implementing interleaved cache banking[32]. The proposed cache organization does not require data replication and allows splitting the process of disambiguating memory accesses. This scheme assigns a first-level data cache bank per cluster and uses a bank predictor to optimize the instruction steering, reduce inter-cluster communications and improve the scheduling of resources.

We have also paid special attention to the branch prediction with the aim of improving the efficiency of the fetch engine. Fetch efficiency, and therefore the ability of exploiting parallelism by a processor, directly depends on the accuracy of the branch predictor. “If-conversion” is a compilation technique that turns control dependences into data dependences by using predication. However, predication incurs in two important drawbacks when it is implemented in an out of order processor: multiple definitions of renaming registers, and waste of resources when executing instructions whose predicate becomes false. These two drawbacks have been faced by proposing a novel technique based on predicate prediction for out of order processors[22]. Since predicates are known (predicted) at renaming time, this technique solves the problem of having multiple definitions of renaming registers and it also eliminates those instructions whose predicate is predicted as false, reducing indeed energy consumption. However, “if-conversion” also reduces the

available information for correlation among branches, reducing the accuracy of the branch predictor as a side effect. This problem has been alleviated by proposing a predictor able to predict predicates as well as branches[20]. Therefore, this predictor contains all the information regarding branch correlation without additional hardware complexity compared to a conventional predictor.

Conversely to high performance computing, instructions executed in embedded processors and DSPs consisting of multiple clusters are statically scheduled with the aim of simplifying the core design, lowering manufacturing costs and shortening their time to market. In this case, the compiler is in charge of the instruction scheduling, register allocation and workload balance among clusters. In this area, a novel scheme to assign registers has been proposed that reduces spill code for cyclic code, significantly improving the current state of the art[17][39]. Besides, a novel technique to schedule instructions and perform workload balance among clusters for general purpose code in a unified step has been developed. The novelty of this technique comes from a powerful deduction process that identifies the consequences that certain scheduling decisions may have on performance[18]. This process allows a wider knowledge of the impact some scheduling decisions which revert in an improved schedule.

Finally, energy consumption is an important constraint in embedded processors. One of the main sources of energy consumption is cache memory. For this reason, partitioned cache memory schemes have been studied where each cache module is located into a different cluster. Moreover, each of these modules can be configured either as a fast-access and high-power component or as a slow-access low-power module[33]. It has been shown that an effective use of these heterogeneous partitioned modules can be reached by means of new compilation techniques.

2.3. Speculative Multithreaded Processors

In the last two years, we achieved several milestones in the task of speculative multithreading (SpecMT), both in the areas of compilation and microarchitecture design. We have developed mechanisms and tools in two different ISA (Intel and Alpha), although we most of the techniques are ISA agnostic.

On one hand, we have investigated on a compiler/architecture combined scheme to perform SpecMT based on pre-computation slices to deal with inter-thread dependences. On the compiler side, we have investigated mechanisms to handled inter-thread dependences in general regions of code selected though an analysis of the control independence of these regions. For the dependences, the objective in to find the best trade-off between the overhead and the accuracy of the technique used. We have implemented an analysis on top of the ORC compiler which is able to analyze the dependences for selected regions of a program and to pick the most effective mechanism to predict these values. Our core technique is based on pre-computation slices that are able to quickly compute the live-ins values of a new spawned thread. We have proposed several mechanisms to reduce the overhead of this slice by means of speculative techniques such as branch and dependence pruning[31][41].

In addition to that, we have developed a trace analyzer tool based on Atom that analyses the dynamic trace collected from the application for selecting loop-boundaries where threads can be spawned. This tool may be used to guide the compiler to select the spawning points and to detect the live-ins among concurrent threads.

On the side of the architecture, we have defined a new microarchitecture and developed a simulator of it for the proposed SpecMT scheme based on pre-computation slices. Due to the lack of a simulator infrastructure supporting SpecMT available in the public domain, the first step has been to develop an infrastructure supporting SpecMT execution. The SMTSIM from University of

California, San Diego was used as a starting point and was modified to support SpecMT execution. The architecture is inspired in a multi-core architecture, where each core executes a thread. For that, we have defined the organization of the register file and the memory hierarchy in order to support multiple versions and detect inter-thread register/memory violations[36]. Memory dependences are tracked at a fine granularity to avoid unnecessary squashes due to false sharing. We have implemented mechanisms that are completely invisible to the software, whereas some others simplify the hardware but require additional analysis of the compiler. For instance, we have proposed schemes in which all the register dependences are completely handled by the software whereas the memory dependences remain checked by the hardware.

2.4. Reliability

Parameter variation in scaled technologies beyond 65nm will pose a major challenge for design of future high-performance microprocessors. It is very important to accurately estimate the impact of parameter fluctuations on circuit performance; an overestimation increases the design complexity and may compromise performance, whereas an underestimation can compromise the reliability and yield. Our work has focused on two main areas: (i) modeling: we have built a statistical model [4] for the impact of different sources of variations such as process, T, voltage and input vectors, (ii) X-Pipe [25]: we have designed a fault detection and recovery scheme that allows removing most of the safety margins designed for in-order and out-of-order processors.

On the other hand, we have exploited the fact that many values in the datapath can be represented by a lower number of bits than the full machine width. We have designed an 8-bit Helper Cluster [27] that complements a 32-bit monolithic processor with a low-complexity 8-bit cluster, which allows obtaining important performance benefits.

Current microprocessors are becoming more vulnerable to cosmic particle strikes, which may cause soft (transient) errors. Our current research focuses on microarchitectural techniques that allow decreasing the vulnerability of the processor to particle strikes, and detect errors. For reducing vulnerability, we have proposed two different approaches: (i) identifying narrow values in data storage instructions [3], and (ii), invalidating cache lines when they hold data not likely to be reused in the short term.

While techniques to reduce vulnerability are very useful, they cannot avoid errors. We consider that error mitigation techniques are still important. However, any technique that detects errors should employ some kind of redundancy, either temporal or spatial. Temporal redundancy impacts performance, whereas spatial redundancy increases area and complexity. So far, we have evaluated two possible directions: full replication (similar to Simultaneous Redundant Threading - SRT) in a clustered environment [29] and optimizations that do not replicate all instructions in order to reduce the power and execution time overhead.

As technology evolves, the size (length, width and thickness) of the transistors and wires shrinks. This size reduction translates into higher vulnerability of the circuits, which degrade faster and are more likely to experience failures. Current solutions like burn-in or testing may not be useful for avoiding such errors.

We are working towards a complete platform that is aware of processor degradation. We call it Canigo, a Wearout-Aware Reliable Platform, which has the ability to keep operating at full-throttle on the face of errors. Canigo incorporates a set of techniques that provide an accurate estimate of the state of the system, and a strategy to diagnose and reconfigure the cores. In particular, we have proposed (i) a technique (fuses) to detect failures in circuits before they appear, (ii) a model (mileage) to measure block degradation, (iii) core servicing, a diagnosis that checks and

reconfigures the cores built on top of fuses and mileage, and (iv) the interface to the OS so that it can steer tasks in a multi-core environment in such a way lifetime and performance are maximized.

On the other hand, we are also working on microarchitectural techniques to mitigate the degradation caused by the different sources of failures, such as electromigration in bidirectional buses and in power/ground grids, and NBTI (Negative Biased Temperature Instability). This phenomenon happens in PMOS transistors whenever the voltage at the gate is negative. In order to mitigate the impact of this degradation, we have proposed a new design for memory cells that significantly reduces the NBTI degradation.

3 Result indicators

3.1. Journal Publications

[1] R. Ronnen and A. González, "Micro's Top Picks from Microarchitecture Conferences", IEEE Micro. Accepted for publication in forthcoming issues.

[2] A. Settle, D. Connors, E. Gibert and A. González, "A Dynamically Reconfigurable Cache for Multithreaded Processors", Journal of Embedded Computing, special issue on Embedded Single-Chip Multicore Architectures and Related Research, from System Design to Application Support. Accepted for publication in forthcoming issues.

[3] O. Ergin, O. Unsal, X. Vera and A. González, "Exploiting Narrow Values for Soft Error Tolerance", IEEE Computer Architecture Letters, Accepted for publication in forthcoming issues.

[4] O. Unsal, J. Tschanz, X. Vera, O. Ergin, K. Bowman, V. De and A. González, "Impact of Parameter Variations on Circuits and Microarchitecture", IEEE Micro, vol. 26, no. 6, Nov.-Dec. 2006, pp. 30-39.

[5] E.Gibert, J.Sánchez and A.González, "Instruction Scheduling for a Clustered VLIW Processor with a Word-Interleaved Cache", Concurrency and Computation: Practice and Experience, vol. 18,no. 11,Sept. 2006,pp. 1391-1412.

[6] A. González, "Multi-core & Multi-threaded: A New Era in Computing - Helping advance a new era in computing through new methods of thread-level parallelization", Dr. Dobb's Journal (electronic journal), May 10th, 2006. http://www.ddj.com/dept/architect/187201936.

[7] J. L. Aragón, J. González and A. González, "Control Speculation for Energy-Efficient Next-Generation Superscalar Processors", IEEE Transactions on Computers, vol. 55, no. 3, March 2006, pp. 281-291.

[8] A. González, "Speculative Threading: Creating New Methods of Thread-Level Parallelization", Technology@Intel Magazine, Intel Corporation , vol. 3, no. 11, December 2005-January 2006, pp. 3-10.

[9] X. Vera, J. Llosa and A. González, "Near-Optimal Padding for Removing Conflict Misses", Lecture Notes in Computer Science, vol. 2481/2005, 2005, pp. 329-343.

[10] T. Monreal, V. Viñals, A. González and M. Valero, "Hardware Support for Early Register Release", Int. Journal of High Performance Computing and Networking (IJHPCN), vol. 3, no. 2/3, 2005, pp. 83-94.

[11] E. Gibert, J. Sánchez and A. González, "Distributed Data Cache for Designs for Clustered VLIW Processors", IEEE Transactions on Computers, vol. 54, no. 10, October 2005, pp. 1227-1241.

[12] X. Vera, J. Abella, J. Llosa and A. González, "An Accurate Cost Model for Guiding Data Locality Transformations", ACM Transactions on Programming Languages and Systems (TOPLAS), vol. 27, no. 5, September 2005, pp. 946-987.

[13] R. Canal, A. González and James E. Smith, "Value Compression for Efficient Computation", Lecture Notes in Computer Science, vol. 3648/2005, August 2005, pp. 519-529.

[14] A. Pajuelo, A. González and Mateo Valero, "Speculative Execution for Hiding Memory Latency", ACM SIGARCH Computer Architecture News, vol. 3, issue 3, June 2005, pp. 49-56.

[15] J. Abella, A. González, X. Vera and M. O'Boyle, "IATAC: a Smart Predictor to Turn-off L2 Cache Lines", ACM Transactions on Architecture and Code Optimization, vol. 2, issue 1, March 2005, pp. 55-77.

[16] J.M.Parcerisa, J.Sahuquillo, A.González and J.Duato, "On-Chip Interconnects and Instruction Steering Schemes for Clustered Microarchitectures", Transactions on Parallel and Distributed Systems, vol. 16,no 2,Feb. 2005,pp. 130-144.

3.2. Symposia Publications

[17] A. Aletà, J.M. Codina, A. González and D. Kaeli, "Heterogeneous Clustered VLIW Microarchitectures", the 2007 Int. Symp. on Code Generation and Optimization, San Jose, CA (USA), March 11-14, 2007, to appear

[18] J.M. Codina, J. Sánchez and A. González, "Virtual Cluster Scheduling Through the Scheduling Graph", the 2007 Int. Symp. on Code Generation and Optimization, San Jose, CA (USA), March 11-14, 2007, to appear

[19] T. Jones, M. O'Boyle, J. Abella, A. González and O. Ergin, "Designing Efficient Processors Using Compiler-Directed Optimisations", Proc. of the 11th Ann. Workshop on the Interaction between Compilers and Computer Architecture, Phoenix, AZ (USA), Feb. 10th, 2007, to appear

[20] E. Quiñones, J.M. Parcerisa and A. González, "Improving Branco Prediction and Predicated Execution in Out-Of-Order Processors", Proc. of the 13th Int. Symp. on High-Performance Computer Architecture Phoenix, AZ (USA), February 10-14, 2007, to appear

[21] G. Magklis, P. Chaparro, J. González and A. González, "Independent Front-end and Back-end Dynamic Voltaje Scaling for a GALS Microarchitecture", Proc. of the Int. Symp. on Low Power Electronics and Design, Tegernsee (Germany), October 4-6, 2006, pp. 49-54.

[22] E. Quiñones, J.M. Parcerisa and A. González, "Selective Predicate Prediction for Out-Of-Order Processors", Proc. of the 20th ACM Int. Conference on Supercomputing, Cairns (Australia), June 28-July 1, 2006, pp. 46-54.

[23] M. Monchiero, R. Canal and A. González, "Design Space Exploration for Multicore Architectures: A Power/Performance/Thermal View", Proc. of the 20th ACM Int. Conference on Supercomputing, Cairns (Australia), June 28-July 1, 2006, pp. 177-186.

[24] J. Abella and A. González, "Heterogeneous Way-Size Cache", Proc. of the 20th ACM Int. Conference on Supercomputing, Cairns (Australia), June 28-July 1, 2006, pp. 239-248.

[25] X. Vera, O. Unsal and A. González, "X-Pipe: An Adaptive Resilient Microarchitecture for Parameter Variations", Proc. of the Workshop on Architectural Support on Gigascale Integration, Boston, MA (USA), June 17, 2006.

[26] P. Chaparro, J. González, G. Magklis and A. González, "Using MCD-DVS for dynamic thermal management performance improvement", Proc. of the ITHERM 2006, San Diego, CA (USA), May 30-June 2, 2006, pp. 140-146.

[27] O. Unsal, O. Ergin, X. Vera and A. González, "Empowering a Helper Cluster through Data-Width Aware Instruction Selection Policies", Proc. of the 20th Int. Parallel and Distributed Processing Symp., Rodhes (Greece), Apri.l 25-29, 2006, pp. 113-114.

[28] J. Abella and A. González, "SAMIE-LQS: Set-Associative Multiple-Instruction Entry Load/Store Queue", Proc. of the 20th Int. Parallel and Distributed Processing Symp., Rodhes (Greece), April 25-29, 2006, p. 52.

[29] X. Vera, J. Abella, O. Unsal, A. González, O. Ergin, "Checker Backend for Soft and Timing Error Detection and Recovery", Proc. of the 2nd Workshop on System Effects of Logic Soft Errors, Urbana, IL (USA), Apr. 11-12, 2006.

[30] E. Gibert, J. Abella, J. Sánchez, X. Vera and A. González, "A Heterogeneous Multi-Module Data Cache for VLIW Processors", Proc. of the Explicitly Parallel Instruction Computing, Manhattan, NY (USA), March 26, 2006.

[31] C. García, C. Madriles, J. Sánchez, P. Marcuello and A. González, "MITOSIS: Speculative Parallelization based on Pre-Computation Slices", Proc. of the 12th Int. Workshop on Compilers for Parallel Computers, A Coruña (Spain), Jan. 9-11, 2006, pp. 341-355.

[32] S. Bieschewski, J.M. Parcerisa and A. González, "Memory Bank Predictors", Proc. of the IEEE Int. Conference of Computer Design, San Jose, CA (USA), October 2-5, 2005, pp. 666-670.

[33] E. Gibert, J. Abella, J. Sanchez, X. Vera and A. González, "Variable-Based Multi-Module Data Caches for Clustered VLIW Processors", Proc. of the 14th Int. Conference on Parallel Architectures and Compilation, St Louis (USA), Sept 17-21, 2005, pp. 207-217.

[34] T. M.Jones, M.F.P.O'Boyle, J.Abella, A.Gonzalez and O.Ergin, "Compiler Directed Early Register Release", Proc. of 14th Int. Conf. on Parallel Architectures and Compilation Techniques, St. Louis (USA), Sept 17-21, 2005, pp. 110-119.

[35] A. González, "The Right-Hand Turn to Multi-Core Processors", Parallel Computing Málaga (Spain), September Sept. 13-16, 2005. **Keynote.**

[36] C. Madriles, C. García Quiñones, J. Sánchez, P. Marcuello and A. González, "The Mitosis Speculative Multithreaded Architecture", Proc. of Parallel Computing 2005, Málaga (Spain), Sept. 13-16, 2005, pp. 27-38.

[37] C. Molina, J. Tubella and A. González, "Reducing Misspeculation Penalties in Trace-Level Speculative Multithreaded Architectures", Proc. of the the 6th. Int. Symp. on High Performance Computing, Nara (Japan), September 7-9, 2005.

[38] R. Canal, A. González, James E. Smith, "Value Compression for Efficient Computation", Proc. of the Int. Conference on Parallel and Distributed Computing, Lisbon (Portugal), Aug. 30th, Sept. 2nd, 2005, pp. 519-529.

[39] A. Aletà, J. M. Codina, A. González and D. Kaeli, "Demystifying on-the-fly spill code", Proc. of the ACM SIGPLAN Conf. on Programming Language Design and Implementation, Illinois (USA), June 12-15, 2005, pp. 180-189.

[40] A. González, "Multi-Core Chips. The Next Wave of Processor Microarchitecture", International Conference on Parallel Processing (ICPP 2005), Oslo (Norway), June 14-17, 2005. **Keynote**.

[41] C. García, C. Madriles, J. Sánchez, P. Marcuello, A. González and D. Tullsen, "Mitosis Compiler: An Infrastructure for Speculative Threading using Pre-Computation", Proc. of the ACM SIGPLANConf. on Programming Language Design and Implementation, June 12-15, 2005, pp. 269-279.

[42] A. Pajuelo, A. González and M. Valero, "Control-Flow Independence Reuse Via Dynamic Vectorization", Proc. of the 19th Int. Parallel and Distributed Processing Syposium, Denver, CO (USA), April 4-8, 2005, p. 21.

[43] J. Abella, A. González, "Inherently Workload-Balanced Clustered Microarchitecture", Proc. of the 19th Int. Parallel and Distributed Processing Syposium, Denver, CO (USA), April 4-8, 2005, p. 20.

[44] C. Molina, J. Tubella and A. González, "Compiler Analysis for Trace-Level Speculative Multithreaded Architectures", Proc. of the 9th Workshop on Interaction between Compilers and Computer Architectures, San Francisco, CA (USA), Feb. 12-16, 2005, pp. 2-10.

[45] T. Jones, M. O'Boyle, J. Abella and A. González, "Software Directed Issue Queue Power Reduction", Proc. of the 11th Int. Symp. on High Performance Computer Architecture, San Francisco, CA (USA), Feb. 2005 pp.144-153.

[46] P.Chaparro, G.Magklis, J.González and A.González, "Distributing the Frontend for Temperature Reduction", Proc. of 11th Int. Symp. on High Performance Computer Architecture, San Francisco, CA (USA), Feb, 2005, pp. 61-70.

3.3. Book Chapters

[47] Y. Sazeides, P. Marcuello, J. Smith and A. González, "Data Speculation", Speculative Execution in High Performance Computer Architectures, D. Kaeli and P-C. Yew editors, CRC Press, pp. 215-244, November 2005.

[48] P. Marcuello, J. Sánchez and A. González, "Multithreading and Speculation", Speculative Execution in High Performance Computer Architectures, D. Kaeli and P-C. Yew editors, CRC Press, pp. 333-354, November 2005.

3.4. PhD Students

During these two years, there were four PhD students (co-)advised by members of the team that successfully completed their PhD (Jaume Abella, Enric Gibert, Alex Pajuelo and Carlos Molina), and 14 more that are still working on it (Alex Aletà, Llorenç Cruz, Carlos Aliagas, Eduardo Quiñones, Stefan Bieschewski, Rakesh Ranjan, Indu Bhagat, Abhishek Deb, Marc Pons, Enric Herrero, Carlos Madriles, Fernando Latorre, Pedro López and Pedro Chaparro).

3.5. Patents

The members of the team have (co-)authored 16 patents that were filed during the last two years.

3.6. Technology Transfer

The research group has been extremely active in the area of technology transfer. Most of the members of the team have been working on a joint research lab of the Universitat Politècnica de Catalunya and Intel Corporation. The group has collaborated with various product groups in Intel in the design of future microprocessors and compilers.

3.7. Collaboration with Other Groups

In addition to multiple groups in Intel, the group has collaborated with other research groups in the following universities: Universidad Politécnica de Valencia, Univeristy of Wisconsin-Madison, University of Edinburgh, Northeastern University and Politecnico de Milano.

3.8. International Projects

The team has collaborated in an EPSRC project ("Low Power High Performance Microarchitecture and Compilation") led by the University of Edinburgh, and has a joint project with the University of Milano (Bi-lateral Action Program).

3.9. Other Indicators

The members of the team have participated in more than 25 program committees of symposia in the areas of computer architecture and compilers, have chaired the program committee for one of the top symposia in the area of computer architecture (Int. Symp. on Microarchitecture, 2004), have recently been nominated to chair another of them (Int. Symp. on High-Performance Computer Architecture, 2008), and have given more than 10 invited talks in academia and industry.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

U-CAT: Ubiquitous Collaborative Adaptive Training
TIN2004-03140

Pilar Rodriguez*

Abstract

The goal of the U-CAT project is the development of an integrated environment for the realization of educational activities through the use of different devices (personal computers, laptops, mobile phones and PDAs) in different contexts and situations. The aim is to adapt activities, tools and contents to the current context and the available devices, taking advantage of the possibilities offered by the use of ubiquitous computing to do both individual and collaborative teaching activities.

The U-CAT research group has formerly worked on adaptive hypermedia systems, collaborative learning, active environments and virtual laboratories. This work, which has given rise to several systems and tools, has been used as the starting point in the design and implementation of the U-CAT environment. With the aim of validating and checking the effectiveness of this adaptive environment, several teaching activities are being performed within it, in the Escuela Politécnica Superior of the Universidad Autónoma de Madrid (UAM). Classrooms and labs are being used as scenarios for the proposed activities, complementing both traditional lectures and practical work.

Keywords: intelligent environments, adaptive hypermedia systems, collaborative learning, context adaptation, quality evaluation, e-learning, mobile and ubiquitous computing, computer supported cooperative work / learning, network management

1 Project Goals

The U-CAT* project continues work from several prior projects, some examples are ENSENADA (TIC2001-0685-C02) focused on adaptive learning tools and environments, or INTERACT (TIC2000-0464), with emphasis on ambient intelligence, and several others (see project proposal). The goal of the project is the integration of recent technologies in the fields of Intelligent Environments, Ubiquitous Systems, Adaptive Hypermedia Systems, and Computer Supported Collaborative Work in the context of higher education. This implies

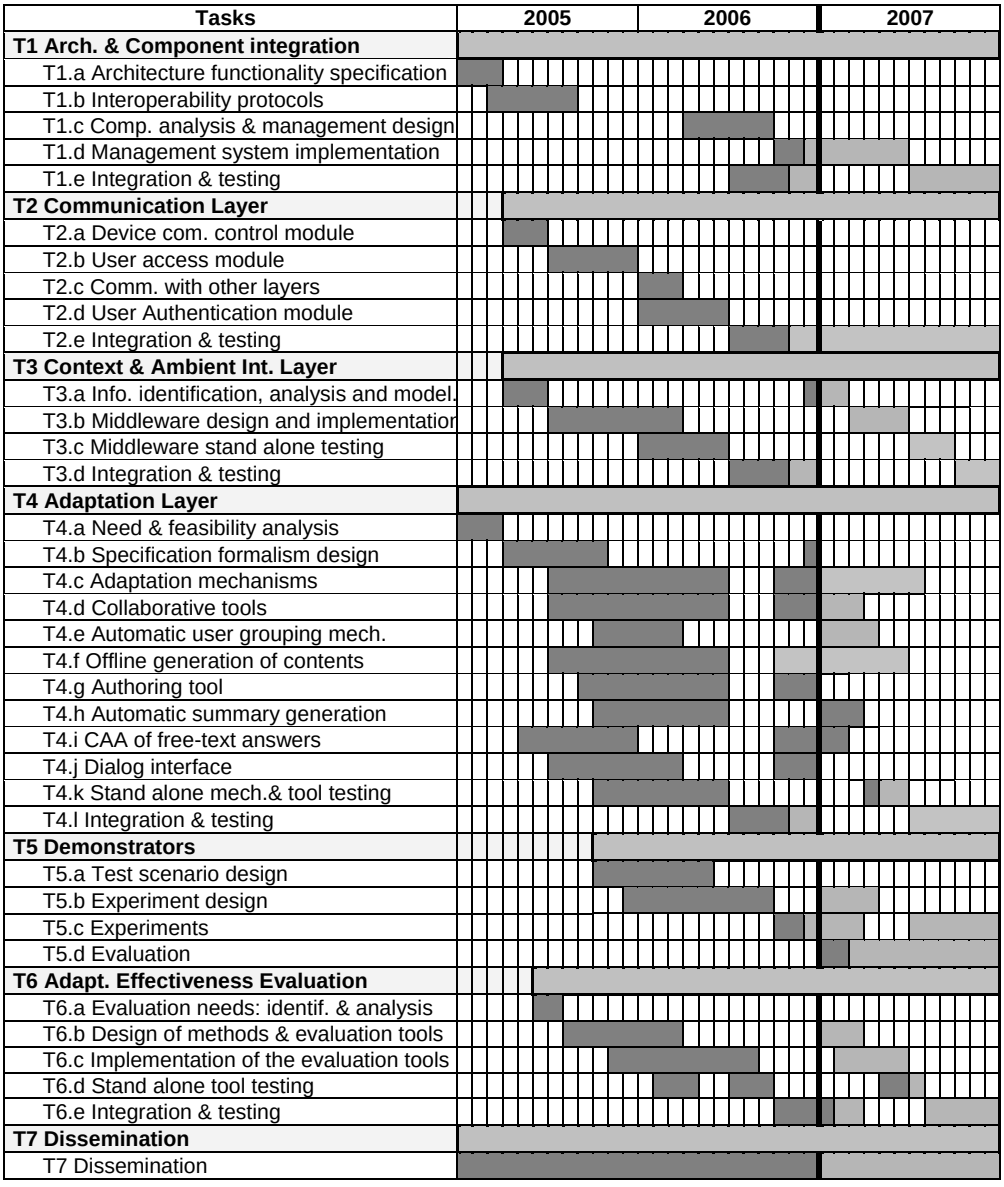
- An extension of the concept of adaptive learning systems to take into account "context" in the sense of so-called "ambient intelligence" systems.
- The application of Adaptive Hypermedia and CSCW user/group models and adaptation frameworks to problems present in traditional intelligent environments.

* Email: pilar.rodriguez@uam.es

* IN MEMORIAM Javier Martínez. He worked hard on the U-CAT proposal and made it possible, but could not enjoy it.

- The design of scenarios within a higher education setting that can be used to experiment with the proposed integration and analysis of the results in an effort to determine general guidelines.

To achieve these goals, seven tasks were proposed, organized in the following chronogram.



Given the high degree of uncertainty considering how the integration of these technologies would work out, the project's duration was divided into two incremental development cycles. After an

initial three-month period of architectural design, the development would start on all tasks, leading to a first system in month 18. This would include preliminary versions of all tests and scenarios, so that a round of experiments with the system could be performed.

The second development cycle would focus on tasks T3 (context and ambient intelligence) and T4 (adaptation), which would be revised to take into account the results gathered from the first cycle. This would lead, at the end of month 30, to a revised version of T3 and T4. The project's last 6 months will cover a final round of integration, including all tests, to measure both the degree of success and the effectiveness of the second cycle. Regardless of the end-of cycle tests, research and experiments with scenarios are being performed continuously throughout the project.

2 Task review and indicators

In general, the development of mechanisms and tools are well within schedule, and some of them started even earlier than planned. However, the integration of the whole system is advancing slower than expected. This is mainly due to the fact that the implementation of the proposed demonstrator is largely dependant on the technology available to the students. Even although the technology and devices required for some demonstrators are already available at the market, their high prices prevent a wider adoption of them by the students. For that reason, some experiments were simplified to fit the available technology, while others were downsized in order to use only devices owned by the research group.

Therefore, task T1 is only partially fulfilled and slightly behind schedule. The first phase of this task finished with a meeting with all the project participants at a workshop held in Granada at the CEDI'05 conference. During this phase, a common architecture was designed and interoperability protocols were agreed upon. The development of the different mechanisms and tools follows these architectural guidelines. Some of these tools have been tested independently and integrated with others. The integration and testing of the complete system are still pending.

With respect to T2, it follows the schedule as planned. The system capabilities regarding this layer will be specifically tested in an experiment to be carried out in February, as explained below (T5).

Regarding task T3, the middleware layer has been implemented and the benchmarking of the platform has already been performed. The following step is the integration with the rest of developments.

Concerning the adaptation layer (task T4), the analysis of the adaptation needs and feasibility was already done (T4.a). During this task execution, we decided to leave it open in case that new needs were detected. Currently we consider this task as finished. We have also finished both the specification of the adaptation formalism (T4.b) and the development of the corresponding adaptation mechanisms (T4.c), which could be refined in the following months, if needed. Regarding task T4.d, a collaborative tool to design and deliver educational materials has been developed. With respect to task T4.e, we have carried out one experiment dealing with the detection of grouping criteria, and are currently carrying out a more complete one (task T5.c). The results obtained feed the mechanism for automatic user grouping already developed, which could be improved in the case that new needs are detected after the next experiments, as it was foreseen

in this task. Concerning tasks T4.g, T4.h, T4.i and T4.j, all of them started before it was planned and have already been finished. In contrast, task T4.f was postponed because the offline tool does not need to be integrated with the rest of the system and its development would be easier after the main components of the adaptive layer are implemented. Finally, each mechanism and tool implemented has been tested independently (T4.k), and new ones will be checked accordingly.
Demonstrators are being developed (T5), as it was explained above, with some difficulties. Demonstrators testing the whole system functionality have not been implemented yet. However, some experiments to test individual developments and partial integrations have already been carried out:

- Three different experiments were carried out with high-school students and teachers, testing both the advantages of introducing Adaptive Educational Hypermedia (AEH) technology for supporting attention to diversity, and the tools to support AEH authoring (T4.g).
- A large experiment was conducted with more than 160 university students to investigate the impact of learning styles on the way in which students group themselves and also on the results of the collaboration tasks performed by them. The results obtained feed the mechanism for automatic user grouping developed (T4.e).

Other experiments are currently being carried out, associated to tasks T4.e (student grouping) and T4.i (CAA of free-text answers). Also, some other experiments are scheduled during the next two months aimed at:

- Testing of the adaptation mechanisms developed in task T4.c, which supports context-based activity recommendations for students.
- Testing the use of mobile phones for supporting student participation in large groups (more than 100 students). This experience will take place within our traditional classrooms The integration of the software developed in T2 and T4 will be used.

Finally, activities of task T6 have followed the planned evolution. Particularly, the integration of the evaluation methods and tools with the adaptive layer, as well as its testing, are even more developed than was originally intended. The main reason is the development of a simulation tool to generate synthetic logs. In this way, it was possible to test how evaluation tools would analyze and provide feedback to the adaptation layer even before this layer were developed. The main pending activity of this task is to apply the evaluation tools to the student data (to be obtained from the experiments mentioned above).

The research carried out during the project has given rise to relevant scientific and technological results, such as four doctoral thesis and the publications enumerated in the next section.

3 Project Results

3.1 Research Staff

The members of the research groups that carry out this project come from two Spanish universities: U. Autónoma de Madrid (UAM) and U. de Granada (UGR). Besides, other three researchers coming from two European universities are also taking part in U-CAT. The Ph.D. researchers initially involved in the UAM group were: X. Alamán, E. Alfonseca, R.M. Carro, F. Gómez, S. Holgado, J. López de Vergara R. Moriyón, A.Ortigosa and P. Rodríguez. In UGr, the

participants are M.Gea, J.L.Garrido and R. López-Cózar. Also, since the beginning of this project, G. Montoro, P. Haya and M. Mora have been working on dynamic dialogs in ambient intelligence, collaborative analysis of the learning process and context information in ambient intelligence, respectively, in the framework of the project. They all defended their doctoral thesis and got their PhD degrees in 2005-2006. Moreover, M. Freire, P. Paredes and D. Pérez are expected to defend their PhD thesis in 2007. Their works centers in complex graph-based structure visualization, learning styles and adaptive computer assisted assessment of free-text answers, respectively. Also, J.A.S. Itmazi defended his Ph.D. Thesis in 2005 at UGr in the U-CAT framework, dealing with e-learning in open and traditional universities.

Finally, other 9 Ph. D. students have already got their research qualification (DEA) at UAM and UGr within the U-CAT project framework..

3.2 Publications

Work on this project has given rise to 14 publications in international journals [1-14] and 39 presentations and publications in international and national conferences [15-53] (see section 4).

3.3 Technological Transfer

The design of the "DeepTest" computer application (2005), transferred to Active Documents S.L, is a direct result of technological transfer from our research lines. This application focuses on the resolution of exercises based on interactive documents. Other companies that stand to benefit from technological transfer are Ikerland and Coloplast S.A., which are closely following project results for their employee training needs.

A professorship sponsoring agreement has been reached with "Soluziona Consultoría and Tecnología S.L.", a leading IT consulting company. The "Cátedra de Patrocinio UAM-Soluziona CyT de Inteligencia Ambiental, Seguridad y Privacidad" will fund a full professorship in the area of ambient intelligence environments, focusing on security and privacy aspects, at UAM.

3.4 Participation in International Projects

Project researchers are participating in two relevant Integrated European Projects, AceMedia and COST 278. AceMedia is centered on the development and use of "Autonomous Content Entities" that would contain and identify semantically any type of multimedia contents. The project seeks to generate such entities automatically using advanced knowledge extraction and classification techniques, and enable advanced processing of these entities through knowledge-based techniques.

The "Spoken Language Interaction in Telecommunication" project, supported by the European Commission as COST 278, seeks to develop spoken language support, dialog management and multimodal communication for distributed systems. Support for pronunciation variants and multilingual speakers are other important research lines in this project.

3.5 Collaboration with Other International and National Groups

As already mentioned, three researchers from two European universities also take part in U-CAT project: J. Schlichter, chair of the "Group XI: Applied Informatics - Cooperative Systems" of the Department of Informatics of the Technical University of Munich (TUM), and S. Garlatti and Y.

Kemarrek, from the Department of Artificial Intelligence and Cognitive Sciences, *Ecole Nationale Supérieure des Télécommunications de Bretagne* (ENST), France. Both J. Schlichter and S. Garlatti regularly visit our group in Spain, as well as members of our group visit their research centers (M. Freire in June 2006, J. Bravo since April 2007 to June 2007).

Additional relations have been established, among others, with the ITC-irst research centre in Trento (Italy), where D. Pérez spent 5 months (Aug. 2004-Dec. 2004); University of Texas (USA); Tokyo Institute of Technology Precision, were E.Alfonseca spent one an a half year (Jul. 2005-Dec. 2006); and University of Lancaster, where M.A. Mora is spending one year (Feb. 2007 – Jan. 2008). These relations are giving raise to other possibilities of collaboration in the context of the U-CAT project, as well as to joint publications.

Also, U-CAT researchers keep relationships with other Spanish groups that share the same interest, such as the LIPNE group at UPV/EUH (Laboratorio de Interacción Persona-Computador para Necesidades Especiales) or the CHICO group (UCLM).

4 References

International journals

[1] E. Alfonseca, R.M. Carro, E. Martín, A. Ortigosa, P. Paredes, *The Impact of Learning Styles on Student Grouping for Collaborative Learning: A case study*. User Modeling and User-Adapted Interaction, Vol. 16 (3-4), pp. 377-401, September 2006.

[2] E. Alfonseca, R.M. Carro, M. Freire, A. Ortigosa, D. Pérez, P. Rodríguez, *Authoring of Adaptive Computer-Assisted Assessment of Free-text Answers*. Educational Technology & Society, Vol. 8 (3), pp. 53-65, July 2005.

[3] E. Alfonseca, P. Rodriguez, D. Perez. *An approach for automatic generation of adaptive hypermedia in education with multilingual knowledge discovery techniques*, Computers and Education. In press, online version available since Nov. 2005.

[4] J Bravo, A. Ortigosa, *Validating the Evaluation of Adaptive Systems by User Profile Simulation*.. Lecture Notes in Learning and Teaching. Eds: Weibelzahl, S. & Cristea, A. (National College of Ireland), 479-483, 2006.

[5] M. Freire, P. Rodríguez, *A Graph-Based Monitoring Tool for Adaptive Hypermedia Course Systems,* Adaptive Hypermedia and Adaptive Web-Based Systems. LNCS 4018, Vol. 1. Eds: V. Wade, H. Ashman and B. Smyth (Springer Verlag), 279-282, 2006.

[6] P.A. Haya, G. Montoro, A. Esquivel, M. García-Herranz, X. Alamán, *A Mechanism for Solving Conflicts in Ambient Intelligent Environments*, Journal of U. Computer Science, Vol. 12, No. 3, pp. 284-296, 2006.

[7] E. Martín, R.M. Carro,, P. Rodríguez, *A Mechanism to Support Context-based Adaptation in M-Learning.* Innovative Approaches for Learning and Knowledge Sharing. LNCS 4227, Eds: Nejdl, W. and Tochtermann, K. (Springer Berlin / Heidelberg), 302-315, 2006.

[8] E. Martín, N. Andueza, R.M. Carro,, P. Rodríguez, *Recommending Activities in Collaborative M-Learning.* Lecture Notes in Learning and Teaching, Eds: S. Weibelzahl and A. Cristea, 197-204, 2006.

[9] G. Montoro, P.A. Haya, X. Alamán, R. López-Cózar,, Z. Callejas. *A proposal for an XML definition of a dynamic spoken interface for ambient intelligence*. International Conference on Intelligent Computing. LNCS (Springer-Verlag), 711-716, 2006.

[10] R. Moriyón, F. Saiz, *A Tool for the Reinforcement of Conceptual Learning: Description and Use Experiences*. Journal of U. Computer Science, Vol. 11 (9), 2005.

[11] F. Muñoz, A. Ortigosa, *An Adaptive Course on Template-based Adaptive Hypermedia Design*. Lecture Notes in Learning and Teaching, Eds: S. Weibelzahl and A. Cristea, 2006.

[12] D. Pérez, E. Alfonseca, P. Rodríguez, *¿Pueden los ordenadores evaluar preguntas abiertas?* Novatica journal, 183, ATI (Association of Computer Engineers), 2006.

[13] D. Pérez, A. Gliozzo, E. Alfonseca, C. Strapparava, B. Magnini, P. Rodríguez, *About the effects of combining Latent Semantic Analysis with other Natural Language Processing techniques to assess open-ended questions*. Signos Journal, Vol 38, Num 59, December 2005.

[14] D. Pérez-Marín, E. Alfonseca, P. Rodríguez, *On the dynamic adaptation of Computer Assisted Assessment of free-text answers*. Proc. of the Int. Conf. of Adaptive Hypermedia, LNCS 4018, Springer-Verlarg, Dublin, Ireland, June 2006.

International and national conferences

[15] N. Andueza, R. M. Carro, *Supporting Configurable Alerting Services in Mobile Learning Environments*, Proc. 8th Int. Symp. on Information and Communication Technologies Applied to Education, León, Vol. 1, 326-333, October 2006.
[16] R.F. Arroyo, M. Gea, J.L. Garrido, P. Haya, *A Task-Driven Design Model for Collaborative AmI System*. Proc. of the 18th Int. Conf. on Advanced Information Systems Engineering (Luxemburgo), Presses Universitaires de Namur, 969-983, June 2006.
[17] R.F. Arroyo, J.L. Garrido, M. Gea, P. Haya, *A Design Model Applied to the Development of AmI Systems*. Proc. of the Int. Conf. on Ubiquitous Computing. Alcalá de Henares (Spain), pp: 95-102, June 2006.
[18] J. Bravo, A. Ortigosa*, Integración y Prueba de Herramientas de Evaluación en un Entorno Hipermedia Adaptativo,* Proc. 8th Int. Symposium on Information and Communication Technologies Applied to Education, Vol. 2, León, October 2006.
[19] C. Bravo, J.L. Garrido, M.A. Redondo, M. Gea*, Soporte a la decisión en entornos CSCL: Modelado y Análisis del Comportamiento del Grupo en el Sistema DOMOSIM-TPC mediante AMENITIES,* AIPO, Thomsom Paraninfo S.A. Madrid, 2005.
[20] A. Esquivel, P. A. Haya, G. Montoro, X. Alamán. *Una propuesta para un modelo de privacidad en entornos activos*. Proc. of the Simposio de Computación Ubicua e Inteligencia Ambiental (UCAmI 2005), J.Bravo y X. Alamán (eds.), Thomson, 381-388, Spain, 2005.
[21] M. Freire, P. Rodríguez, *Preserving the mental map in interactive graph interfaces*, Proceedings of the International Conference on Advanced Visual Interfaces (AVI'06), 270-273, Italy, 2006.
[22] M. Freire, P. Rodríguez, *Comparing Graphs and Trees for Adaptive Hypermedia Authoring,* 3rd Int. Workshop on Authoring of Adaptive and Adaptable Educational Hypermedia (A3EH) at the 12th International Conference AIED, 4-12, The Netherlands, July 2005.
[23] M. García-Herranz, P.A. Haya, G. Montoro, A. Esquivel, X. Alamán, *Adaptación automática de entornos activos mediante agentes basados en reglas*, 2nd Int. Workshop on Ubiquitous Computing and Ambient Intelligence, 189-204, Puertollano, November 2006.
[24] M. García-Herranz, P.A. Haya, A. Esquivel, G. Montoro, X. Alamán, *Semi-automation in Perceptive Environments: A rule-based agents proposal*, VII Cong. Int. Interacción Persona-Ordenador, November 2006.
[25] J.L. Garrido, P. Paderewski, P. Rodríguez, M.L. Hornos, M.J. Noguera, *A Software Architecture Intended To Design High Quality Groupware Applications*. Int. Conf. On Software Engineering Research And Practice. CSREA Press, USA, 59-65, 2005.
[26] F.J. Gómez-Arribas, S. Holgado, J.E. López-de-Vergara, Integración de experimentos de laboratorio remotos en un entorno ubicuo de aprendizaje. Proc. of the I Simposio sobre Computación Ubicua e Inteligencia Ambiental (UCAmI'2005), 337-344, 2005.
[27] M. González, J.L. Garrido, M. Noguera, *Estudio y Desarrollo De Sistemas CSCW Bajo RM/ODP: Una Aproximación Desde La Metodología AMENITIES*. Interacción Persona-Ordenador (AIPO), ISBN: 84-9732-436-6. Thomson Paraninfo S.A. Madrid, Spain, 411-412, 2005.
[28] P.A. Haya, A. Esquivel, G. Montoro, M. García-Herranz, X. Alamán, R. Hervás, and J. Bravo, *A prototype of context awareness architecture for AmI at home*. Int. Symp. on Intelligent Environments. Cambridge, ISBN 1-59971-529-5, 49-54, UK. April 5-7, 2006.
[29] P.A. Haya, G. Montoro, A. Esquivel, M. García-Herranz, X. Alamán, *Desarrollo de aplicaciones sensibles al contexto en Entornos Activos,* 2nd Int. Workshop on Ubiquitous Computing and Ambient Intelligence (WUCAmI'06), 205-217, Puertollano, November 2006.
[30] P.A. Haya, G. Montoro, Xavier Alamán, *Un mecanismo de resolución de conflictos en entornos de Inteligencia Ambiental*. Proc. of the Simposio de Computación Ubicua e Inteligencia Ambiental (UCAmI 2005), J.Bravo y X. Alamán (eds.), Thomson, 11-18, Spain, 2005.
[31] R. López-Cózar, Z. Callejas, G. Montoro, *New Multimodal Dialogue System For An Academic Intelligent Environment,* 4th Int. Conf. on Smart Homes and Health Telematics (ICOST2006), Belfast, Ireland, June 2006.

[32] R. López-Cózar, Z. Callejas, G. Montoro, *DS-UCAT: A New Multimodal Dialogue System for An Academic Application*. Interspeech Satellite Workshop "Dialogue on Dialogues – Multidisciplinary Evaluation of Advanced Speech-Based Interactive Systems", USA, 2006.
[33] R. López-Cózar, Z. Callejas, G. Montoro, P. Haya, *DS-UCAT: Sistema de Diálogo Multimodal y Multilingüe Para un Entorno Educativo*. Proc. IV Jornadas en Tecnología del Habla, 135-140, Zaragoza, November 2006.
[34] R. López-Cózar, Z. Callejas, M. Gea, J. L. Garrido, M. Cabrera, *Setting Up a Multimodal Dialogue System for Ubiquitous Environments*, "Interacción 2004", R. Navarro-Prieto, J. Lorés (Eds.), Springer Dordrecht, 2005.
[35] R. López-Cózar, Z. Callejas, *Fusión de Computación Ubicua y Sistemas de Diálogo: Un caso de Estudio*. Proc. VI Congreso Interacción Persona-Ordenador, Spain, 14-16, September 2005.
[36] R. López-Cózar, Z. Callejas, M. Gea, Germán Montoro, *Multimodal, Multilingual and Adaptive Dialogue System for Ubiquitous Interaction in an Educational Space*, Proc.s of the Int. Workshop ISCA Tutorial and Research Workshop on ASIDE, Denmark, 2005.
[37] E. Martín, N. Andueza,, R.M. Carro, *Architecture of a System for Context-based Adaptation in M-Learning*. Proceedings of the 6th IEEE Int. Conf. on Advanced Learning Technologies, IEEE Computer Society Press, 252-254, The Netherlands, 2006.
[38] E. Martín, N. Andueza, R.M. Carro, P. Rodríguez, *Adaptación de Actividades a Usuarios y Grupos en Entornos Hipermedia Colaborativos*. Proc. of the VII Cong. Internacional Interacción Persona Ordenador, 289-294, 2006.
[39] N. Medina, Mª V Hurtado,. J.L. Garrido, L. García, *Sistemas Hipermedia Evolutivos: Una Extensión para su Aplicación a Sistemas Cooperativos*. IADIS Conf. Iberoamericana WWW/Internet 2005, 279-286. Portugal, 2005.
[40] J. Moreno, X. Alamán, R. Cobos, *SKC: Midiendo la Intensidad de la Interacción de los Usuarios*. Proceedings VI Congreso Interacción Persona Ordenador (Interacción 2005) . Thomson, 89-96,Granada, Spain, 2005.
[41] G. Montoro, P.A. Haya, X. Alamán. *Propuesta de definición de una interfaz dinámica para la interacción con el entorno*. VI Congreso Interacción Persona Ordenador (Interacción 2005), Granada, Spain, September 2005.
[42] R. Moriyón, F. Saiz, *A Novel Type of Interactive Exercises for the Detection of Errors in Documents*.. Proc. of E-Learn 2005, Vancouver, Canada, November 2005.
[43] R. Moriyón, F. Saiz, *Aprendizaje basado en corrección mediante edición de errores en documentos interactivos (Learning by editing errors in interactive documents)*. Proc. of the Int. Symp. on Technologies for Computer Education, SINTICE, ADIE, Spain, 2005.
[44] F. Muñoz, A. Ortigosa, *Using Adaptive Hypermedia to Support Diversity in Secondary Schools*,
Proc. of the 6th International Conference ICALT, IEEE Computer Society Press, The Netherlands, 2006.
[45] M. Noguera, J.L. Garrido, M. González, M.J. Hornos, J.A. Gómez, *Aplicación Del Modelo De Coordinación Linda Al Desarrollo De Sistemas Groupware*. XIII Jornadas De Concurrencia y Sistemas Distribuidos. Thomson Paraninfo, Madrid, Spain, 277-288, 2005.
[46] P. Paredes, P. Rodriguez, *The Application of Learning Styles in both Individual and Collaborative Learning*, Proceedings of the 6th Int. Conf. ICALT, IEEE Computer Society Press, 252-254, The Netherlands, 2006.
[47] D. Pérez, O. Postolache, E. Alfonseca, D. Cristea, P. Rodríguez, *About the effects of using Anaphora Resolution in assessing free-text student answers*. Proceedings of the Int. Conf. RANLP, Bulgaria, September 2005.
[48] D. Pérez, Enrique Alfonseca. *Adapting the automatic assessment of free-text answers to the student*. Proceedings of the 9th Int. Conf. on Computer Assisted Assessment (CAA), Loughborough, United Kingdom, July 2005.
[49] D. Pérez, A. Gliozzo, C. Strapparava, E. Alfonseca, P. Rodríguez, B. Magnini, *Automatic Assessment of Students free-text Answers underpinned by the Combination of a BLEU-inspired algorithm and Latent Semantic Analysis*. AAAI Press, FLAIRS, Florida, U.S.A., May 2005.
[50] D. Pérez-Marín, I. Pascual-Nieto, E. Alfonseca, P. Rodríguez, *Automatic Identification of Terms for the Generation of Students Concept Maps*. Proc. .International Conference MICTE, Sevilla, Spain, November 2006.
[51] D. Pérez-Marín, E. Alfonseca, P. Rodríguez, I. Pascual-Nieto, *Willow: Automatic and adaptive assessment of students free-text answers*. Proceedings 22nd Int. Conference of the SEPLN, Zaragoza, Spain, September 2006.
[52] D. Pérez-Marín, E. Alfonseca, M. Freire, P. Rodríguez, J.M. Guirao, A. Moreno-Sandoval, *Automatic Generation of Students' Conceptual Models underpinned by Free-Text Adaptive Computer Assisted Assessment*.. Proceedings of the 6th International Conference ICALT, IEEE Computer Society Press, 252-254, The Netherlands, 2006.
[53] D. Pérez-Marín, E. Alfonseca, P. Rodríguez, *A free-text scoring system that generates conceptual models of the students knowledge with the aid of clarifying questions*. Proc. Int. Workshop SWEL, Dublin (Ireland), June 2006.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Desarrollo, Integración y Evolución de sistemas Software: un Enfoque de Líneas de producto (DIESEL) TIN2004-03145

Miguel A. Laguna*

Abstract

The quality factors related to software reusability, adaptability and maintenance are crucial for the efficiency of the software systems development and maintenance processes. Theoretical and operative support to achieve the mentioned quality factors can adopt an a priori (or systematic) approach or an a posteriori (or transformational) approach. This project covers both approaches in the context of product lines development and maintenance. In the first approach, the problem of requirements variability is treated from all the possible aspects: goals, functional and not functional aspects, characteristics of the final user, etc. Once a model that envisages all the aspects is established (if possible in an orthogonal way), it is necessary to mutually relate these requirements in a way that discover the dependencies between the different types of requirements and their traceability relationships to the product line design elements (and their variants). These relationships will be the base for the previous and necessary (tool supported) decisions process when concrete software system is going to be developed in the product line. On the other hand, we propose the application of diagnosis techniques to the product line evolution, and the automated elaboration and execution of refactoring plans. The necessary transformations in the databases used by the applications and the information these can provide to the process must be also taken into account. Similar approach and techniques can be used to tackle the subsequent product line evolution.
Keywords: product line, requirements, variability, software evolution, refactoring

1 Project Goals

The main goal of the project is to define techniques that increase the quality of the software in the context of product lines development and evolution. In a more detailed way, regarding the systematic approach (development):

- To represent the domain knowledge, by means of separation of concerns in the requirements phase (including goals, soft-goals, features, etc. and the traceability and dependencies between them).
- To determine the set of successive model transformations to achieve the final model.
- To build a tool that supports the previous goals in industrial environments.

*Email: mlaguna@infor.uva.es

And from the point of view of the transformation of legacy software systems into product lines (evolution):

- To explore the viability of techniques such as Formal Concept Analysis (FCA) or metrics to help the deduction of opportunities of software refactoring
- To guide the utilization of these techniques to obtain a diagnosis mechanism that suggests the modifications needed to build an initial version of a software product line. and to give support to its subsequent evolution
- To establish a process and a development environment that supports these techniques.
- To support the automation of this refactoring process. The transformation sequence is understood as a chain of refactorings or "Refactoring Plan".

The following Work plan and schedule describes the modules and tasks that were originally identified:

Module 1. Requirement Engineering for Product Lines

The first part of the project focuses into the requirements and their various aspects. The aim is to represent the domain knowledge, by means of separation of concerns in the requirements phase (including goals, soft-goals, features, skills of future users, etc.). The traceability, incompatibilities and dependencies between these types of requirements must be established. The operative support must be improved by the design of tools.

Responsible: Miguel Ángel Laguna

- Task T1.1: Definition and application of techniques to represent the variability and dependencies, from [year 1 , month 1] until [year 1, month 12]
- Task T1.2: Definition and application of techniques to represent the traceability, from [year 2,mes1] until [year 2, month 12]
- Task T1.3: Design and construction of a tool that supports the treatment of variability, the traceability, and the transformation of requirements, from [year 2 , month 1] until [year 3, month 12]
- Task T1.4: Aspect extraction techniques, from [year 2 , month 1] until [year 2, month 12]
- Task T1.5: Definition of a process of reverse engineering for system adaptation, from [year 3, month 1] until [year 3, month 12]

Module 2. Inference and execution of transformations to build the support to the product line

The goal is to determine, step by step, the transformation techniques to convert a system in another more reusable, adaptable and maintainable. In second place, we must build a domain framework that supports the line of product. This module is divided in three phases and includes the tasks T2.1 until T4.5. The last phase includes as activity the point of final unification of all the modules, with the analysis of results and the final conclusions of the project.

Phase 1: Diagnosis techniques

Responsible: Yania Crespo and María Esperanza Manso

- Task T2.1: Collection and study of real systems, from [year 1, month 1] until [year 1, month 3]
- Task T2.2: Application of techniques based on FCA and/or empirical methods, from [year 1, month 1] until [year 2, month 3]
- Task T2.3: Analysis of techniques of diagnosis, from [year 1, month 10] until [year 2, month 3]
- Task T2.4: Conclusions, from [year 2, month 3] until [year 2, month 6]

Phase 2: Diagnosis and assessment tool and refactoring plans

Responsible: Yania Crespo

- Task T3.1: Study and selection of entries of phase 2 , from [year 1, month 7] until [year 1, month 9]
- Task T3.2: Tools for the diagnosis automation, from [year 1, month 7] until [year 2, month 9]
- Task T3.3: Definition of developer/diagnosis results and diagnosis results/refactoring tool interactions, from [year 2, month 1] until [year 2, month 6]
- Task T3.4: Achievement and execution of refactoring plans, from [year 1, month 10] until [year 2, month 9]
- Task T3.5: Conclusions, from [year 2, month 10] until [year 2, month 12]

Phase 3: Definition of a domain construction and evolution process

Responsible: José Manuel Marqués

- Task T4.1: Definition of a process of domain frameworks construction from initial applications. from [year 2, month 7] until [year 3, month 4]
- Task T4.2: Development of the support necessary for the construction of the first version of a domain framework, from [year 2, month 7] until [year 3, month 9]
- Task T4.3: Definition of a process of domain frameworks evolution, from [year 2, month 10] until [year 3, month 7]
- Task T4.4: Development of the support necessary for domain frameworks evolution, from [year 2, month 8] until [year 3, month 10]
- Task T4.5: Analysis of the domain framework and project conclusions. Final point unification with the rest of the project, from [year 3, month 10] until [year 3, month 12]

Module 3. Reverse engineering and re-engineering of data bases to support the product line

Development of a methodology that supports the process of re-engineering of data bases and the construction of a series of tool to automate some of the tasks of this process. We will make fundamental use of the Formal Concepts Analysis. The process will be oriented, in the first place, to obtain important information in the task of the previous modules and, in second place, to transform the data base according to the requirements of the product line

Responsible: Yania Crespo and Carmen Hernández

- Task T5.0: Collection and study of the data bases of selected systems, from [year 1, month 1] until: [year 1, month 3]
- Task T5.1: Analysis of data, from [year 1, month 1] until: [year 2, month 3]
- Task T5.2: Conceptual abstraction, from [year 2 , month 1] until: [year 2, month 12]
- Task T5.3: Transformation of the schema and migration, from [year 2, month 10] until: [year 3, month 10]

2 Level of success achieved by the project

The main achievements are organized in blocks related to the structure of the project. The degree of advancement is proportional to the work done in the PhD theses associated to each module and phase.

2.1 Requirement Engineering for Product Lines (Module 1)

The advances in this module have defined techniques to represent and analyze the variability and traceability in product lines requirements and correspond to the PhD thesis work of Bruno González-Baixauli, that is in its final phase (Main references: [2, 3, 7, 12, 13, 14, 20, 21, 23, 24, 31]). As part of the objectives of the module, some tools and prototypes have been developed. The collaboration with the NFR group of the Toronto University has produced a tool that presents graphically to the requirements engineer the results of the analysis of relationships between goals and soft-goals, allowing the optimization of variant selection process. In the same field, experimental prototypes to generate transformations between features and UML class diagrams and composite views of requirement concerns have been built (References: [2, 6, 21, 31]).

2.2 Inference and execution of transformations to build the support to the product line, Diagnosis techniques (Module 2, Phase 1)

The work done in this module has been redirected mainly to empirical techniques based on metrics. The articles of M. Esperanza Manso on experimental research in collaboration with other research groups have established the foundations (References: [4, 5, 9, 10, 16, 26]). The application to the diagnosis of software and search for refactoring opportunities has generated several results (References: [8, 18, 19, 22, 25]) that covers the goals of this phase and corresponds to the PhD thesis of Raúl Marticorena (Refactoring catalog with language independence)

2.2 Inference and execution…, Diagnosis and assessment tool and refactoring plans (Module 2 Phase 2)

A framework based tool is being developed as an eclipse plug-in. The refactoring engine has been designed with language independence and extensions for Java are added. The results until this moment include the global architecture of the environment and the metrics and detection modules. A refactoring catalogue has been included to support the engine. This work is basically the PhD thesis of Carlos López, Framework support for refactoring reuse (References: [15, 18, 22, 25, 28]).

The last part of this phase, Refactoring plans, is the work that corresponds to the PhD thesis of Javier Pérez and is in its initial stage (References: [29, 30])

2.2 Inference and execution…, Definition of a domain construction and evolution process line (Module 2, Phase 3)

The pending work in this phase corresponds to the last year of the project and is in preliminary status. A general schema of process for/with reuse, applicable to a product line approach, was defined and published in LCNS and IEEE at the end of a previous project (DOLMEN, 2000-2003). Now we are adapting the process incorporating the evolution aspects.

2.5 Reverse engineering and re-engineering of data bases to support the product line (Module 3)

The work done with the collaboration of an undergraduate student† has generated a tool to apply the FCA to a case study of a legacy data base given by a local company (G1 Ingenieros). The theoretical previous work was published at the end of a previous project (DOLMEN, 2000-2003). Other case studies are in process as work corresponding to the PhD thesis of Cármen Hernández.

3 Achievement indicators

3.1 PhD theses

Funded by this project there are 6 people doing research for their PhD. Mª Esperanza Manso is going to defend her thesis in a few months. All the experimental part is concluded and published. Raul Marticorena, Carlos López and Bruno González-Baixauli could finish their work at the end of this year. The dissertation is expected at the end of this year (Bruno) or during 2008. The other two people, Carmen Hernández and Javier Pérez have done only preliminary work.

3.2 Publications

The references section lists the referred articles in journals, books, conferences and workshops from the end of 2004 until now. Other few works have been published internally as technical reports of the Department of Informatics of the University of Valladolid.

Articles in Journals and Chapters in Books

References [1] to [11] are articles in journals and chapters in books:

- International Journals: 3 (Empirical Software Engineering, Electronic Notes in Theoretical Computer Science, IEEE América Latina)

† Alberto Nistal Calvo. "Reingeniería de bases de datos mediante análisis de conceptos formales: un caso de estudio". PFC de Ingeniería Informática. Octubre 2005.

- International books (chapters): 7 (mainly in IEEE proceedings, 3 and Springer LCNS, 2)
- National books (chapters): 1

Congress and workshop contributions

References [12] to [31] are articles presented in national (7) and international events (13, 20 in totals):

- National conferences: 7 (mainly JISBD, 5 and DSDM, 2)
- International conferences and workshops: 13 (Apart from Requirements Engineering RE, ICSR and MoDELS conferences published in IEEE or LCNS, our woks have been presented in ACM Symposium on Applied Computing, Workshop on Requirements Engineering (WER), Quantitative Approaches in Object-Oriented Software Engineering, International Symposium on Empirical Software Engineering, etc.)

3.3 Collaboration with national and international groups

We maintain closed relation with other national groups, as consequence of past coordinated projects (MENHIR and DOLMEN, coordinated by Dr. Isidro Ramos). In this moment we are integrated in two national research nets funded by the MEC:

- Red de Desarrollo de Software Dirigido por Modelos (DSDM), TIN2005-25866-E, coordinated by Dr. Antonio Vallecillo.
- Líneas de Producto Software, TIN2007- (pending), coordinated by Dr. Oscar Díaz.

M. Esperanza Manso works with groups specialized on empirical software engineering in Castilla la Mancha (Dr. Mario Piattini) and UPV (Dr. Javier Dolado) and has published many join articles. (References: [4, 5, 9, 10, 16, 26]).

In the international context we work with Dr. Julio Leite (PUC of Rio de Janeiro, Brasil) and indirectly with the NFR group of Dr. Mylopoulos, Toronto University (Canada). These collaboration has result in three stays of Bruno González (one in Toronto, two in PUC-Rio) and has been materialized in several joint articles (References: [2, 3, 7, 13, 14, 23, 31])

3.2 Technological Transfer

We have collaborations with some local companies and organizations via article 83 LOU contracts with the aim of implementing some of the advances produced in software product lines field (PromaTIC system, Javier Pérez, 8 000€).

References

[1] Miguel A. Laguna, Oscar López, Yania Crespo, Reuse, Standardization, and Transformation of Requirements, Lecture Notes in Computer Science, LNCS 3107, pag. 329-338, 2004

[2] Bruno González-Baixauli, Julio Cesar Sampaio do Prado Leite, John Mylopoulos, Visual Variability Analysis for Goal Models, 12th IEEE International Conference on Requirements Engineering (RE 2004), Kyoto, Japan, ISBN 0-7695-2174-6, page 198--207 - sep 2004

[3] Bruno Gonzalez-Baixauli; Miguel A. Laguna; Julio Cesar Sampaio do Prado Leite, Aplicación de la Teoría de Constructos Personales a la Elicitación de Requisitos, IEEE América Latina, ISSN 1548-0992 Volumen 3:1, March 2005

[4] Marcela Genero, Geert Poels, Esperanza Manso, Mario Piattini, Defining and Validating Metrics for UML Class Diagrams, IMPERIAL COLLEGE PRESS, page 99--160 - 2005

[5] José A. Cruz-Lemus, Marcela Genero, Esperanza Manso, Mario Piattini, Evaluating the Effect of Composite States on the Understandability of UML Statechart Diagrams,Model Driven Engineering Languages and Systems: 8th International Conference (MoDELS 2005), Montego Bay, Jamaica, Springer LNCS Volume 3713, page 113--125 - oct 2005

[6] F. Javier Pérez García, Miguel A. Laguna, Yania Crespo González-Carvajal, and Bruno González-Baixauli, Requirements variability support through MDD and graph transformation, Electronic Notes in Theoretical Computer Science, ISSN: 1571-0661 pag. 161-173(2006)

[7] Bruno Gonzalez-Baixauli; Julio Cesar Sampaio do Prado Leite; Miguel A. Laguna, Eliciting Non-Functional Requirements Interactions Using the Personal Construct Theory, IEEE proceedings of the RE 2006 Conference, September 2006

[8] Yania Crespo, Carlos López, M. Esperanza Manso Martínez, Raúl Marticorena, from Bad Smells to Refactoring: Metrics Smoothing the way, in Object-Oriented design knowledge Principles, Heuristics and Best Practices. Chapter vii pp 193-249, Idea Group Publishing Hershey USA ISBN 1-59140-897-0, 2006

[9] Bocco Genero, Luis Reynoso, Esperanza Manso, Measuring Object Coupling in OCL Expressions: A cognitive Theory – Based Approach, Proceedings of The 23rd IEEE Instrumentation and Measurement Technology Conference, Page(s):1087 – 1092, 2006

[10] M. Esperanza Manso Martínez, J. Javier Dolado Cosín, Cap.25. Utilidad del Análisis de Componentes Principales en la Experimentación en Ingeniería del Software, in press

[11] Marcela Genero, Esperanza Manso, Aaron Visaggio, Gerardo Canfora and Mario Piattini, Building a measure-based prediction model for UML class diagram maintainability. Acepted for publication in Empirical Software Engineering, 2007

[12] Bruno González-Baixauli, Miguel A. Laguna, Yania Crespo: Product Line Requirements based on Goals, Features and Use cases, International Workshop on Requirements Reuse in System Family Engineering (International Conference on Software Reuse 2004)

[13] Bruno González-Baixauli, Miguel A. Laguna, Julio Cesar Sampaio do Prado Leite: Análisis de Variabilidad con Modelos de Objetivos. VII Workshop on Requirements Engineering (WER-2004). Diciembre 2004

[14] Bruno González-Baixauli, Miguel A. Laguna, Julio Cesar Sampaio do Prado Leite: Aplicaciones de la Teoría de Constructos Personales a la Elicitación de Requisitos, IX Jornadas de Ingeniería del Software y Bases de Datos, JIDBD, nov. 2004

[15] Yania Crespo, Carlos López, Raul Marticorena, Un Framework para la Reutilizacion de la Definicion de Refactorizaciones, IX Jornadas Ingenieria del Software y Bases de Datos (JISBD 2004), Malaga, Spain ISBN: 84-688-8983-0, page 499--506 - nov 2004

[16] Marcela Genero, Mario Piattini, Esperanza Manso, Finding "Early" Indicators of UML Class Diagrams Understandability and Modifiability,2004 International Symposium on Empirical Software Engineering (ISESE 2004) , Redondo Beach, CA, USA. , ISBN 0-7695-2165-7, page 207--216 - aug 2004

[17] José A. Cruz-Lemus, Marcela Genero, Esperanza Manso, Mario Piattini, Do Composite States Improve the Understanding of UML Statechart Diagrams?, X Jornadas Ingenieria del Software y Bases de Datos (JISBD 2005), Granada, Spain ISBN: 84-9732-434-X - sep 2005

[18] Yania Crespo, Carlos López, Raul Marticorena, Esperanza Manso, Language Independent Metrics Support towards Refactoring Inference, 9th ECOOP Workshop on QAOOSE 05 (Quantitative Approaches in Object-Oriented Software Engineering). Glasgow, UK. ISBN: 2-89522-065-4, page 18-29 - jul 2005

[19] Raul Marticorena, Carlos López, Yania Crespo, Parallel Inheritance Hierarchy: Detection from a Static View of the System,6th International Workshop on Object Oriented Reenginering (WOOR), Glasgow, UK., page 6 - jul 2005

[20] Bruno González-Baixauli, Miguel A. Laguna, Yania Crespo, Product Line Requirements based on Goals, Features and Use cases, International Workshop on Requirements Reuse in System Family Engineering (IWREQFAM), page 4--7 - jun 2004

[21] Bruno González-Baixauli, Miguel A. Laguna, Yania Crespo, Product Lines, Features , and MDD, 1st European Workshop on Model Transformation (EWMT 05) - sep 2005

[22] Raul Marticorena, Yania Crespo, Carlos López, Soporte de Métricas con Independencia del Lenguaje para la Inferencia de Refactorizaciones, X Jornadas Ingenieria del Software y Bases de Datos (JISBD 2005), Granada, Spain ISBN: 84-9732-434-X, page 59--66 - sep 2005

[23] Bruno González-Baixauli, Miguel A. Laguna, Julio Cesar Sampaio do Prado Leite, Aplicación de un Enfoque Intencional al Análisis de Variabilidad, VIII Workshop on Requirements Engineering, Junio 2005

[24] Bruno Gonzalez-Baixauli, Miguel A. Laguna, MDA e Ingeniería de Requisitos para Líneas de Producto, II Taller sobre Desarrollo de Software Dirigido por Modelos, MDA y Aplicaciones (DSDM 2005)

[25] Yania Crespo, Carlos López, Raul Marticorena, Relative Thresholds: Case Study to Incorporate Metrics in the Detection of Bad Smells.10 the ECOOP Workshop on QAOOSE'06, Quantitative Approches in Object-Oriented Software Engineering. Nantes, France. ISBN 88-6101-000-8. 2006

[26] Reynoso Luis, Genero Bocco Marcela, Piattini Mario, Manso Martínez M. Esperanza, Does Object Coupling really affect the Understandability and Modifiability of OCL expressions? 21st ACM Symposium on Applied Computing. SAC-SE 2006

[27] Bruno González-Baixauli, Miguel A. Laguna y Julio Cesar Sampaio do Prado Leite: Variabilidad en Requisitos: Un Enfoque Basado en Meta-Modelado y Transformación de Modelos, III Taller sobre Desarrollo de Software Dirigido por Modelos (DSDM 2006), 2006.

[28] Carlos López, Raul Marticorena, Yania Crespo, Caracterización de Refactorizaciones para la Implementación en Herramientas, XI Jornadas de Ingeniería del Software y Bases de Datos, Sitges, Barcelona, 2006. ISBN:84-95999-99-4, page 538--543 - oct 2006

[29] Javier Pérez, Yania Crespo, First Results in Supporting Automated Elaboration of Refactoring Plans, Summer School on Generative and Transformational Techniques in Software Engineering - Braga, 2005

[30] Javier Pérez, Overview of the Refactoring discovering Problem, Doctoral Symposium, 20th edition of the European Conference on Object-Oriented Programming (ECOOP 2006) - Nantes, jul 2006

[31] Bruno Gonzalez-Baixauli, Miguel A. Laguna, Julio Cesar do Prado Sampaio Leite, Using Goal-Models to Analyze Variability, VAMOS - First International Workshop on Variability Modelling of Software-intensive Systems, Limerick, Ireland, January 16–18, 2007

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

LOGICTOOLS
TIN2004-03382

Albert Rubio *
Universitat Politècnica de Catalunya

Abstract

The aim of this project is the development of new techniques, both at the theoretical level and at the implementation level, for the following four kinds of logic-based tools:

1. Automated theorem provers for first-order logic.
2. Efficient decision procedures for certain logics.
3. Environments for automated termination proving.
4. Environments for program verification

where the first three kinds of tools will be applied as well to the fourth one. Furthermore, four additional kinds of specific problems have been selected for which we will intensively work on the applicability of these tools and their underlying techniques (such as data structures and algorithms, or constraint propagation methods): circuit verification problems, planning and scheduling (for industrial problems, or for sport events), bio-informatics, and the semantic web.
Keywords: Logic in Computer Science, Decission procedures, Formal methodos, Verification of computing systems.

1 Goals of the project

Logic-based theory, techniques and tools are having an increasingly big impact on different computer science areas, as well as on finding solutions to numerous computational problems that arise in industry and other sciences like Biology.

McCarthy wrote, back in the sixties, that logics would have an importance in computer science similar to the importance that mathematical analysis had in Physics in the XIX century. Along the same lines, Manna and Waldinger [5] called logic *"The Calculus of Computer Science"*, since its role in computer science, both at the theoretical and practical level, is similar to the one of Mathematics in Physics and, in particular, to the one of calculus in engineering.

Nowadays, it is widely accepted that in all sorts of systems, logic-based methods and tools are able to improve their quality and reduce their cost. This view is documented in the paper *"On the Unusual Effectiveness of Logic in Computer Science"* [3], where the crucial role of logics in areas like databases, programming languages, complexity, agent systems and verification

*Email: rubio@lsi.upc.edu

Task:	**1**st year jfmamjjasond	**2**n year jfmamjjasond	**3**rd year jfmamjjasond
1	xxxxxxxxxxxx	xxx.	. . . . .
2	xxxxxxxxxxxx	xxxxxx.	. . . . .
3	xxxxxxxxxxxx	xxxxxx.	. . . . .
4	. . . . .	. . . xxxxxxxxx	xxxxxxxxxxx
5	. . . . .	 xxxxxx	xxxxxxxxxxx
6	xxxxxxxxxxx	xxxxxxxxxxx	xxxxxxxxxxx
7	xxxxxx.	. . . . .	. . . . .
8	xxxxxxxxxxx	. . . . .	. . . . .
9	xxxxxxxxxxx	xxxxxx.	. . . . .
10	. . . . .	xxxxxxxxxxx	xxxxxxxxxxx
11	xxxxxxxxxxx	xxxxxxxxxxx	xxxxxxxxxxx
12	xxxxxxxxxxx	xxxxxxxxxxx	xxxxxxxxxxx
13	xxxxxxxxxxx	xxxxxxxxxxx	xxxxxxxxxxx
14	xxxxxxxxxxx	. . . . .	. . . . .
15	. . . . .	xxxxxxxxxxx	xxxxxxxxxxx
16	xxxxxxxxxxx	xxxxxxxxxxx	. . . . .
17	. . . . .	xxxxxxxxxxx	. . . . .
18	. . . . .	xxxxxxxxxxx	xxxxxxxxxxx
19	xxxxxxxxxxx	xxxxxxxxxxx	xxxxxxxxxxx

Figure 1: Chronogram

is exposed by worldwide experts (see also `http://www.cs.rice.edu/~vardi/logic/`) The goal of our project is to develop new techniques, both at the theoretical level and at the implementation level, for several kinds logic-based tools, as well as transfer as much of this technology as possible to industry and social services.

The particular tools we want to develop are: automated theorem provers for first-order logic, efficient decision procedures for certain logics, environments for automated termination proving and system verification. As specific problems, where we are interested in applying our results, we are considering: circuit verification problems, planning and scheduling (for industrial problems or for sport events), bio-informatics and the semantic web.

1.1 Specific tasks

In the goal of developing tools for first-order logic and its applications we have six tasks. Task 1: Prototype implementation of the extension to first-order logic of some successful techniques for propositional logic. Task 2: Experimental implementation of the *Model Evolution Calculus* [1]. Task 3: Study of the relationship between some description logics and some decidable logics. Task 4: Efficient implementation of Task 1 and benchmarking. Task 5: Efficient implementation of Task 2 and benchmarking. Task 6: Study and implementation of new indexing data structures and its application to bio-informatics problems.

For the development of efficient decision procedures for some decidable logics we have five more tasks. Task 7: Improve the current implementation of our *congruence closure* algo-

rithm [6]. Task 8: Keep developing the implementation of $DPLL(=)$. Task 9: Further study of theoretical foundations of decision procedures based on term rewriting techniques [2]. Task 10: Study of other instances of $DPLL(X)$, like DPLL(Z), DPLL(FD), o DPLL(R), and its application to verification problems. Task 11: Design and implement better constraint propagation algorithm and apply them to real scheduling problems like the ones of the Spanish professional basketball league (ACB).

For the development of termination proving tools we have four tasks. Task 12: Define and implement new techniques and tools for proving termination of higher-order rules [4], which can be used in proof assistants like Coq. Task 13: Maintain and develop the termination problem data base (TPDB). Task 14: Extend the current term rewriting techniques for proving termination in the presence of *built-in* domains like the integers or lineal data structures. Task 15: Extend our current automatic termination provers to handle richer languages.

Finally, for the development of program verification environments we have the last four tasks. Task 16: Extend the results in [7] for the automatic generation of invariants. Task 17: Study the possibility to include some efficient static analysis techniques. Task 18: Improve the efficiency and power of the symbolic computation and automated deduction subsystems. Task 19: Keep developing and maintaining the system.

1.2 Resources

The project group currently consists of by 11 full-time members of the university (7 have a PhD) and five more researchers from other research centers: University of Girona, University of Vic, University of Iowa, University of New Mexico and the SRI Computer Science Laboratory.

With respect to the equipment, apart from some personal computers, within the context of the project we have recently acquired a cluster of 6 dual core machines for benchmarking and testing for the development of efficient decision procedures and a disk array with 4.5 Tb for the bio-informatics applications. Both the cluster and the array will need to be enlarged in the near future in order to stay the group competitive with respect to other international research teams.

2 Project progress and achievements

As can be observed in the following sections there has been an important progress in the context of the project. Our group has become an international reference in some of the areas of the project, in which we are working only since 2003. We have also succeeded in having industrial contracts by transferring the technology we are developing to very different kind of problems, like in professional sports leagues, in hospitals or in hardware and software companies. Moreover several PhD thesis on the lines of the project have been defended or are in progress.

In what follows we will relate the publications and software developed to each of the four goals of the project.

For the tasks of the development of tools for first-order logic and its applications, apart from the work developed at the University of Iowa by Cesare Tinelli (external member of the project), we have produced the following publications and software: 3.2.6, 3.4.3 and 3.4.4.

For the tasks of the development of efficient decision procedures for some decidable logics, we have produced the following publications and software: 3.2.2, 3.2.3, 3.3.1, 3.3.5, 3.3.6, 3.3.10-14 and 3.4.1.

For the tasks of the development of termination proving tools, we have produced the following publications and software: 3.2.1, 3.3.3, 3.3.7-9 and 3.4.2.

For the tasks of the development of program verification environments, apart from the work throw contracts with Intel and Microsoft for applying our techniques in their verification tools, we have produced the following publications: 3.2.4, 3.3.15-17.

3 Quality indicators

In this section we include several indicator to show the relevance of the results of the project. We have included only those indicators that involve at least one of the members of the project at the Universitat Politècnica de Catalunya (and not the ones including only external members).

3.1 PhD training

Since the project started three PhD's have been defended, Albert Oliveras and Enric Rodríguez, supervised by Robert Nieuwenhuis and Enrique Blanco, supervised by Xavier Messeguer and Roderic Guigó. The PhD of Mirtha Lina Fernandez, supervised by Guillem Godoy and Albert Rubio, is in deposit and it is going to be defended in early spring.

Additionally three more student, Roberto Asín, Bernat Gel and Todd Treangen are developing their PhD within the goals of the project, and two more will start next September.

3.2 International journals

We have publications within the scope of the project in high quality journals, like two papers 3.2.1 and 3.2.2 in the *Journal of ACM*, which is in the top 10 of all the four categories it appears in the JCR, or the paper 3.2.6 in *BMC Bioinformatics* which is indexed as number 13 out of 139 in the *Biotechnology and applied microbiology* category of the JCR, or the paper 3.2.3 in *Information and Computation* indexed as number 32 out of 151 in the *Applied mathematics* category.

1. J.-P. Jouannaud and A. Rubio. Polymorphic Higher-Order Recursive Path Orderings. Journal of the ACM. To appear. 2007.

2. R. Nieuwenhuis, A. Oliveras and C. Tinelli. Solving SAT and SAT Modulo Theories: From an Abstract Davis-Putnam-Logemann-Loveland Procedure to DPLL(T). Journal of the ACM, to appear. 2007.

3. R. Nieuwenhuis and A. Oliveras Fast Congruence Closure and Extensions. Information and Computation, to appear. Elsevier Science, 2007.

4. E. Rodríguez-Carbonell and D. Kapur. Automatic Generation of Polynomial Invariants of Bounded Degree using Abstract Interpretation. In Science of Computer Programming, Volume 64, Issue 1, January 2007.

5. E. Blanco, X. Messeguer, T. Smith and R. Guigó. Transcription Factor Map Alignment of Promoter Regions. PLoS Computational Biology , 2(5), 2006.

6. T. Treangen and X. Messeguer. M-GCAT: Interactively and efficiently constructing large-scale multiple genome comparison frameworks in closely related species. BMC Bioinformatics, 7:433, 2006.

7. E. Blanco, D. Farre, M.M. Alba, X. Messeguer and R. Guigó. ABS: a database of Annotated regulatory Binding Sites from orthologous promoters. Nucleic Acid Research34, D63-D67, 2005.

8. D. Dominguez-Sal, X. Messeguer and J.A. Subirana FibreHelix: A programe for calculating the X ray diffraction pattern of macromolecules with helical symmetry: Aplications to DNA coiled-coils. Acta Cryst. D61, 2005.

3.3 Published papers in international conferences

As for journals, we have published in high quality conferences. We have two new papers 3.3.5 and 3.3.12 in the main international conference on verification, *Computer-Aided Verification (CAV)* (which belongs to the top 3.5% in the Citeseer Venues list[1]). We also have the paper 3.3.17 in SAS (in the top 12% of the list), the paper 3.3.9 in CADE (in the top 15% of the list), the papers 3.3.4, 8 and 13 in RTA (in the top 22% of the list) and all others in the top 40% of the list.

1. C. Barrett, R. Nieuwenhuis, A. Oliveras and C. Tinelli. Splitting on Demand in Satisfiability Modulo Theories. 13th Int. Conf. Logic for Programming, Artif. Intell. and Reasoning (LPAR). LNAI 4246:512-526, Springer, 2006.

2. F. Blanqui, J.-P. Jouannaud and A. Rubio. Higher-Order Termination: from Kruskal to Computability (Invited Paper). 13th Int. Conf. on Logic for Programming Artificial Intelligence and Reasoning (LPAR). LNAI 4246:1-14, Springer, 2006.

3. A.E. Darling, T. Treangen, L. Zhang, C. Kuiken, X. Messeguer and N.T. Perna. Procastination leads to efficient filtration for local multiple alignment. 6th Int. Workshop on Algorithms in Bioinformatics (WABI'06), LNCS 4175:126-137. Springer, 2006

4. J.-P. Jouannaud and A. Rubio. Higher-Order Orderings for Normal Rewriting. 17th Int. Conf. on Rewriting Techniques and Applications, LNCS 3467:387-399, Springer, 2006.

5. S. Lahiri, R. Nieuwenhuis and A. Oliveras. SMT Techniques for Predicate Abstraction. 18th Int. Conf. on Computer Aided Verification (CAV). LNCS 4144:424-437. Springer, 2006.

6. R. Nieuwenhuis and A. Oliveras. On SAT Modulo Theories and Optimization Problems. 9th Int. Conf. on Theory and Applications of Satisfiability Testing (SAT). LNCS 4121:156-169. Springer, 2006.

[1] we are using the estimated impact of publication venues in Computer Science given in CiteSeer (`http://citeseer.ist.psu.edu/impact.html`)

7. G. Godoy, M.-L. Fernandez and A. Rubio. Recursive Path Ordering can also be incremental. 12th Int. Conf. Logic for Programming, Artif. Intell. and Reasoning (LPAR). LNAI 3835:230-245. Springer, 2005.

8. G. Godoy, M.-L. Fernandez and A. Rubio. Orderings for Innermost Termination. 16th Int. Conf. on Rewriting Techniques and Applications (RTA). LNCS 3467:17-31, Springer, 2005.

9. G. Godoy and A. Tiwari. Termination of Rewrite Systems with Shallow Right-Linear, Collapsing, and Right-Ground Rules. 20th Int. Conf. on Automated Deduction. LNCS 3632:164-176, Springer, 2005.

10. G. Godoy and A. Tiwari. Confluence of Shallow Right-Linear Rewrite Systems. 19th Int. Workshop on Computer Science Logic (CSL), 14th Annual Conf. of the EACSL. LNCS 3634:541-556. Springer, 2005.

11. R. Nieuwenhuis and A. Oliveras. Decision procedures for SAT, SAT Modulo Theories and Beyond. The BarcelogicTools. (Invited Paper). 12th Int. Conf. Logic for Programming, Artif. Intell. and Reasoning (LPAR). LNAI 3835:23-46, Springer, 2005.

12. R. Nieuwenhuis and A. Oliveras. DPLL(T) with Exhaustive Theory Propagation and its Application to Difference Logic. 17th Int. Conf. on Computer Aided Verification (CAV). LNCS 3576:321-334, Springer, 2005.

13. R. Nieuwenhuis and A. Oliveras. Proof-Producing Congruence closure. 16th Int. Conf. on Rewriting Techniques and Applications. LNCS 3467:453-468. Springer, 2005.

14. R. Nieuwenhuis, A. Oliveras and C. Tinelli. Abstract DPLL and Abstract DPLL Modulo Theories. 11th Int. Conf. on Logics for Programming, AI and Reasoning (LPAR). LNAI 3452:36-50. Springer 2005.

15. E. Rodríguez-Carbonell and A. Tiwari. Generating Polynomial Invariants for Hybrid Systems. 8th Int. Workshop on Hybrid Systems: Computation and Control (HSCC'05). LNCS 3414:590-605. Springer, 2005.

16. R. Clarisó, E. Rodríguez-Carbonell and J. Cortadella. Derivation of Non-structural Invariants of Petri Nets Using Abstract Interpretation. 26th Int. Conf. on Application and Theory of Petri Nets and Other Models of Concurrency (ICATPN'05). LNCS 3536:188-207. Springer, 2005.

17. R. Bagnara, E. Rodríguez-Carbonell and E. Zaffanella. Generation of Basic Semi-algebraic Invariants Using Convex Polyhedra. 12th International Symposium on Static Analysis (SAS'05). LNCS 3672:19-34. Springer, 2005.

3.4 Software tools

1. Barcelogic. This tool won all four categories of the 2005 SMT-Competition in which it participated. In the 2006 SMT-Competition it came second in all categories in which it participated. Moreover, the BarceLogic SAT solver beat well-known systems like Chaff or Berkmin in the last competition over industrial problems (see SAT Race in Google).

2. HORPO. It is an implementation written in GNU-Prolog of the Higher-Order Recursive Path Ordering.

3. M-GCAT: Multiple Genome Comparison and Alignment tool (Academic software demo). 13th Annual Int. Conference on Intelligent Systems for Molecular Biology (ISMB 2005).

4. M-GCAT: Multiple Genome Comparison Software (Academic software demo). 4th European Conference on Computational Biology (ECCB 2005).

3.5 Funded projects and contracts

1. Project name: Instituto Nacional de Bioinformática. Funded by Genoma España. From april 2004 to April 2007. Xavier Messeguer has been project manager until April 2006. Total amount: 600.000 Euros.

2. Contract name: Diseño de un Gestor de la Base de Dades del CGMM (Centre de Genètica Mèdica i Molecular de l'Institut de Recerca Oncològica . Funded by Institut Català d'Oncologia. Year:2005. Responsable: Xavier Messeguer. Total amount: 18.000 Euros.

3. Contract name: Configuración del calendario de competición de la ACB. Funded by Asociación de Clubes de Baloncesto (ACB). Since 2003. Responsable: Albert Rubio. Total amount: 41.000 Euros

4. Contract name: Development of the KNVB calendar. Funded by the Royal Netherlands Football Association (KNVB) and Hypercube Business Inn. Since 2005. Responsable: Robert Nieuwenhuis. Total amount: 51.000 Euros

5. Contract name: SMT solvers for high-level hardware verification. Funded by Intel Corporation. Since 2005. Responsable: Robert Nieuwenhuis. Total amount: 59.000 Euros

6. Contract name: Efficient SMT solvers. Funded by Microsoft Research. Year 2006. Responsable: Albert Oliveras. Total amount: 12.000 Euros.

3.6 Collaborations

The group have a lot of national and international collaborations as can be seen from the list of publications given in sections 3.2 and 3.3. Apart from this, we would like to mention that Xavier Messeger was heading the Bio-informatics area, called ALGGEN (Algorithmics and Genetics Group), of the BSC (Barcelona Supercomputering Center) until Septembre 2006. ALGGEN is dedicated to research and teaching in the area of Computational Biology and Bioinformatics and its web site (`http://alggen.lsi.upc.es/`) is the virtual laboratory that contains the bioinformatic tools designed by the group.

3.7 Technology transfer

Apart from the technology transfer made throw contracts (see section 3.5) and collaborations, we would like to mention a work in progress with Sant Joan de Déu Hospital in Barcelona

to optimize their operating rooms schedules, a topic of huge social impact due to the problem of usually having too long waiting lists. Currently, we have a prototype system working satisfactorily.

Other ongoing applications, but in a less advanced state, include constructing timetables (schools, universities, hospitals, companies, transport) and some others involving planning sport events.

3.8 Other research activities

As an additional hint of the relevance of the obtained results in the respective communities, we would like to mention other activities like, chairing or participating in Program Committees (PC) of international workshops and conferences.

In 2005, Robert Nieuwenhuis was PC Chair of the 20th International Conference on Automated Deduction (CADE), and co-organized the Deduction and Applications Dagstuhl Seminar (Proceedings 05431). In 2006, several team researchers were members of PC of international conferences and workshops: Guillem Godoy (17th RTA conference), Robert Nieuwenhuis (PDPAR 2006 workshop) and Albert Oliveras (6th IWIL workshop). In 2007, Robert Nieuwenhuis is co-organizing the Deduction and Decision Procedures Dagstuhl Seminar and Albert Oliveras is co-chair of the 5th International Workshop on Satisfiability Modulo Theories (SMT 2007; successor of PDPAR). As PC members in 2007: Robert Nieuwenhuis (5th SMT workshop), Albert Oliveras (21st CADE conference and 6th FroCoS whorkshop) Enric Rodriguez (WING'07 workshop) and Albert Rubio (4th HOR workshop and 8th FTP workshop).

References

[1] P. Baumgartner and C. Tinelli. The Model Evolution Calculus. In *19th Int. Conf. on Automated Deduction*, LNAI 2741. Springer, 2003.

[2] H. Comon, G. Godoy, and R. Nieuwenhuis. The confluence of ground term rewrite systems is decidable in polynomial time. *42nd Annual IEEE Symp. on Foundations of Computer Science*, 2001.

[3] J. Y. Halpern, R. Harper, N. Immerman, P. G. Kolaitis, M. Y. Vardi, and V. Vianu. On the unusual effectiveness of logic in computer science. *The Bulletin of Symbolic Logic*, 7(2):213–236, 2001.

[4] J.-P. Jouannaud and A. Rubio. The higher-order recursive path ordering. In *14th IEEE Symposium on Logic in Computer Science*, 1999.

[5] Z. Manna and R. Waldinger. *The Logical Basis for Computer Programming.* Addison-Wesley, Reading, 1985.

[6] R. Nieuwenhuis and A. Oliveras. Congruence Closure with Integer Offsets. *10h Int. Conf. Logic for Programming, Artif. Intell. and Reasoning*, LNAI 2850. Springer, 2003.

[7] E. Rodríguez-Carbonell and D. Kapur. Automatic generation of polynomial loop invariants: algebraic foundations. *2004 Int. Symp. on Symbolic and +Algebraic Computation*, 2004.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

High performance processing: architecture, development environments and applications TIN2004-03388

Emilio Luque*
Computer Architecture and Operating System Department (CAOS)
University Autonoma of Barcelona (UAB). Spain
Investigador Principal

Abstract

The objective of this project is to study different aspects of parallel and distributed computing, from the processor level to applications. The project is organized in four research lines:
1- Environment for performance evaluation and applications tuning. This research covers parallel program development using programming frameworks and templates, including techniques for run-time automatic performance evaluation and improvement of those programs and task mapping and performance prediction in distributed systems. **2- Strategies and tools for distributed systems management.** Four different approaches in parallel/ distributed computing are evaluated: Grid environments, dedicated cluster based on user-shared machines, geographically distributed dedicated clusters and web-computing. **3- Multiprocessor architecture.** The advanced "fetch unit" of high performance processors and the interconnection network behaviour are the selected topics for concentrating our research effort in order to improve the system performance. **4- Applications.** Three applications demanding high performance computing are investigated. The first one is concerned with the analysis of DES (Distributed Event Simulation) applied to Individual oriented Models (IoM) in the field of biology/ecology. The second application is focused on the design of the Video-on-Demand (VoD) service, based on distributed systems. The last application concerns the use of advanced simulation techniques to face the problem of forest fires propagation simulation using parallel/distributed systems.

Keywords: High Performance Computing. Parallel Computing. Programming environments for parallel systems. Tools for performance evaluation and tuning. Static scheduling and dynamic load balancing. Cluster, Multicluster and Grid computing. Distributed heterogeneous systems. Fault Tolerant in clusters.Parallel Computers Architecture. Interconnection networks. Parallel and Distributed Simulation. Multimedia Architecture. Video-on-Demand. Computational Science and Engineering

1 Objectives of the project

The research group developing this project is composed of 17 PhD researchers and 18 graduate students working on their PhD thesis. The research objectives are described in following four basic research lines:

* Email: emilio.luque@uab.es

1.1 Environment for performance evaluation and applications tuning

This research covers parallel program development using programming frameworks and templates, including techniques for run-time automatic performance evaluation and improvement of those programs and task mapping and performance prediction in distributed systems.

The goals of this research line can be summarized in the following points:

1. Design and development of static and dynamic performance analysis tools.
2. Dynamic tuning of applications developed using a programming framework for parallel/distributed applications specification
3. Dynamic tuning of applications based on mathematical libraries.
4. Dynamic tuning of parallel applications on grid environments.
5. Generalization of the modelling phase and the task mapping phase to represent applications with different kinds of parallelism
6. Develop a prediction environment to respond to some important questions prior to execution

1.2 Strategies and tools for distributed systems management

Four different approaches in parallel/distributed computing are evaluated: Grid environments, dedicated cluster based on user-shared machines, geographically distributed dedicated clusters and web-computing.

GRID computing

The goals of this research line can be collected in the following points:

1. Evaluation, design and implementation of advanced resource-management mechanisms and policies that could provide automatic and reliable support for interactive and parallel jobs over grid environments.
2. Research on mechanisms that enable run-time monitoring tools to be launched under the control of resource management systems used on large-scale distributed environments.

Dedicated cluster based on user-shared machines

Using the computing power of the idle computer resources of Network of Workstations (NoW) to execute distributed applications, with a performance equivalent to a MPP and without perturbing the performance of the local user's applications on each workstation.

Geographically distributed dedicated clusters

The objective is to improve the execution performance of parallel applications adapting them from their original single cluster to heterogeneous multi-cluster environments. The goal is to reduce the execution time guaranteeing a minimum level of efficiency on both local and distance resources usage.

Web-computing

The goal of this approach (called Parasite) is to develop the technology and the framework to develop parallel applications to be executed, during the idle Internet navigation time, on the distributed infrastructure of not-dedicated personal computers, used as computing nodes of a distributed cluster machine.

1.3 Multiprocessor architecture

The performance of the advanced "fetch unit" of superscalar processors limited by Control Flow dependences, the configuration of the first-level cache and its strong influence on the performance and energy consumption of a processor, and the interconnection network behaviour are the goals of this research line and can be summarized in the following points:

1. Develop and evaluated different strategies of the prediction of the control flow with the compromise between seemingly incompatible goals: accuracy and speed.
2. Adaptation of the first-level cache based on the dynamic characteristics of the running application.
3. Evaluation of several alternatives to current DRB definition.

1.4 Applications

Three applications demanding high performance computing are investigated: 1) Analysis of DES (Distributed Event Simulation) applied to Individual oriented Models (IoM) in the field of biology/ecology; 2) Design of the Video-on-Demand (VoD) service, based on distributed systems; 3) Use of advanced simulation techniques to face the problem of forest fires propagation simulation using parallel/distributed systems.

Simulation of complex systems in the field of biology/ecology

This research line covers the application of Distributed Event Simulation (DES) to Individual oriented Models (IoM) in the field of biology/ecology, specifically to the Fish Schools (fish species movement simulation).

The goals of this research line can be summarized in the following points:

1. Create a three dimensional model of Fish School
2. Efficient and scalable use of HPC for Distributed Event Simulation of the Fish School
3. Create a tool for biologist to study the behaviour of Fish Schools by simulation
4. Elaborate a methodology to solve similar problems using the Individual oriented Models

Video-on-Demand (VoD) service

Developing a new P2P architecture based on the use of proxies and multicast communications, to improve the performance of the P2P scheme based unicast forward or application level multicast (ALM), which are limited by the high network resource requirement of the clients' communications.

Advanced simulation techniques of forest fires propagation

The goals of this research line can be summarized in the following points:

1. Design and development of new methods to improve the prediction of forest fire propagation.
2. Application of the new methods on "prescribed real fires" on a post-mortem phase.
3. Application of the new methods on real time.
4. Design and development of a DSS (Decision Support System) prototype.

2 Description of the Project Achievements

To describe the Project results we will follow the structure of the four basic areas.

2.1 Environment for performance evaluation and applications tuning.

The achievements of this research line can be summarized in the following points:

1. Design and development of a static performance analysis tool.
The design and development of a static automatic performance analysis tool (Kappa-Pi 2 Knowledge-based Automatic Parallel Program Analyzer for Performance Improvement) has been completed. The tool is open and allows the inclusion of new performance bottlenecks. The specification read is used to build a decision tree that is used in the automatic performance analysis phase to identify the occurrence of bottlenecks and generate suggestions concerning the improvement of the behaviour of the application.

2. Design and development of a dynamic performance tuning tool.
A dynamic automatic performance tuning environment (MATE Monitoring Analysis and Tuning Environment) integrating an analyzer module that loads the information concerning performance bottlenecks, has been developed. The performance bottleneck information is stored in a component called "Tunlet" which contains information concerning the required measurements to evaluate the behaviour of the application, the performance model to determine the optimal behaviour and the tuning actions that should be done improve the behaviour. A "Tunlet" specification language has been developed to allow the user to develop new "tunlets" corresponding to new performance bottlenecks.

3. Dynamic tuning of applications developed using a programming framework for parallel/ distributed applications specification
The M/W pattern has been studied and the performance parameters have been identified. The "tunlets" corresponding to the performance behaviour of the M/W pattern have been developed using the "tunlet" specification language and they have been integrated in the MATE environment. Currently, the pipeline pattern is being studied.

4. Dynamic tuning of applications based on mathematical libraries.
Mathematical libraries such as PETSc and SCALAPACK, commonly used in parallel application development include several parameters that cannot be specified before the execution of the application since they depend on the input data. A set of tunlets that dynamically tune such applications has been developed. The first step in this work has been to determine the performance models related to different mathematical functions and the effect of different parameters on the application behaviour.

5. Dynamic tuning of parallel applications on grid environments.
Grid environments are very relevant nowadays but they are very dynamic too, and it is necessary to adapt the application to the actual state of the environment using the dynamic tuning approach. Therefore, the first step that we tackled was the development of a monitoring environment based on dynamic instrumentation that is able to monitor parallel applications executed on a grid environment.

6 Task mapping and performance prediction in distributed systems
An application model to represent parallel applications with task and data parallelism that execute for a continuous input data stream has been developed. We developed a new task mapping algorithm to optimize the latency and throughput for these applications running in a pipeline fashion. The mapping strategy was integrated into the commercially available simulator DIMEMAS, and its efficiency was tested for message-passing applications. Regarding to prediction, we decided for a simulation based prediction process. In this context, we developed a

characterization methodology to extract the parameters of the application behaviour that influence on the execution and that are the most relevant for providing accurate predictions.

2.2 Strategies and tools for distributed systems management

Four different approaches in parallel/distributed computing have been developed

Grid environments

Main achievements of this research line can be summarized as follows:
1. A job-management service has been developed that provides a reliable on-line and batch MPICH-G2 submission to a Grid. It uses agents to take control of remote machines, allowing the implementation of backfilling scheduling policies for sequential jobs, while all the MPICH-G2 application subjobs are waiting for the proper co-allocation of resources. This service guarantees execution without blocking machines, and takes the appropriate decisions in order to guarantee resubmission of failed parallel jobs (due to crashes or failures with the network connection, resource manager or remote resources) and exactly-once execution. The job-management service also provides scheduling support for parallel applications that can be specified as a workflow.

2. We have also developed a solution to support interactive application on a Grid in which no changes to user code are required. Our tools are based on the idea of split execution systems, in which an agent is placed between the application and the operating system and traps some of its input/output system calls. The agent routes the trapped calls to a shadow process that executes these calls on the home machine. With this simple execution model, we are able to run existing, untouched, executable programs and hide all of the implementation details. We have applied interposition agents to sequential and MPI applications, incurring a minimal overhead for the whole system.

3. We have analyzed five well-known heuristics (min-min, max-min, sufferage, HEFT and random) when used as static and dynamic scheduling strategies in a grid environment in which computing resources exhibit congruent performance differences. The analysis shown that non-list based heuristics are more sensitive than list-based heuristics to inaccuracies in timing information. Static list-based heuristics perform well in the presence of low or moderate inaccuracies. Dynamic versions of these heuristics may be needed only in environments where high inaccuracies are observed. Our analysis also shows that list-based heuristics significantly outperform non-list based heuristics in all cases, and therefore they constitute the most suitable strategies to schedule workflows either statically or dynamically.

4. Finally, we have developed TDP-Shell, a generic framework that is able to deal with a wide range of different run-time tools and resource managers. TDP-Shell uses a simple and easy notation mechanism, similar to the one exhibited by most OS shells, to specify the interactions between the run-time tool and the user application when executed by a given resource management system. TDP-Shell is based on two agents that have little impact on the normal execution of the application and introduce minimum overhead (mostly at the application launching time).

Dedicated cluster based on user-shared machines

Several studies have revealed that a high percentage of computing resources (CPU and memory) in a Network of Workstations are idle. With the aim of taking advantage of the idle computer

resources available across the cluster (Network of user-shred Workstations), a new resource management environment, called CISNE, has been developed which combines space sharing and time sharing scheduling techniques. CISNE is set up basically by a dynamic coscheduling technique and a job scheduler.

Geographically distributed dedicated clusters

Main achievements of this research line can be summarized in the following points:
1. A system architecture was developed to provide transparency, scalability and to surpass the inter-cluster communications problems. An analytical model, including the multi-cluster system and the application features, was developed to support the efficiency evaluation, providing the basis for the performance prediction methodology and a guide for the system tuning actions.

2. The performance prediction methodology aims to guide to the analysis of the application and selection of the multi-cluster resources in order to guarantee the execution time reduction within a certain level of efficiency. The methodology also gives the support to the application tuning in order to increase the efficient speedup through the multi-cluster.
The multi-cluster experiments were implemented using clusters from Catholic University of Salvador (Brazil), National University "General Sarmiento" (Argentina) and University Autonoma of Barcelona (Spain).

3. The Fault Tolerant requirement in clusters was implemented developing two different alternatives:
a Fault-Tolerant protection by Data Replication (FT-DR), based on preserving critical functions by on-line dynamic data replication, oriented to MW applications
b RADIC (Redundant Array of Distributed Independent Checkpoints) which is a fully distributed functional architecture, based on rollback-recovery protocol, developed to implement fully transparent fault-tolerance in clusters by using an uncoordinated checkpoint protocol associated with pessimistic message log.

Web-computing

In order to fulfil the proposed objectives an initial prototype has been designed and developed using Java Technologies. This prototype was designed with the purpose of "prove of the idea" and to show the possibilities of the distributed infrastructure

2.3 Multiprocessor architecture

The achievements of this research line can be summarized as follows:
1. For the problem of the prediction of the control flow we have proposed and evaluated different strategies. The general idea is to arrange and codify data more efficiently, so that frequent and simple predictions are faster and consume fewer resources, while less frequent predictions or those that require more information to achieve higher accuracies are slower. The proposals can be classified into three groups. First, we have used three mechanisms to augment prediction speed: way prediction, index prediction, and two-level prediction. Second, an efficient trace prediction algorithm, which enhances previous proposals in the literature, is used to increase prediction width. In both cases, prediction accuracy is not reduced and memory requirements are only slightly increased.

2. For the configuration of the firs-level cache we have proposed static and dynamic algorithms. The identification of patterns or phases inside the application is performed by means of a learning algorithm (K-means) based on basic block vectors. The association of each application phase to a particular configuration of the cache memory could be done off-line or at runtime. Finally, the gathered information can be used to select the best static cache configuration for a given workload, or a dynamic recognition stage can be implemented at runtime, requiring hardware support both to track program behaviour and to enable the on-the-fly reconfiguration of the cache.
Results obtained by simulation show average performance improvements of 12%, with an energy saving around 3%. Alternatively, it is possible to obtain energy and power savings averaging 25% with a performance reduction lower than 2%.

3. We have set up an Infiniband modeller based on OPNET and currently are modelling DRB over Infiniband/OPNET and over OPNET directly. This allows us to evaluate current and future developments with a real scalable platform.
The development of a model of the DRB router at hardware description level, the analysis of the delays and the calculation of the cycle time. In this task, we have defined router structure at block level and indeed the specification of each block

2.4 Applications

Simulation of complex systems in the field of biology/ecology

1. The distributed simulator, a tool for biologist to analyze the fish school behaviour, was designed with the purpose of showing the viability of the ideas and performance analyses. The developed prototype is totally functional and the verification and validation of the tool has been totally satisfactory.
2. Based on the Huth&Wissel's model used represent the interactions among fishes in a 2D space, we developed a generalized 3D model for modelling fish School behaviour, redefining the mathematical model.
3. We developed an analytical model of the simulator that allow to predict computational resources and simulation times

Video-on-Demand (VoD) service

1. We have developed a new P2P architecture that combines proxies and uses multicast communications for peer collaborations. We introduce the concept of n-peers to m-peers cooperative collaboration to achieve high client delivery efficiency as well as unbounded scalable client collaboration capacity.
2. Furthermore, the architecture selects peers by taking into account the underlying network infrastructure in order to avoid network saturation.
3. We compared the new design with video delivery architecture based on Patching and Chaining schemes in a synthetic network model as well as a real Ethernet network topology.

Advanced simulation techniques of forest fires propagation

The main project achievements can be summarized as follows:
1. Development of a simulation framework (S2F2M-Statistical System for Forest Fire Management) that allows analyzing in a statistical way the forest fire behaviour. S2F2M assumes forest fire propagation prediction as an inherent imprecise task due to different factors such as the

imprecision of the applied analytical model, the difficulty of measuring the exact values of the needed parameters, the error introduced due to the numerical methods applied to create the fire simulators and so on. To overcome these difficulties, S2F2M explores a large set of input parameter combinations (called scenarios) in order to obtain the fire propagation map for each scenario. Since the number of executions of the fire simulator for the S2F2M system to be useful is quite huge, S2F2M were developed under the master-worker program paradigm and executed on cluster platforms. The maps provided by S2F2M were not considered in an isolated way but they were combined by applying statistics methods in order to generate probabilistic propagations maps which will be further used as a fire propagation prediction tool. S2F2M was first validated by comparing its predictions against the classical fire prediction strategy where the propagation provided by a unique execution of a certain fire simulator was considered and the obtained results shown the goodness of the methodology.

2. Validation of the S2F2M system using real fires. In particular, the "real fires" used for this experimental part of the project consisted of a set of "prescribed real fires" performed in Gestosa, Portugal. The analysis and post-processing of these fires were done in a post-mortem process consisted of several steps going from the analysis of the recorded film in order to extract the position of the fire front at different instant times, to the translation of the environmental information provided by the meteorological stations into input parameter accepted by our prediction system. The comparison of the obtained results with the prediction results provided by: the classical prediction systems showed that in the 95% of the cases the S2F2M provides better predictions results.

3. Finally, S2F2M has been exploited as a basic DSS (Decision Support System). In particular, the statistical scheme used by S2F2M was alternatively applied to generate fire risk maps of the Mediterranean coast for the summer 2005. The resulting maps could be considered as the information provided by a DSS of level 1, where processed and organized information is provided to the expert user in order to help him/her to decide how to proceed, for example, in the design of fire preventing policies.

3 Results

The results of the research including, PhD Thesis, Master Thesis, International Projects and Research collaborations, and Publications, can be summarized in the following data:

- PhD thesis 9
- Master thesis 10
- Current Graduate students (Master and PhD students) 21 (From 8 countries)
- International Projects and Research collaborations 10
- Publications
  - Journals 10
  - Conferences 54

PhD Thesis

University Autonoma of Barcelona, Computer Architecture and Operating Systems Department

1. J. Jorba Esteve. "Análisis Automático del Rendimiento de Aplicaciones Distribuidas basado en la Especificación de Problemas". (2006).
2. E. César Galobardes. "Definition of Framework-based Performance Models for Dynamic Performance Tuning". (2006).
3. M. Hanzich. "Un Sistema de Planificación Temporal y Espacial para Clusters no Dedicados". (2006)
4. E. Argollo. "Performance prediction and tuning in a multicluster environment". (2006)
5. J. Rodrigues de Souza. "FTDR: Tolerancia a fallos, en clusters de computadores geográficamente distribuidos, basada en Replicación de Datos". (2006)
6. J.C. Moure. "Aumentando las Prestaciones en la Predicción de Flujo de Instrucciones". (2006)
7. Xiaoyuan Yang. "Un Sistema de Vídeo bajo Demanda a gran escala basado en la Arquitectura P2P con Comunicaciones por Multidifusión". (2006)
8. G. Bianchini. "Wildland fire Prediction based on Statistical Analysis of Multiple Solution". (2006) (European Mention)
9. D. J. Mostacchio. "Simulación de Altas Prestaciones para Modelos Orientados al Individuo." (2007)

Master Thesis

University Autonoma of Barcelona, Computer Architecture and Operating Systems Department

1. P. Caymes-Scutari. "Entorno de Desarrollo y Sintonización de Aplicaciones Master/Worker." (2005)
2. E. Fernández del Castillo. "CrossBroker: gestión de recursos en CrossGrid". (2005)
3. A Duarte. "RADIC: Redundant Array of Distributed Independent Checkpoints. (2005)
4. F. Bonàs Vega "P2Parasite. Procesamiento distribuido sobre redes Peer to Peer". (2005)
5. L. Santos. "Extending VoDsim, A Simulation Tool for VoD systems". (2005)
6. M. Torchinsky. "Un Sistema de Planificación de Trabajos en Entornos NOW no Dedicados". (2006)
7. J. Ll. Lérida. "Un Sistema de Planificación para Entornos Multiclusters no Dedicados". (2006)
8. G. Vera Rodríguez. "Diseño e implementación de una aplicación distribuida basada en tecnologías Web para la evaluación del modelo de cómputo oportunista". (2006)
9. J. Balladini. "Un Planificador de Canales Lógicos para un Servidor de VoD en Internet" (2006)
10. G. Costa. "Dynamic monitoring of Grid Applications". (2006).

International Projects and Research collaborations

1. "CROSSGRID: Development of Grid Environment for Interactive Applications" – Project Nº: IST-2001-32243 (EU Commission) 2002-2005
2. "Forest fire spread prevention and mitigation- SPREAD" (EVG1-2001-00027) 2002-2005
3. "Flood Forecasting Computed on Grid Infrastructures" – Project EST.EAP.CLG 981032 (NATO) 2004-2006
4. "Plataforma de servicios multimedia para centros educativos" (PROFIT FIT-330301-2004-1) R+D Project. Spanish Ministry of Industry, Tourism and Commerce. Partners: Telefónica I+D; OBSERVA; Generalitat de Cataluña., 2004-06
5. HiPEAC NoE (European Network of Excellence on High-Performance Embedded Architecture and Compilation) 2004-2008
6. "int.eu.grid - Interactive European Grid"- Project Nº: 031857 (EU Commission) 2006-2008
7. "MAP-IT" R+D Project. Spanish Ministry of Industry, Tourism and Commerce. Partner: Oryzon Genomics, 2006-07
8. HEWLETT-PACKARD Contracts ""Desarrollo de un framework de simulación temporal de redes de comunicación". 2007-2008
9. Research collaboration with the Universidad de Alcalá, the University of Zurich and the Joint Research Centre from Ispra to elaborate forest fire risk maps. The obtained maps were combined with the risk maps caused by human and meteorological factors to finally provide risk global maps of forest fire in the Mediterranean zone. The Ispra Joint Research Center diffused the final maps to the corresponding authorities of the involved countries.
10. Research collaboration with the Computer Science de la Universidad de Wisconsin in the fields of :
    a Grid computing
    b Static and Dynamic performance analysis tools.
    c Programming framework for parallel/ distributed applications specification

4 Publications

Journals (10)

A. Morajko, T. Margalef, E. Luque
Design and Implementation of a Dynamic Tuning Environment
Journal of Parallel and Distributed Computing (accepted) (2007).

A. Morajko, P. Caymes, T. Margalef, E. Luque
MATE: Monitoring, Analysis and Tuning Environment for Parallel/Distributed Applications
Concurrency and Computation: Practice and Experience, vol. 18, pp. 1-15 (2006).

E. Cesar, A. Moreno, J. Sorribes, E. Luque.
Modeling Master/Worker applications for automatic performance tuning.
Parallel Computing, 32, pp 568-589, (2006)

B. Abdalhaq, A. Cortés, T.Margalef, G. Bianchini, E.Luque

Between Classical and Ideal: Enhancing Wildland Fire Prediction using Cluster Computing
Cluster Computing, 9, pp.329-343, 2006

D. Mostaccio, C. Dalforno, R. Suppi, E. Luque.
Distributed Simulation of Large-Scale Individual Oriented Models.
Journal of Computer Science & Technology, Vol.: 6, no. 2. pp: 59-65, 2006

G. Bianchini, M. Denham, A. Cortés, T. Margalef, E. Luque.
"Improving Forest-Fire Prediction by Applying a Statistical Approach"
Forest Ecology and Management, Vol. 234 (Supplement 1), pp. S210, 2006

Baker Abdalhaq, Ana Cortés, Tomás Margalef, Emilio Luque.
Enhancing wildland fire prediction on cluster systems applying evolutionary optimization techniques.
Future Generation Computer Systems, 21, pp. 61-67. 2005

M. Hanzich and F. Giné and P. Hernández and F. Solsona and E. Luque.
What To Consider For Applying Back_lling On Non-Dedicated Environments.
Journal of Computer Science & Technology, vol. 5, pp. 189-195. 2005.

G. Bianchini, A. Cortés, T. Margalef, E. Luque, E. Chuvieco, A. Camia.
Wildland Fire Risk Maps using S2F2M
Journal of Computer Science & Technology (JCS&T) - Special Issue Selected Papers from CACIC 2005, Vol. 5 No. 4.pp. 244-249. 2005

A. Morajko, E. César, P. Caymes-Scutari, J.G. Mesa, G. Costa, T. Margalef, J. Sorribes, E. Luque
Development and Tuning Framework of Master/Worker Applications.
Journal of Computer Science & Technology, vol.5, no. 3, pp 115-120. 2005.

Conferences (54)

F. Guirado, A. Ripoll, C. Roig, A. Hernández, E. Luque.
Exploiting Throughput for Pipeline Execution in Streaming Image Processing Applications.
Proc. Euro-Par 2006. LNCS. Vol. 4128. Pp: 1095-1105. 2006.

V. Ivars, A. Cortés, M. A. Senar.
"TDP_SHELL: An Interoperability Framework for Resource Management Systems and Run-time Monitoring Tools"
Proc. Euro-Par 2006, Lecture Notes in Computer Science, Vol. 4128, pp. 15-24, 2006.

E. Fernández, E. Heymann, M. A. Senar.
"Supporting Efficient Execution of MPI Applications Across Multiple Sites"
Proc. Euro-Par 2006, Lecture Notes in Computer Science, Vol. 4128, pp. 383-392, 2006.

M. Hanzich, J. LL. Lérida, M. Torchinsky, F. Giné, P. Hernández and F. Solsona, E. Luque.
Using On-The-Fly Simulation For Estimating the Turnaround Time on Non-Dedicated Clusters.
Proc. EuroPar 2006, Lecture Notes in Computer Science, Vol. 4128, pp. 177-188. 2006

E. Argollo, A. Gaudiani, D. Rexachs, E. Luque
"Tuning Application in a Multi-cluster Environment."
Proc. Euro-Par, 2006, Lecture Notes in Computer Science, Volume 4128, pp. 78-88. 2006

X.Y. Yang, P. Hernández, F. Cores, L. Souza, A. Ripoll, R. Suppi, E.Luque.
Providing VCR in a Distributed Client Collaborative Multicast Video Delivery Scheme.
Proc. EuroPar 2006, Lectures Notes on Computer Science, Volume 4128, pp. 777-787. 2006
Distinguished Paper.

G. Costa, A. Morajko, T. Margalef, E. Luque
Automatic Tuning in Computacional Grids
Proc. PARA'06, LNCS (Pending of publication) (2006).

J. Jorba, T. Margalef, E. Luque
Search of Performance Inefficiencies in Message Passing Applications with KappaPI 2 Tool
Proc. PARA'06, (Pending of publication) (2006).

G. Costa, A. Morajko, T. Margalef, E. Luque
Process Tracking for Dynamic Tuning Applications on the Grid
Proc. CACIC'06, Universidad de San Luis, Argentina, 2006.

P. Fritzsche, C. Roig, A. Ripoll, E. Luque, A. Hernández.
A Performance Prediction Methodology for Data-dependent Parallel Applications.
IEEE Proc. Int. Conf. on Cluster Computing (Cluster 2006). CD-ROM proceedings. Oct. 2006.

M. M. López, E. Heymann, M. A. Senar.
"Analysis of Dynamic Heuristics for Workflow Scheduling on Grid Systems",
Proc. IEEE Int. Symposium on Parallel and Distributed Computing (ISPDC 06), pp.199-207, 2006

E. Fernández, E. Heymann, M. A. Senar
"Resource Management for Interactive Jobs in a Grid Environment",
Proc. of IEEE Int. Conf. On Cluster Computing (Cluster 2006), CD-ROM, 2006.

J. LL. Lérida, F. Solsona, F. Giné, M. Hanzich, p. Hernández, E. Luque.
" MetaLoRaS: A Predictable MetaScheduler for Non-Dedicated Multicluster".
Proc. Int. Sym. ISPA 2006, Lecture Notes in Computer Science, Vol. 4330, pp. 630-641. 2006

M. Torchinsky , M. Hanzich, F. Giné, F. Solsona J. LL. Lérida, P. Hernández, E. Luque.
"CISNE P- A Portable and Integral Approach for Scheduling Parallel Jobs on Non-Dedicated Clusters".
Proc. CACIC 2006, Workshop de Procesamiento Distribuido y Paralelo (WPDP), Argentina, 2006

M. Hanzich, F. Giné, P. Hernández, F. Solsona, J. LL. Lérida, E. Luque.
Using Simulation, Historical and Hybrid Estimation Systems for Enhancing Job Scheduling on NOW´s.
Proc.Cluster 2006, CD ROM IEEE-ACM. 2006.

E. Argollo, D. Rexachs, F.G. Tinetti, E.Luque
"Efficient Execution of Scientific Computation on Geographically Distributed Clusters"
Proc. PARA 2004, Lecture Notes in Computer Science Volume 3732, pp. 691-698. 2006

A. Duarte, D. Rexachs, E. Luque
"An Intelligent Management of Fault Tolerance in Cluster Using RADICMPI".
Proc. Euro PVMMPI, Lecture Notes in Computer Science, Volume 4192; pp. 150-157. 2006

J. C. Moure, D. Benítez, D. I. Rexachs, and E. Luque.
Wide and Efficient Trace Prediction using the Local Trace Predictor.

Proc. of the 20th ACM Intl. Conf. on Supercomputing (ICS06), pp. 55-65. 2006.

D. Benítez, J. C. Moure, D. I. Rexachs, E. Luque
A Reconfigurable Data Cache for Adaptive Processors.
Proc. ARC-2006, Lecture Notes on Computer Science, Volume3985, pp. 230–242, 2006.

D. Benítez, J.C. Moure, D. I. Rexachs, E. Luque.
Evaluation of the Field-Programmable Cache: Performance and Energy Consumption.
Proceedings of the 3rd ACM Int. Conf. on Computing Frontiers (CF'06), pp. 361-372, 2006

J. Balladini, L. Souza, R. Suppi.
A Network Traffic Scheduler for a VoD Server on the Internet
Proc. Int. Conf. on Signal Processing and Multimedia Applications (SIGMAP 2006), 2006.

X.Y. Yang, P. Hernández, F. Cores, L. Souza, A. Ripoll, R. Suppi, E. Luque.
LDPn2Pn: Distributed P2P assisted multicast VoD architecture.
Proc. IEEE Int. Parallel & Distributed Processing Symposium (IPDPS'06), CD ROM, 2006.

X.Y. Yang, P. Hernández, F. Cores, L. Souza, A. Ripoll, R. Suppi, E. Luque.
Multi-collaboration domain multicast P2P delivery architecture for VoD system.
Proc. IEEE Int. Conference on Communications (ICC'06), 2006.

G. Bianchini, A. Cortés, T. Margalef, E. Luque.
"Improved Prediction Methods for Wildfires Using High Performance Computing: A Comparison"
Proc ICCS 2006, Lecture Notes on Computer Science, Volume 3991, pp. 539-546. 2006.

G. Bianchini, M. Denham, A. Cortés, T. Margalef, E. Luque
"Improving Forest-Fire Prediction by Applying a Statistical Approach"
Proc. V International Conference on Forest Fire Research, 2006 (CD-ROM)

A. Morajko, P. Caymes, T. Margalef, E. Luque
Automatic Tuning of Data Distribution Using Factoring in Master/Worker Applications
International Conference on Computational Science, LNCS, 3515, pp. 132-139. 2005.

A. Morajko, E. César, P. Caymes-Scutari, T.Margalef, J. Sorribes, E.Luque
Automatic Tuning of Master/Worker Applications
Proc. EuroPar 2005. Lecture Notes in Computer Science, 3648, pp. 95-103.2005.

E. Cesar, A. Moreno, J. Sorribes, E. Luque.
Modeling Pipeline Applications in POETRIES
Proc. EuroPar 2005. Lecture Notes in Computer Science, 3648, pp 83-92. 2005

J. Jorba, T. Margalef, E. Luque
Automatic Performance Analyisis of Message Passing Applications using the KappaPi 2 tool
Proc. EuroPVM/MPI 2005. Lecture Notes in Computer Science, 3666, pp. 293-300. 2005

J. Jorba, T. Margalef, E. Luque
Performance Analysis of Parallel Applications with KappaPi 2
Proc. Parco 2005, NIC Series Vol. 33, pp. 155-162. 2005

A. Moreno, E. César, J. Sorribes, T. Margalef, E. Luque
Balanceo de carga en Sistemas distribuidos Master-Worker

Proc. XVI Jornadas de Paralelismo, Granada. 2005

P. C. Fritzsche, J. J. Fernández, A. Ripoll, I. García, E. Luque.
A Performance Prediction Model for Tomographic Reconstruction in Structural Biology.
Proc. VECPAR 2004. Lecture Notes on Computer Science Volume 3402. pp: 90-103. 2005.

F. Guirado, A. Ripoll, C. Roig and E. Luque.
Exploitation of Parallelism for Applications with an Input Data Stream: Optimal Resource-Throughput Tradeoffs.
Proc. IEEE 13th Parallel, Distributed and Network-based Processing (PDP).pp:170-178. 2005.

P. C. Fritzsche, A. Ripoll, E. Luque, J. J. Fernández, I. García.
A Performance Prediction for Iterative Reconstruction Techniques on Tomography.
Proc. IEEE 13th Parallel, Distributed and Network-based Processing (PDP). pp: 92-99. 2005

F. Guirado, A. Ripoll, C. Roig, E. Luque.
Optimizing Latency under Throughput Requirements for Streaming Applications on Cluster Execution.
Proc. IEEE Int. Conf. on Cluster Computing (Cluster 2005). CD-ROM proceedings. 2005.

A. Morajko, E. Fernández, A. Fernández, E. Heymann, M. A. Senar.
"Workflow Management in the CrossGrid Project"
Proc. European Grid Conf., Lecture Notes on Computer Science Vol. 3470, pp. 424-433, 2005.

J. Gomes et alter, "Experience with the International Testbed in the CrossGrid Project"
Proc. European Grid Conf., Lecture Notes on Computer Science Vol. 3470, pp. 98-110, 2005.

M. Hanzich; F. Giné, P. Hernández, F. Solsona, E. Luque
A Space and Time Sharing Scheduling Aproach for PVM Non-Dedicated Clusters.
Proc. EuroPVM/MPI 2005, Lecture Notes in Computer Science, Volume 3666, pp. 379-387. 2005.

M. Hanzich and F. Giné and P. Hernández and F. Solsona and E. Luque.
CISNE: A New Integral Approach for Scheduling Parallel Applications on Non-Dedicated Clusters.
Proc. EuroPar 2005. Lecture Notes in Computer Science, Volume 3648, pp. 220-230. 2005.

J. Rodrigues de Souza, E. Argollo, A. Duarte, D. Rexachs, E. Luque
"Fault Tolerant Master-Worker over a Multi-Cluster Architecture".
Proc. ParCo 2005. NIC Series vol. 33. pp. 465-472. 2005

A. Duarte, D. Rexachs, E. Luque
"A Distributed Scheme for Fault-Tolerance in Large Clusters of Workstations".
Proc. ParCo 2005. NIC Series vol. 33. pp. 473-480; 2005

J.C. Moure, D. Benítez, D. Rexachs, E. Luque.
Target Encoding for Efficient Indirect Jump Prediction.
Proc. Euro-Par 2005. Lecture Notes on Computer Science, Volume 3648, , pp. 497 -507, 2005.

D. Benítez, J. C. Moure, D. I. Rexachs, E. Luque.
Performance and Power Evaluation of an Intelligently Adaptive Data Cache.
Proc. HiPC-2005. Lecture Notes on Computer Science, Volume 3769, pp. 363–375. 2005.

D. Mostaccio, R. Suppi, E. Luque.

Simulation of Ecologic Systems Using MPI.
Proc. Euro PVMMPI. Lecture Notes in Computer Science, Vol.: 3666, pp: 449-456. 2005

D. Mostaccio, R. Suppi, E. Luque.
Distributed Simulation of Ecologic Systems
Proc. XVI Jornadas de Paralelismo, Vol.: I, pp: 671-676, 2005

X.Y. Yang, P. Hernández, F. Cores, A. Ripoll, R. Suppi, E. Luque.
Distributed P2P merging policy to decentralize the multicasting delivery.
Proc EuroMicro Conference on Software Eng. and Advanced Applications, pp. 322–329, 2005.

X.Y. Yang, P. Hernández, F. Cores, A. Ripoll, R. Suppi, E. Luque.
Dynamic distributed collaborative merging policy to optimize the multicasting delivery scheme.
Proc. Euro-Par 2005. Lecture Notes on Computer Science, Volume 3684, pp. 879–889. 2005.

L. Santos, A. Ripoll, E. Luque.
Designing a Large-scale Distributed VoD system utilizing the VoDSim
Proc. IEEE International Conference on Distributed Computing Systems, p. 46, ICDCSW, 2005.

G. Bianchini, A. Cortés, T. Margalef and E. Luque.
S2F2M - Statistical System for Forest Fire Management
Proc. ICCS 2005. Lecture Notes on Computer Science, Volume 3514, pp. 427-434. 2005.

E. Chuvieco, A. Camia, G. Bianchini, T. Margalef, N. Koutsias and J. Martínez.
Using remote sensing and GIS for global assessment of fire danger
Proc XXII International Cartographic Conference (ICC2005)

G. Bianchini, A. Cortés, T. Margalef, E. Luque, E. Chuvieco, A. Camia.
Wildland Fire Propagation Danger Maps Based on Factorial Experimentation
Information Technologies in Environmental Engineering (ITEE'2005) pp.173-185. 2005.

G. Bianchini, A. Cortés, T. Margalef, E. Luque.
S2F2M – Sistema Estadístico para la Predicción de Incendios Forestales.
Proc. I Congreso Español de Informática (CEDI 2005), pp.623-629. 2005

G. Bianchini, A. Cortés, T. Margalef, E. Luque, E. Chuvieco, A. Camia.
Mapas de Riesgo de Incendios Forestales Basados en Experimentación Factorial
Proc.CACIC'05, Universidad de Entrerios, CD-ROM, 2005

P. Caymes-Scutari, A. Morajko, E. César, J.G. Mesa, G. Costa, T. Margalef, J. Sorribes, E. Luque
Entorno de Desarrollo y Sintonización de Aplicaciones Master/Worker
Proc CACIC'05, Universidad de Entrerios, CD ROM, 2005

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

TIN2004-03453: Dynamic Weaving Aspect System over a Reflective Platform

Juan Manuel Cueva[*]
Main Researcher

Francisco Ortín[**]
Research Member

Abstract

This project is aimed at investigating how different reflective technologies can be used to develop a dynamically weaving aspect-oriented computing system, without any dependency of a concrete programming language, built over a heterogeneous computing platform.

An abstract machine with a reduced instruction set has been employed as the root computation system's engine; it offers the programmer basic reflection computation primitives. Its reduced size and its introspective capabilities, make it easy to be deployed in heterogeneous computational systems, becoming a platform-independent computational system.

By using the reflective features offered by the abstract machine, the running applications can be adapted and extended. This would be programmed on its own language, without needing to modify the virtual machine's source code and, therefore, without loosing code portability. As an example of its extensiveness, new software aspects can be programmed achieving facilities such as persistence, distribution, logging or trace. All this new abstractions are adaptable at runtime to any application.

By using a reflective language-neutral computing platform, a framework has been created to develop dynamic weaving aspect scenarios. No needing to modify application's functional code, new crosscutting concerns will be weaved to any program at runtime. Following the same scheme, when this dynamic and language-neutral aspects are needed no more, they could be suppressed at runtime in a programmatically way –i.e., not only a human could adapt an application at runtime but another program or even itself could do it.

Keywords: Reflection, virtual machine, aspect-oriented programming, dynamic weaver.

1 Project Objectives

In many cases, significant concerns in software applications are not easily expressed in a modular way. Examples of such concerns are transactions, security, logging or persistence. The code that addresses these concerns is often spread out over many parts of the application. It is commonly tangled with the application logic and, therefore, it is more difficult to maintain, understand and reuse.

In the Software Engineering area, the principle of separation of concerns, SoC [1],[2] has been used to manage the complexity of software development; it separates main application algorithms from special purpose concerns (typically orthogonal to the main functionality) –e.g. authentication, logging, memory management, performance, etc. Final applications are built by means of its main functional code plus their specific problem-domain concerns.

[*] Email: cueva@uniovi.es
[**] Email: ortin@uniovi.es

This principle has been performed following several approaches where Aspect Oriented Software Development (AOSD) [3] is one of them. The main AOSD aim is the construction of concern-adaptable programs. Most existing tools lack adaptability at run-time: once the final application has been generated (woven), it will not be able to adapt its concerns (aspects) at runtime. This is known as static aspect weaving. There are certain cases in which the adaptation of application concerns should be done dynamically, in response to changes in the runtime environment –e.g., distribution concerns based on load balancing. Another example where dynamic AOSD can be used is what has been recently called autonomic software: software capable of being itself repaired, managed, optimized or recovered.
To overcome the static-weaving tools limitations, different dynamic-weaving AOSD approaches have appeared (e.g. PROSE [4]). However, they limit the set of join points they offer (in comparison with static weaving tools), restricting the way applications can be adapted at runtime. In addition, both static and dynamic weavers use fixed programming languages, so aspects and concerns are not reusable regardless of its programming language.
Reflection is a programming language technique that achieves dynamic application adaptability. It can be used to reach aspect adaptation at runtime. As a previous research work [5][6][7], we have applied reflection (computational) to a virtual machine, obtaining a non restrictive program adaptation at runtime regardless the language and platform.
We have applied the concepts we have learned to the language and platform independent Microsoft .NET platform, which offers good runtime performance at the same time as being language and hardware neutral. We have developed a reflective treatment of .NET applications offering a dynamic weaving framework.

2 Project Status

In order to accomplish the objectives described in the previous section, we have defined two different approaches: code instrumentation and virtual machine extension.

1. Code Instrumentation. Taking a binary application, we inject new code in order to inject join points of a dynamic AOSD framework. The main benefit of this approach is that the runtime is the original one, i.e. it is not necessary to install a new platform.
2. Virtual machine extension. Taking a commercial virtual machine implementation, we have added structural reflection primitives in order to support dynamic weaving features in a native way. This approach is less standard but it offers a higher runtime performance.

2.1 Code Instrumentation

The system architecture is shown in Fig. 1 and Fig. 2. Before a compiled application is executed, it is processed and altered by our system. In order to offer computational reflection, the join point injector (JPI) inserts MOP-based computational reflection routines following the next steps:

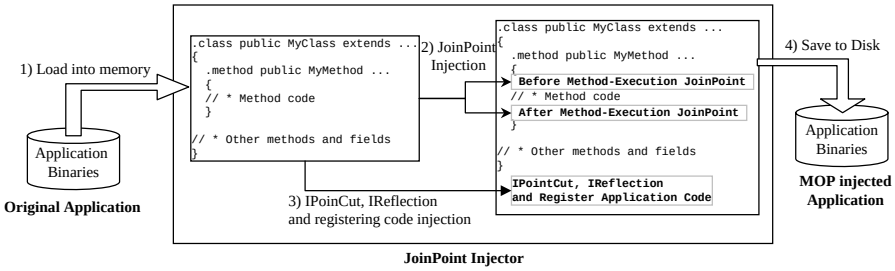


Fig. 1 System Architecture.

1. Application is loaded into memory by means of the System.Reflection namespace. From the application binaries, this namespace is used to load the program MSIL representation into memory. We will then manipulate this code, as it was data.
2. Control routines injection by means of the System.Reflection.Emit namespace. Conditional routines to call future aspect's advice are injected in every join point. At first, this code does not invoke any routine, but modifying its condition later on, its behavior can be extended – as happens in most MOP reflective systems.
   IPointCut and IReflection interfaces and application-registering code injection. IPointCut provides the interface to define different join points to be employed in future adaptation. This functionality is the same as the one offered by AspectJ's pointcut designators.
   By using IPointCut, aspects may select a range of specific join points of an application to be called back each time the join point is reached (at method invocation, field access, or whatever). This aim is achieved modifying conditions injected in the second point, assigning them true values and, therefore, causing aspect advice invocation.
3. IReflection interface allows other applications to access program's information (intra-application introspection) –for security reasons, in .NET platform an application can only inspect itself; it cannot inspect another one. When aspects are being developed, it is commonly needed to access the structure of the application to be adapted: that is exactly the functionality offered by IReflection.
   This interface has been developed using mainly the System.Reflection namespace, permitting that aspects could gain access to application's information, invoking a method or even creating an object. Finally, necessary code to achieve application registration itself in the server at start up is injected –explained afterwards. Finally, necessary code to achieve application registration itself in the server at start up is injected –explained afterwards.
4. Save to disk the modified application. The modified program binaries are finally saved to disk.

At start up the modified application will register itself in the server (application registry) with a unique GUID, allowing aspects interact with it. When an aspect needs to adapt an application, the following steps are followed:

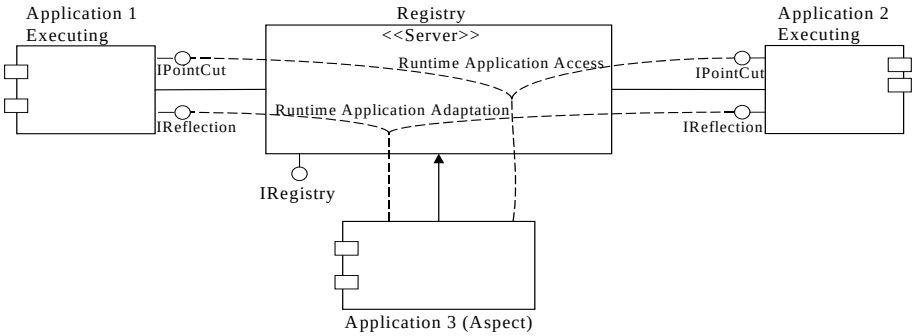


Fig. 2 Dynamic Application Adaptation.

1. The aspect accesses to the server and informs him about which joinpoints of the target application (the one to be adapted) is interested in –in order to be notified when these joinpoints are reached.
2. The server, by means of the IPointCut interface, requests the target application to notify him when the specific joinpoints are reached. What this interface does internally is to modify the appropriated conditions previously injected. So, when the join points are reached, the condition is satisfied and the server will be notified.
   With these first two steps, the aspect has designated the pointcuts.
3. When any of the requested joinpoints are reached in the target application, the server is notified (by the target application) of this, and receives a reference to the application.
4. When the server receives a notification, it informs the aspects that have requested it, and pass them the reference. The server implements the Observer design pattern, to achieve this behavior.
5. By means of this reference and the IReflection interface implemented by the target application, the aspect can access the target application, obtain its structure, invoke target application's code, or running its own routines.
6. If an aspect is no more needed, it informs the server in order to not been notified any more –when the joinpoint would be reached in the target application.
7. If the server detects that no aspects left has to be notified in the occurrence of a joinpoint, informs the target application of this, and, using the IPointCut interface, modifies necessary conditions to false values. This would obtain better run-time performance.

As we do all our work in code written in intermediate language, MSIL, we obtain two main benefits: language independence and platform independence in both the application to be adapted and the aspects that adapt it. In addition, it is not necessary at design time to know if the application is going to be adapted or not.

With our system, it is also possible to unweave (delete) an aspect when it is necessary no more. A simple example is execution time metrics: 1) when we need to measure something in a running application we will develop an aspect to do it, 2) once the aspect has been executed, the application is adapted showing the metrics, and 3) when the metric is no more needed, we can unweave the aspect to make the application returns to its initial state.

Another important benefit we achieve is that we do not need the source code of an application to modify it, as all the process is done from binary code that is translated to MSIL. Furthermore, we do not modify the abstract machine so we can use our system with any standard .net application.

2.1.1 System Benefits

Our research work states that applying computational reflective techniques to a platform and language virtual machine can be used to implement a dynamic aspect weaver with a rich set of join points. Using the .net platform, this system has the following benefits:

- Language independence. As we do all the process over intermediate language (MSIL) both the application that can be adapted and the aspects that can adapt an application can be written in any language, as the weaving process is done only with the code translated into MSIL.
- Aspects and core functionality follow full independent lifecycles. Thanks to this we obtain a greater reusability of both, and it is easier to debug a program because the code it is not tangled.
- Platform independence. By doing all the process at the virtual machine level we obtain complete independence from the platform over it is implemented.
- We do not need source code to adapt an application. As we work only with MSIL, there is no need to use the source code. This would be a very useful feature if we have to modify code from third parties.
- Use of a standard virtual machine. Our system does not need to modify the virtual machine, so it is ECMA compliant. Therefore, we can introduce any CLR standard application into our system.
- The system offers a rich set of join points. The mechanism identified in our system does not limit the range of join points to capture in any application. Therefore, it can be developed aspects the same way as in AspectJ, but weaving them at runtime.
- Finally, it is possible to adapt an aspect by means of another aspect. Aspects are common applications that work in our system, so they follow the same steps than other applications and could be adapted the same way.

2.2 Virtual Machine Extension

This second approach is still under development, but we have concluded the majority of its work. In order to overcome the runtime performance drawback of the previous approach, this work has been focused on applying JIT compilation techniques to optimize structural reflection primitives used in our dynamic AOSD system.

Compiling languages to a virtual machine's intermediate code offers many benefits such as platform neutrality, compiler simplification, application distribution, direct support of high-level paradigms and application interoperability [8]. In addition, compiling languages to a lower abstraction level virtual machine improves runtime performance in comparison with direct interpretation of programs. Therefore, we have used the Microsoft .Net platform as the targeted virtual machine to benefit from all the advantages mentioned above. The main reason why we have selected the .Net abstract machine was its design focused on supporting a wide number of languages [9, 10]. Moreover, extending the .Net platform to support structural reflection facilitates future interoperation between these languages and any .Net application or component.

Microsoft SSCLI, also known as Rotor, is a source code distribution that we have taken to develop our project. It includes fully functional implementations of the ECMA-334 C# language standard and the ECMA-335 Common Language Infrastructure specification, various tools, and a set of libraries suitable for research purposes [11]. The source code can be built and run under Windows XP, FreeBSD 4.5 or Mac OS X.

The following is a summary of the most significant reflective primitives added to the System.Reflection.Structural namespace (all of them, static methods of the NativeStructural utility class):

- The {add, remove, alter, get, exist}Method methods receive an object or class (System.Type) as a first parameter indicating whether we want to modify (inspect) a single object or a shared behavior. The second parameter is a MethodInfo object of the System.Reflection namespace. This object uniquely describes the identifier, parameters, return type, attributes and modifiers of a method. The IsStatic property of this object is used to select the schema evolution behavior or class (static) member adaptation.
  If the programmer needs to create a new method, she can generate it by means of the System.Reflection.Emit namespace, and add it later to an object or class using its MethodInfo.
- The invoke primitive executes the method of an object or class specifying its name, return type and parameters. If no reflection has been used (we are running "static" programs) a fast concatenation strategy is used. However, in the execution of reflective dynamic languages method invocation is based on delegation: when a message is passed to an object, it is checked whether the object has a suitable method or not; in case it exists, it is executed; otherwise, the message is passed to its class (its trait object) in a recursively way. A MissingMethodException is thrown if the message has not been implemented in the hierarchy.
- The {add, remove, alter, get, exist}Field methods modify the runtime structure of single objects (prototype model) or their common schema (classes or traits) passed as the first parameter. The second parameter is an instance of a new RuntimeStructuralFieldInfo class (derived from the .Net's FieldInfo class) that describes the field's type, visibility, and many other attributes. Once again, the Static attribute of the second parameter selects the schema evolution behavior (class-based and Python models) or class (static) member adaptation (class-based and Ruby semantics).

We have also modified the native code generated by the JIT compiler when executing the following CIL statements:

- ldfld, ldsfld and ldflda: Loads into the stack the (instance or static) field value (or address) following our computational model.
- stfld and stsfld: Stores a value into a (instance or static) field, deciding its appropriate memory location at runtime.
- call and callvirt: Method execution following both the concatenation and delegation inheritance strategies.

2.2.1 System Benefits and Future Work

The assessment of our reflective platform implementation has shown that our approach, in comparison with dynamic languages that offer structural reflection, is the fastest when running reflective tests and its memory consumption is appropriate. When executing static code, we are more than 12 times faster than our interpreter-based counterparts, requiring only 28% more memory resources. This benefit is due to design of the JIT compiler, aggressively optimized to generate non-reflective code. Finally, we have also evaluated the cost of our reflective enhancements. When running real applications that do not use reflection at all, it seems that the performance cost is near to 100%, using 7% more memory.

Future work will be developing services of dynamic AOSD using structural reflection, the same way we did in the previous approach (see Section 2.1).

3 Results

Main results of this project are PhD Thesis, publications and collaborations in other international project.

3.1 Publications

3.1.1 Journals

- Optimizing Reflective Primitives of Dynamic Languages. José M. Redondo, Francisco Ortín Soler, Juan M. Cueva. International Journal of Software Engineering and Knowledge Engineering. Accepted, to be published in 2007.
- Adding Structural Reflection to the SSCLI. Francisco Ortín Soler, José M. Redondo, Luis Vinuesa, Juan M. Cueva. Journal of .Net Technologies, Volume 3, Number 1-3, pp. 151-162. May 2005.
- Designing an Adaptable Heterogeneous Abstract Machine by means of Reflection. Francisco Ortín Soler, Diego Díez Redondo. Elsevier Information and Software Technology, Volume 47, Issue 2, pp. 81-94. February 2005.
- Separating Adaptable Persistence Attributes through Computational Reflection. Francisco Ortín Soler, Benjamín López, J. Baltasar G. Pérez-Schofield. IEEE Software, Volume 21, Issue 6. November 2004.
- Reflection as the basis for Developing a Dynamic SoC Persistence System. Benjamín López Pérez, Francisco Ortín Soler, Javier Noval Arango. Journal of Object Technology, Volume 3, Issue 8. September 2004.
- Dynamic Adaptation of Application Aspects. Francisco Ortín Soler, Juan Manuel Cueva Lovelle. Elsevier Journal of Systems and Software, Volume 71, Issue 3. May 2004.
- A Dynamic Aspect Weaver over the .Net Platform. Luis Vinuesa Martínez, Francisco Ortín Soler. Springer-Verlag Lecture Notes in Computer Science 3002. April 2004.

3.1.2 Conferences

- Designing Structural Reflection over the SSCLI (spanish). José M. Redondo, Francisco Ortín Soler, Juan M. Cueva. 11th Conference on Software Engineering and Databases (JISBD). Sitges (Spain). October 2006.
- Optimizing Structural Reflective features of Dynamic Languages (spanish). José M. Redondo, Francisco Ortín Soler, Juan M. Cueva. 6th Programming and Languages Conference (PROLE). Sitges (Spain). October 2004.
- Adding Structural Reflection to the SSCLI. Francisco Ortín Soler, José M. Redondo, Luis Vinuesa, Juan M. Cueva . International Conference in.Net Technologies, Pilzen (Check Republic), May 2005.
- Separating the Persistence Aspect by means of Computational Reflection (spanish). Benjamín López Pérez, Francisco Ortín Soler, Juan Manuel Cueva Lovelle. 9th Conference on Software Engineering and Databases (JISBD). Malaga (Spain). November 2004.
- Dynamic Aspect Weaving for .Net (spanish). Luis Vinuesa Martínez, Francisco Ortín Soler. Workshop on Aspect Oriented Software Development (DSOA), 9th Conference on Software Engineering and Databases (JISBD). Malaga (Spain). November 2004.
- A Practical Approach to Develop Type Checkers based on Design Patterns (spanish). Francisco Ortín Soler, Luis Vinuesa, Juan Manuel Cueva Lovelle. 4th Programming and Languages Conference (PROLE). Malaga (Spain). November 2004.

3.2 PhD Theses

- Benjamín López Pérez, "Dynamic Adaptation of Object Persistence by means of Computational Reflection ", June 2006.
- Luis Vinuesa Martínez, "Platform and Language-Neutral Dynamic Aspect Weaving by means of Computational Reflection". To be presented in June, 2007.
- José Manuel Redondo López, "Improving the performance of Structural Reflection by means of Just In Time Compilation". To be presented in March, 2007.

3.3 Collaborations in other International Projects

Part of this project (the "virtual machine extension" approach described in Section 2.2) has been developed in collaboration with Microsoft Research, under the project "Extending Rotor with Structural Reflection to support Reflective Languages" awarded in the second Microsoft Research Rotor Request for Proposals, in March 2004.

4 References

[1] Parnas, D., 1972. On the Criteria to be Used in Decomposing Systems into Modules. Communications of the ACM, Vol. 15, No. 12.
[2] Hürsch, W.L., Lopes, C.V., 1995. Separation of Concerns, Technical Report UN-CCS-95-03, Northeastern University, Boston, USA.
[3] Kiczales, G., Lamping, J., Mendhekar, A., Maeda, C., Lopes, C.V., Loingtier, J.M., Irwin, J., 1997. Aspect Oriented Programming. In: Proceedings of European Conference on Object-Oriented Programming (ECOOP), vol. 1241 of Lecture Notes in Computer Science.
[4] Popovici, A., Gross, Th., Alonso, G., 2001. Dynamic Homogenous AOP with PROSE, Technical Report, Department of Computer Science, ETH Zürich, Switzerland.
[5] Ortin, F., and Cueva, J. M. Building a Completely Adaptable Reflective System. 2001. European Conference on Object Oriented Programming ECOOP'2001. Workshop on Adaptive Object-Models and Metamodeling Techniques, Budapest, Hungary, June 2001.
[6] Francisco Ortin and Juan Manuel Cueva. Non-Restrictive Computational Reflection. Elsevier Computer Software & Interfaces. Volume 25, Issue 3, Pages 241-251. June 2003.
[7] Francisco Ortin, Juan Manuel Cueva. Implementing a real computational-environment jump in order to develop a runtime-adaptable reflective platform. ACM SIGPLAN Notices, Volume 37, Issue 8, August 2002
[8] Diehl, S., Hartel, P., and Sestoft, P. 2000. Abstract machines for programming language implementation. *Future Generation Computer Systems 16,* 7. 739—751.
[9] Meijer, E., and Gough, J. 2000. Technical Overview of the Common Language Runtime. *Technical Report.* Microsoft.
[10] Singer, J. 2003. JVM versus CLR: a comparative study. In *ACM Proceedings of the 2nd international conference on Principles and practice of programming in Java.* 167–169.
[11] Stutz, D., Neward, T., and Shilling, G. 2003. *Shared Source CLI Essentials.* O'Reilly.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Development of E-Services for the New Digital Society TIN2004-03534

Vicente Pelechano
Departamento de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia

Abstract

Software as a service is becoming more and more the new production and distribution model for applications using Internet. The Information Society should provide citizens and companies with a series of accessible services based on computer technologies that use the infrastructure that the Internet provides. At the moment, this kind of service is provided in an ad-hoc way. In this context, rigorous methods are necessary for allowing us to design and develop services that are accessible through the Internet (E-Services).
Keywords: Model Driven Development, Web Engineering, Web Services, Interoperability, Quality, Metrics, Domain Specific Languages, Smart Homes.

1 Introduction

Software as a service is becoming more and more the new production and distribution model for applications using Internet. The Information Society should provide citizens and companies with a series of accessible services based on computer technologies that use the infrastructure that the Internet provides. At the moment, this kind of service is provided in an ad-hoc way, with a scarce methodological support and in an unproductive way. In this context, rigorous methods are necessary for allowing designing and developing services that are accessible through the Internet. We should provide: (1) models and conceptual modelling primitives that are necessary for the requirements elicitation of this kind of applications, where service usability and personalization is going to be a decisive factor for its full acceptance; (2) service-oriented software architectures that constitute the infrastructure to develop E-services; (3) transformation processes that allow systematic generation of software products, and finally (4) metrics and estimation mechanisms that allow us to evaluate the quality of the developed E-services.
In this report, we present an insight into the current development of the DESTINO (TIN2004-03534) project. In Section 2, we summarise our goals and the tasks involved; in Section 3, we report on its current status and our main achievements and finally, in section 4 we provide some hints on the relevance of our contributions.

2 Project Goals

This project has as a main goal the development of conceptual modelling methods, techniques and tools to develop Quality Web Applications and E-Services in a systematic and correct way. To achieve this purpose we define the following goals:

1. Introduce a requirements model to support the early specification of Web application requirements.
2. Introduce conceptual mechanisms to support Personalization of Web Applications.
3. Define techniques to model complex web services.
4. Define an architectural model for service oriented architectures (SOA) and develop software patterns to implement this model.
5. Define conceptual model transformation processes to automate the software application and web service generation.
6. Define mechanisms for the integration and interoperability of applications and web services at the semantic level.
7. Development of Metrics and Evaluation of quality of web applications and services.
8. Define a Modelling Language for the Specification of Smart Home Services.

As an underlying goal of the research project, we will develop tools for supporting our ideas and we will apply the research results to real environments.

Finally, the approximate schedule of the project is the following:

Tasks	First year												Second year												Third year											
T 1.1	x	x	x																																	
T 1.2				x	x	x	x	x	x	x	x	x																								
T 1.3													x	x	x	x	x	x																		
T 1.4																			x	x	x	x	x	x	x	x	x									
T 2.1	x	x	x																																	
T 2.2				x	x	x	x	x	x	x	x	x																								
T 2.3													x	x	x	x	x	x																		
T 2.4																			x	x	x	x	x	x												
T 3.1	x	x	x	x	x	x	x	x	x																											
T 3.2										x	x	x	x	x	x																					
T 3.3																x	x	x	x	x	x															
T 3.4																						x	x	x	x	x	x	x	x	x						
T 4.1	x	x	x	x	x	x																														
T 4.2							x	x	x	x	x	x	x	x	x																					
T 4.3																x	x	x	x	x	x	x	x	x	x	x	x									
T 5.1	x	x	x																																	
T 5.2				x	x	x																														
T 5.3				x	x	x	x	x	x																											
T 5.4				x	x	x	x	x	x	x	x	x	x	x	x																					
T 5.5																x	x	x	x	x	x	x	x	x	x	x	x									

T 5.6			x|x|x|x|x|x|x|x|x|x|x
T 5.7			|x|x|x|x|x
T 6.1	x|x|x|		
T 6.2	|x|x|x|x|x|x|x|x|x		
T 6.3		x|x|x|x|x|x|x|x|	
T 6.4		|x|x|x|x|x|x	x|x|x	
T 6.5			|x|x|x|x|x|x|x|x
T 7.1	x|x|x|		
T 7.2	|x|x|x|x|x|x|		
T 7.3	|x|x|x	x|x|x|x|x|	
T 7.4	|x|x|x	x|x|x|x|x|	
T 7.5		|x|x|x|x|x	
T 7.6		|x|x|x|x|x	
T 7.7			|x|x|x|x|x|x|x|x
T 8.1	x|x|x|x|x|x|		
T 8.2	|x|x|x|x|x	x|x|x|	
T 8.3		|x|x|x|	
T 8.4		|x|x|x|x|x	x|x|x|	
T 8.5		|x|x|x|x|x	x|x|x|x|x|	
T 8.6			|x|x|x|x|x|x|x|x

Each task corresponds to each of the eight main goals mentioned previously. In this table, bold crosses denote tasks that have already been completed.

3 Level of Success

The level of success achieved along the project has been important and very high. We have solved nearly all the goals.

We have not found any remarkable incidence during the project execution. The planned tasks are being developed following the predicted plan and most of the main goals of the project are maintained and achieved. We only want to comment that the goal 3 (*Define techniques to model complex web services*) has been modified to adapt the goals to the current research interests and orientation of the group. As we are developing and extending the OOWS method, we have focused our research in how to generate navigational models from complex web service specifications. In this way we can generate/implement web interfaces that give support to business processes that have been defined as a complex web service.

The DESTINO project was awarded with budget to hire one or two technicians. The budget did not allow us to hire two people during the whole project. In this way, during the first year we hired María Victoria Torres and Javier Muñoz. Javier got a FPU grant and we hired José Ignacio Panach until March in year 2006. From this date, María Victoria was the only technician that was hired until now. Due to the delay that has passed between the proposal preparation and the development

of the project, the group has grown with the incorporation of new grant holders and new engineers that are currently working in the research group. These new members were hired by other funding mechanisms different to those provided by the DESTINO project such as technology transfer projects with software companies; In addition, the group has also got invited lecturers that are doing their PhD in our research group. This new information has been processed following the instructions: "Instrucciones para Proyectos Aprobados bajo la convocatoria 2004 del Programa Nacional del Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica".

In the following subsections, we analyze briefly how we have solved each goal, the main problems that we have detected, and the relevant results generated in the project.

3.1 Introduce a requirements model to support the early specification of Web application requirements.

First, we have studied the discipline of Requirements Engineering (RE) as well as the discipline of Web Engineering (WE). The goal of this study has been to know how navigational requirements are supported by the different approaches that belong to each discipline. Two conclusions have been obtained from this study: (1) RE approaches are mainly focused on traditional software (non-Web) and they do not properly consider unique characteristics of the Web application development; (2) WE approaches are focused on capture the navigational aspects at the design level. The specification of navigational requirements are handled with less relevance or not considered at all.

Next, we have defined a requirements model for Web applications in order to solve the detected lack of techniques for the specification of navigational requirements. To do this, we have analyzed the different aspects of a Web application that need to be considered to capture navigation at the requirement model. We have also defined a graphical notation which facilitates the definition of these aspects. Furthermore, a model-to-model transformation has been defined in order to allow us to automatically obtain the navigational model proposed by the OOWS method from the new requirements model for Web applications. Currently, we are developing the last task, which is the development of tools for supporting the creation of this requirements model. To do this task, we are using two development environments which support the Model Driven Development: Eclipse and DSL Tools.

The main results of our work are the following: (1) We have defined a study of the current approaches that belong to both disciplines (RE and WE). We have identified their strengths as well as their main weakness. (2) We have proposed a requirements model based on the concept of tasks. This model is made up of: (a) A task taxonomy which hierarchically describe the different tasks that users need to perform by interacting with the Web application. This taxonomy includes temporal relationships among the different task based on those proposed by Paternò. (b) A set of activity diagrams which complement the task taxonomy. These diagrams extend the UML activity diagram in order to allow us to describe each task from the interaction that the user requires form the system in order to perform it. (c) A set of information templates that allow us to describe it in detail the information that is exchanged between the user and the system when they interact to each other. (3) We have identified a set of mappings between the task-based requirements model and the OOWS navigational model in order to define a model-to-model transformation between both models. Finally, this work has produced publications in international conferences such as ICWE 2005 [1] , ER 2005 [2], ECWeb 2005 [3], WER 2005 [4] and CAiSE 2006 [5] as well as in the International Journal on Web Engineering and Technology IJWET 2007 [6].

3.2 Introduction of conceptual mechanisms to support Personalization of Web Applications.

We have studied and analyzed the adaptive methods considered in the implementation of Adaptive Hypermedia Systems, with focus on the taxonomy of adaptive methods and techniques made by the Adaptive Hypermedia community and the analysis of some existing Adaptive Web Applications. We have also analyzed the approaches to Adaptivity proposed by the most important Model-Driven Web development approaches, such as OO-HDM, WebML, WSDM, OO-H, Hera, capturing valuable information about how Adaptivity is being tackled from a model-oriented perspective.
New conceptual primitives have been introduced into OOWS Navigational and Presentational Models, to properly define the characteristics of different types of users, as well as the user-centred constraints defined on structural and operational views of the application, which allow modelling multiple adaptive methods. An implementation framework has been built, in order to obtain implementations of the adaptive characteristics defined at conceptual level, following a model-driven strategy. As a final task, a CASE tool is currently under development to support the proposal.
As noticeable results, we made a detailed study of the current approaches of Adaptivity in Web Applications. A set of conceptual structures that support de modelling of adaptive methods has been incorporated to OOWS conceptual models. A User Modelling strategy, based on three views of the application users (personal, domain-dependent and interaction behaviour characteristics), has been proposed to serve as a basis of the modelling of adaptation features. The development of this work has produced multiple articles in international conferences specially dedicated to Adaptive Hypermedia and Model-Driven Web Development, as IWWOST 2005 [7], WISE 2005 [8] and AH 2006 [9].

3.3 Define techniques to model complex web services.

Service integration and composition is a vital issue to achieve the industrial adoption of the Web service technology. In this goal we initially studied the different modelling languages for the specification of complex processes and the dynamic interaction between web services. This study was focused on UML Sequence and Activity Diagrams and the BPMN notation. Finally, we opted by the BPMN notation to model complex processes. Then, we extended the OO-Method/OOWS approach with the intention of modelling these complex interactions. This extension was based on a Business Process Model (BPM) that was graphically depicted by using the BPMN notation. Moreover, we defined the necessary extensions to the method (both in the navigational and presentation models) to cope properly with the navigation and presentation of complex processes when they required human participation. Then, as the method follows a model driven approach, we have defined the necessary model-to-model and model-to-text transformations to generate automatically from a business process specification (1) the necessary graphical user interfaces that allow users to interact with the process and (2) the corresponding representation of the processes modelled in the new BPM in terms of an executable language (in particular to the WS-BPEL language). The fact that WS-BPEL relies on the Web Services technology complicates the utilization of this execution language in a context where tasks are not supported by services (for instance manual tasks). Therefore, in order to solve this limitation we generate the necessary infrastructure to interact with the process engine in charge of executing the process. Nowadays, we are implementing the extension to the tool that provides support to the OOWS method.

As a result of this goal we have produced several publications in national and international conferences such as BPM 2006 [10], JISBD 2006 [11], JSWEB 2005 [12] and IDEAS 2005 [13].

3.4. Define an architectural model for service oriented architectures (SOA) and develop software patterns to implement this model.

SOA and their conceptualization proposals, for example Web Service Architecture (WSA) from W3C or OASIS, have been studied. In this step we have developed a Service Model. This model is a conceptual framework for SOA where a web service definition is proposed with the characteristics and common elements that define a SOA and their relationships. An architectural style for SOA and a set of frameworks and design patterns are developed. They allow implementing this architecture giving support to some design principles like loose coupling, coarse-grained, no connexion oriented, stateless, etc.

As a result, a MOF metamodel for SOA has been obtained. This metamodel includes the definition and the relationships between SOA elements and a set of design patterns to implement the characteristics of this architecture. These ideas have been published in several works in national and international conferences like JSWEB 2005 [12], IDEAS 2005 [13], MDWE 2005 [14] and LA-WEB 2005 [15].

3.5. Defining conceptual model transformation processes to automate the software application and web service generation.

A strategy to apply MDA and the ideas of the Model Driven Development (MDD) to OOWS and SOA has been developed. This strategy automatically generates OOWS models and SOA architectures over different technologies. Two models have been designed: (1) the PIM metamodel to the proposed notation in the OOWS method, expressed by means of a MOF metamodel (eCore, making use of the EMF tools of the Eclipse environment), (2) the PIM metamodel to the SOA architectures, also expressed by means of a MOF metamodel. The starting point to define these PIMs has been the Service Model proposed in the previous objective. This metamodel defines in a generic way, the essential characteristics and components of SOA architectures. A set of transformation patterns has been proposed in order to define the mappings between the OOWS conceptual models, SOA architectures and the web service platforms to be implemented. In this case, technological metamodels are not defined (contrary to the ideas initially presented) to carry out the proposal. Here, the code is generated for a specific platform by means of model transformations.

As results, MOF metamodels for OOWS and SOA have been obtained. These metamodels are used by the transformations in order to generate code, therefore these metamodels appear in most of the works and international publications of the group like ICSOC 2005 [16], JSWEB 2005 [17], CAiSE Forum 2005 [18], IDEAS 2005 [13], MDWE 2005 [14], LA-WEB 2005 [15], WEWST 2006 [19], JSWEB 2006 [20], CoSS 2006 [21] and SPRINGER-BOOK 2006 [22].

3.6 Define mechanisms for the integration and interoperability of applications and web services at the semantic level.

The OOWS method has been extended in order to support the development of interoperable web applications over service oriented architectures (SOA). This extension has been performed in two steps. In a first step, we have introduced a service model that allow us to handle at the modelling level both the external functionality and the own functionality derived from the specification of our

system. The main goal of this step is (1) to make transparent to the modeller/designer the technological aspects introduced by external services and (2) to specify the functionality that is going to be provided as a web service to external systems. In a second step, we have introduced a new dimension to the method that allows us to describe the system (its data and functionality) from the point of view of the Semantic Web. The main objective of this new dimension is to generate content for the web that facilitates its search, invocation and composition in an automatic way. Nowadays, we are developing extensions to the OlivaNova CASE tool in order to generate semantic data and functionality.
As a result we have introduced three new models to the OO-Method/OOWS method. One of these models has been introduced to describe services (both external and own) at the modelling level. The other two models have been introduced to specify the system from the semantic web point of view. In these two models we specify what parts (data and functionality) from the system are going to be accessible as semantic web content. Then, these parts are transformed into a specific semantic web language. We have produced several publications in international conferences such as CoSS 2006 [23], MDWE 2005 [14], CAiSE Forum 2005 [24] and LA-WEB 2005 [15].

3.7 Development of Metrics and Evaluation of quality of web applications and services.

A comparative analysis was carried out of functional size metrics for Web applications. The main limitation of these metrics was their late application (software products already implemented), which implies an inefficient estimation in relation to resource management. We defined a set of functional size metrics by means of the adaptation of function points for the measurement of OOWS conceptual models. In order to do this, we identified some primitives from navigation and presentation models that are relevant to obtain their functional size and one or more metrics being defined for each of these primitives. In addition, with the aim of estimating opportunely, another set of metrics from requirements specifications was defined. These metrics were defined on the basis of the COSMIC-FFP method.
A theoretical validation of the metrics defined on the basis of function point method was validated only partially. The main reason was the use of different measurement scales for the quantification of the functional size of the OOWS conceptual schemas. Nevertheless, the metrics defined on the basis of COSMIC-FFP could indeed be validated successfully using the DISTANCE framework.
Various experimental studies were carried out to evaluate and validate metric quality. The results of these studies empirically demonstrated that the metrics defined for conceptual schemas and requirements specification were accurate and reproducible. Furthermore, these metrics were perceived to be easy to use and useful for the estimation of other indicators (effort and productivity).
In this last year the automation of the estimation of the functional size of Web applications from conceptual schemas is being carried out and will be embedded in the OlivaNova tool as an additional module. The automation of the estimation of functional size from requirements specification is also being implemented.
Members of our group are active members of the Spanish Metrics Society (AEMES). A number of articles have been published in various journals, conferences and workshops, such as Journal of Information and Software Technology [25], the Journal on Systems and Modelling (SOSYM) [26], JCST [27], QSIC 2006 [28], MENSURA 2006 [29], QoIS 2006 [30], ACOSM 2005 [31], WER 2006 [32], LA-WEB 2006 [33], IDEAS 2006 [34], INTERACCION 2005 [35] and IDEAS 2005 [36].

3.8 Definition of a Modelling Language for the Specification for Smart Home Services.

Current home automation systems and future trends have been analyzed, and the existing technologies and standards have been studied. A UML-based modelling language, which we have called PervML, has been defined with the goal of providing the capabilities for modelling the functional requirements of home automation systems and other kinds of pervasive systems. A systematic way for the development of the conceptual model of the home automation has been proposed. A software framework for the development of residential gateways which is based on the Java-OSGi technology has been defined including the transformation rules from the specification language to the framework. Currently, a CASE tool supporting the language is under development.

As a result: (1) a metamodel of PervML including their OCL constraints has been obtained. With this metamodel, a graphical syntax has been specified using as notation several UML 2.0 diagrams, (2) a set of steps (system modelling, detailed specification, generation and configuration of software components, etc.) and the software development process assets that must be produced and consumed as input and output of each step has been defined, (3) an architectural style and a set of transformation rules which define the mappings from the conceptual primitives and their software representations are provided. This work has been published in several national and international conferences like CAiSE 2005 [37], JISBD 2005 [38], UCAmI 2005 [39], DSDM 2005 [40], RE 2006 [41], CICU 2006 [42], IWUC 2006 [43], IDEAS 2006 [44], ICEIS 2006 [45], OT4AmI 2006 [46] and SEPS 2006 [47], international journals like ERCIM News 2006 [48] and a book chapter published at PERVASIVE INFORMATION SYSTEMS [49].

4 Result Indicators

This section summarises the main results of the DESTINO project including a description of the PhD theses under development during the project, the main publications, collaborations with national and international groups, technology transfer and finally participation with related projects and networks.

4.1 PhD Theses and Human Resources

Our research group has a broad experience in training PhD students, who have joined the work developed in the group in a very effective way. Moreover, the R&D projects in which we are involved together with companies of the Software Development industry (as concrete example we can quote CARE Technologies S.A., INTEGRANOVA S.A. and Consoft S.A.) which allows us having PhD students who have reached an advanced state in their Thesis, working in R&D projects co-funded by the participant companies. Our PhD students have the possibility of focusing their research work towards either the academy or more applied; implementing the ideas resulted from their research works in industrial contexts. At the same time, this activity generates new work in a continuous way, what increases the personnel necessities in order to continue with the advancement in all the opened research lines.

The following 3 PhD Theses have being developed in the context of this project:

- Manoli Albert (advised by Vicente Pelechano and Oscar Pastor)
- Maria Angeles Pastor (advised by Matilde Celma)
- Silvia Abrahao (advised by Oscar Pastor)

Currently, 10 theses are being advised, and four or them are planned to be finished during the year 2007, which will increase the formative capacity of the group. These theses are the ones from:

- Javier Muñoz, Pedro J. Valderas, Victoria Torres and Marta Ruiz (advised by Vicente Pelechano)
- Joan Fons and Gonzalo Rojas (advised by Oscar Pastor and Vicente Pelechano)
- Nelly Condori Fernández and Isabel Díaz (advised by Oscar Pastor and Silvia Abrahao)
- Alicia Martínez and Hugo Estrada (advised by Oscar Pastor and John Mylopoulos).

Several PhD and master students are starting their research career and they probably will be the next PhD theses that are going to be developed in the context of our research projects. These are the following students:

- Carlos Cetina, Estefanía Serral, Pau Giner, Sergio España, Francisco Valverde, José Ignacio Panach, Inés Pederiva, David Anes.

4.2 Publications

Our research group has a solid tradition on publication in scientific journals and international conferences of high impact and prestige. Concretely, our group attends and publishes in well know Software Engineering and Data Bases conferences like CAiSE, ER, ICSE, DEXA, MODELS, RE, BPM, BP-UML, IDEAS, JISBD, and the DSDM Spanish workshop. In the Web Engineering field we actively participate in conferences like WWW, ICWE, EC-WEB, LA-WEB, AH, WISE, ICSOC, ECOWS, CoSS, JSWEB, etc. In the Human-Computer Interaction area, we publish in conferences like CADUI, INTERACCIÓN and HWID. In the Empirical Software Engineering and Quality areas we publish in conferences like METRICS, ISESE, ISERN and IFPUG. Finally, in the Smart Homes and Pervasive Systems areas we have published in conferences like IWUC, OT4AmI, SEPS, MOMPES, UCAmI, CICU and ICEIS. These forums constitute the first level of basic spreading for our research group. We believe, as a group policy, that our works should be publicly presented and defended in front of international experts.

We also want to note the participation of several member of our group in the Program Committee of relevant international conferences like CAiSE, ER, WWW, RE, ICEIS, DSV/IS, ICWE, EC-WEB. Professor Oscar Pastor is member of the Steering Committee of ICWE, he was the PC Chair in the year 2002 edition, he has been PC Chair of the ER and CAiSE International Conferences in year 2005, and he is the Deputy Chair of the Web Engineering track in the WWW 2007 conference. The Main Researcher (MR) of this project is the PC Chair of IDEAS 2007 Ibero-American Conference, member of the executive committee and workshop chair of the JISBD 2007 (the Spanish conference on Software Engineering and Data Bases), PC Chair and Organizer of the DSDM workshop (on model driven development) and PC Member of the WWW, ICWE and EC-WEB international conferences among others.

Our research group has published in international and national scientific journals with a noticeable impact in the areas that we are currently working. We are publishing in Information Systems Journal, Data and Knowledge Engineering, IEEE Multimedia, Requirements Engineering Journal, Information and Software Technology, SoSyM, International Journal on Web Engineering (JWE), International Journal of Web Engineering and Technology (IJWET), etc.

Summarizing, the main publications derived from this project during these 2 years are:

- 6 Books.
- 7 Chapters in Books
- 8 Articles in International Journals
- 42 papers in International Conferences

- 13 papers in National Conferences

4.3 Collaboration with other International and National Groups

We have reliable contacts with significant researchers with international prestige in the fields of Software Engineering, Requirements Engineering, Web Engineering and Model Driven Development. We keep in touch with them and send several PhD students to stay with them to allow students to acquire a complementary training that we consider really important for their career development. For instance, our PhD student Nelly Condory Fernández has been working with Prof. Alain Abran from the University of Quebec, Montreal (Canada); Gonzalo Rojas is currently working with Geert Jan Houben from the Vrije University of Brussels, Belgium; Inés Pederiva is working actively in a series of different visits with Prof. Jean Vanderdonckt, from the University of Louvaine-La Neuve, Belgium. Our close relationship with the community of the Web Engineering and the participation of the ALFA-WEE-NET project allows us to keep interchange relationships and very fruitful collaborations with the groups led by Prof. Dr. Paolo Paolini, Franca Garzotto and Luciano Baresi from the Politecnico de Milán, with Prof. Nora Koch, from the Ludwig-Maximilians-University o Munich, Germany; with Martin Gaedke, from the University of Karlsruhe, Germani, with Gerti Kappel, from the University of Linz, Austria, with Olga De Troyer, from the Vrije University of Brussels, Belgium, with Paolo Atzeni, from the University Roma Tre, Roma, Italy, with Stephan Reiff-Marganiec, from the University of Leicester, United Kingdom, with Wieland Schwinger from the Johannes Kepler University of Linz, Austria, with Prof. Symeon Retalis from the University of Piraeus, Greece, with Prof. Dr. Daniel Schwabe, from the PUC of Rio de Janeiro (Brazil), with Dr. Luis Olsina from the University Nacional de la Pampa (Argentina) and with Prof. Dr. Gustavo Rossi, from the University of La Plata (Argentina). In particular we have a very close relationship with Prof. Dr. Gustavo Rossi, with whom the MR of this project co-advises the thesis of the PhD student Andrés Fortier in the scope of the Pervasive Systems. Moreover, we have also applied together for an AECI cooperation project, in the frame of the Latin-American Cooperation Programme, in order to collaborate in aspects related to Web Engineering and Pervasive Systems.

Finally, it is worth to mention the relationship based on technological transfer projects of the company CARE Technologies with Prof. Dr. Heinrich Maier, from the University of Klagenfurt, Austria, the group led by Prof. Dr. Steve Liddle from the BYU, Salt Lake City, Utah, USA, with Prof. Brian Henderson-Sellers, from the University of Technology in Sydney, Australia and with Prof. Dr. John Mylopoulos, from the University of Toronto, Canada, with whom Prof. Oscar Pastor is co-managing two thesis (the ones from the PhD students Alicia Martínez and Hugo Estrada).

At a national level we have been in contact with several research groups which work, direct or indirectly, either in the project matter or in similar ones. Among them we should quote the group from the University of Seville (with Prof. Antonio Ruiz, Miguel Toro, Rafael Corchuelo and Amador Durán among others) which work in research lines such as Product Lines, Web Services and Quality, Requirements Specification and Software Factories; the groups from the University of Murcia (lead by Ambrosio Toval and Jesús García Molina) who are working in conceptual modelling using UML (properties formalization and verification), and in model driven development and model transformation respectively; the group from the University of Valladolid (leaded by Miguel Angel Laguna), which is working in the integration of features models with goal-driven requirement models; the group from the University of Alicante (whose person in charge is Jaime Gómez) which works in Web application conceptual modelling; the group from the University of Malaga, to which belongs José María Troya, Ernesto Pimentel and Antonio Vallecillo,

which work, among others, in fields such as Web services, components, aspect orientation, MDA and metrics; the group lead by Dr. Mario Piattini from the University of Castilla-La Mancha whose members work in areas related to metrics, modelling, information system security and support following a model driven approach. The KYBELE group, headed by Prof. Esperanza Marcos from the University Rey Juan Carlos, whose members are working in the definition of methods for the development of web applications within the scope of the MDSD. We also keep a productive collaboration with the group leaded by Prof. Natalia Juristo from the Technical University of Madrid within the scope of Usability and the Empirical Software Engineering. Moreover, we have contacts with members from the group leaded by Prof. Jesús Lorés from the University of Lerida in the scope of the Human-Computer Interaction; at a national level we are involved with the academic section of the Human-Computer Interaction Association (AIPO). We also keep in touch with the groups headed by Prof. Antoni Olivé and Ernest Teniente, and Pere Botella and Xavier Franch, in the scope of conceptual modelling and non-functional requirements respectively, from the Technical University of Catalunya. Within the scope of the development of home automation and middleware, we have contact through projects (such as ATENEA or OSIRIS) with the group leaded by Prof. Juan Carlos Dueñas from the UPM. These relationships are basically focused in the monitoring of phD Thesis, in the support to different workshops held periodically or in the scope of different events such as the Jornadas de Ingeniería del Sofware y Bases de Datos, the National Network on MDSD (Red Nacional sobre DSDM) and the MDSD workshops (Taller de DSDM).

4.4 Technology Transfer

In this project the companies CARE Technologies S.A. and Consoft S.A participate as interested agents. Our research group has a very good relationship with these companies and these are supporting several research lines of our group. Currently the group is working on four active technology transfer projects for these companies, where the common goal is to incorporate the R&D results from the group into their industrial tools (OlivaNova, the programming machine). Other support mechanisms given by the group to these companies are: the investment on R&D agreements in order to support and develop specific aspects of the implementation, participation as partners in project calls such us PROFIT, support on project preparations for other calls as Torres Quevedo, FP7, etc.

Other remarkable activity of the OO-Method group, whose main goal is to improve the technology transfer from the group to the industry, is its active participation in related Spanish Technology Platforms. Our group is an active member of the following Spanish Technology Platforms: INES (*Iniciativa española de software y servicios)*, PROMETEO (*Plataforma Tecnológica Española de Sistemas con Inteligencia Integrada*), and eMOV (*Plataforma de Comunicaciones Inalámbricas*). All the activities of the group in these forums, where the main companies working on the similar research fields can be met, trend to improve the knowledge of the group's research activities and capacities among the industry. As a result of the knowledge of the group's capacities and the strengthening of the relations between the group and the specialized industry, new partnerships have raised and the experience of the group is being actively transferred to the companies in the frame of collaborative projects as ATENEA (Architecture, Middleware and Tools - Ref. FIT-340503-2006-5), which is a result of the work performed in the PROMETEO Technology Platform. The aim of the group is to progress in the technology transfer to the Spanish industry by improving its participation in national forums, as the Technology Platforms, and by developing stronger relations with all the Spanish agents that can take advantage of our R&D results.

4.5 Participation in Related Projects and Networks

As a result of our group interests and the reach of our work in this scope its worth mentioning the participation of our group in important projects and networks such as:

- COST MAUSE (Towards the MAturation of Information Technology USability Evaluation, http://www.cost294.org/). European Union COST Action. Lead by Effie Lai-Chong Law, ETH, Zurich.
- WEE-NET:Web Engineering Network of Excellence, Web: http://cosy.ted.unipi.gr/weenet/. Alfa European Union Project. Lead by Symeon Retalis (University of Pyreaus, Greece)
- The ITEA-OSIRIS Project (http://www.itea-osiris.org/), in which we participate as guests with several companies at national and international level such as Telvent, Telefónica, Ericsson, Philips, Sintef, etc.
- Model Driven Software Development National Network of Excellence (Red de Excelencia Nacional en el ámbito del Desarrollo de Software Dirigido por Modelos) (lead by Prof. Antonio Vallecillo).
- The ATENEA Project from the MITYC (FIT-340503-2006) participating with AICIA, CARE-Technologies, DS2, Eliop, Etra, Fidetia, Fundación European Software Institute, Mondragón Corporación Cooperativa, RedIRIS, Telefónica I+D, TELVENT, Universidad Carlos III, Universidad de Cantabria, Universidad de Málaga, Universidad Politécnica de Valencia, Universidad de Vigo, Universidad Politécnica de Madrid, Vodafone.

5 References

[1] P. Valderas, J. Fons, V. Pelechano. From Web Requirements to Navigational Design. A Transformational Approach. Springer-Verlag, Lecture Notes in Computer Science 3579. Proc. Of 5th International Conference on Web Engineering, ICWE 2005. 25-29 July, Sidney, Australia. pp 506-511.

[2] P. Valderas, J. Fons, V. Pelechano. Transforming Web Requirements into Navigational Models: An MDA based Approach. Springer-Verlag, Lecture Notes in Computer Science 3716. Proc. Of the 24th International Conference on Conceptual Modelling (ER 2005). Klagenfurt, (Austria). pp 63-75. 2005

[3] P. Valderas, J. Fons, V. Pelechano. Developing E-Commerce Applications from Task-Based Descriptions. Springer-Verlag, Lecture Notes in Computer Science 3590. Proc. Of the 6th Electronic Commerce and Web Technologies (ECWEB 05). 23-26 August, Copenhagen, Denmark, 2005.

[4] P. Valderas, J. Fons, V. Pelechano. Using Task Descriptions for the Specification of Web Application Requirements. WER 2005. VIII International Workshop on Requirements Engineering. 13-14 June 2005, Porto, Portugal. pp 257-268.

[5] P. Valderas, V. Pelechano, O. Pastor. Towards an End-User Development Approach for Web Engineering Methods. 18th Conference on Advanced Information Systems Engineering (CAiSE 2006). Grand-Duchy of Luxembourg,June 2006. SPRINGER, LNCS 4001, pp 528-543.

[6] P. Valderas, V. Pelechano, O. Pastor. A Transformational Approach to Produce Web Application Prototypes from a Web Requirements Model. International Journal of Web Engineering and Technology 2007 - Vol. 3, No.1 pp. 4 – 42.

[7] G. Rojas, V. Pelechano, J. Fons. A Model-Driven Approach to include Adaptive Navigational Techniques in Web Applications. V International Workshop on Web Oriented Software Technologies - IWWOST'05 (within The 17th Conference on Advanced Information Systems Engineering (CAiSE'05)), Porto, Portugal.

[8] G. Rojas, V. Pelechano. A Methodological Approach for Incorporating Adaptive Navigation Techniques into Web Applications. VI International Conference on Web Information Systems Engineering - WISE 2005. New York, NY, USA, November 20-22, LNCS 3806, Springer.

[9] Rojas, G., Valderas, P., and Pelechano, V.(2006). Describing Adaptive Navigation Requirements Of Web Applications. IV International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems - AH2006. Dublin, Ireland, June 20-23, 2006, Proceedings. Lecture Notes in Computer Science 4018, Springer.
[10] V. Torres, V. Pelechano. Building Business Process Driven Web Applications. BPM 2006 - Fourth International Conference on Business Process Management, Vienna, Austria, 5.-7. September 2006. EDITORES S. Dustdar, J.L. Fiadeiro, and A. Sheth. LNCS 4102. SPRINGER. Pag: 322 – 337
[11]V. Torres, V. Pelechano, P. Giner. Generación de Aplicaciones Web basadas en Procesos de Negocio mediante Transformación de Modelos. XI Jornadas de Ingeniería de Software y Base de Datos (JISBD 2006) October, Barcelona (Sitges) (Spain), 2006
[12] R. Quintero, V. Torres, M. Ruiz, V. Pelechano. Modelado de Aplicaciones Web Colaborativas Basadas en Servicios y Dirigidas por Procesos de Negocio B2B. I Jornadas Científico-Técnicas en Servicios Web (JSWEB 2005) September, Granada. España.
[13] R. Quintero, V. Pelechano, V. Torres, M. Ruiz. Una Solución a la Integración de los Métodos de Ingeniería Web y las Aplicaciones B2B. Un caso de Estudio. VIII Workshop Iberoamericano de Ingeniería de Requisitos y Ambientes Software (IDEAS 2005) Mayo, Valparaíso (Chile)
[14] V. Torres, V. Pelechano, M. Ruiz, P. Valderas. A Model Driven Approach for the Integration of External Functionality in Web Applications. The Travel Agency System. I Workshop on Model-driven Web Engineering (MDWE 2005) - July, Sydney (Australia), 2005.
[15] Victoria Torres, Javier Muñoz, Vicente Pelechano. A Model-Driven Method for the Integration of Web Applications. 3rd Latin American Web Congress (LA-WEB 2005) October, Buenos Aires (Argentina), 2005
[16] M. Ruiz, P. Valderas, V. Pelechano. Applying a Web Engineering Method to Design Web Services. The 3th International Conference on Service Oriented Computing (ICSOC'05), Amsterdam (Holanda), December 12-15, 2005. SPRINGER. LNCS 3826 pp 576-581.
[17] M. Ruiz, P. Valderas, V. Torres, V. Pelechano. Una Aproximación Dirigida por Modelos para el Diseño de Servicios Web en el ámbito de un Método de Ingeniería Web. I Jornadas Científico-Técnicas en Servicios Web, JSWEB 2005. Granada. Páginas 133 – 140.
[18] M. Ruiz, V. Pelechano. Model Driven Design of Web Service Operations using Web Engineering Practices. Workshop on Emerging Web Services Technology, in conjunction with 4th IEEE European Conference on Web Services. Pags: 75 – 90
[19] M. Ruiz, P. Valderas, V. Torres, V. Pelechano. A Model Driven Approach to Design Web Services in a Web Engineering Method. CAiSE Forum 2005. Pags: 183-188.
[20] M. Ruiz, F. Valverde, V. Pelechano. Desarrollo de servicios Web en un método de generación de código dirigido por modelos. II Jornadas Científico-Técnicas en Servicios Web, JSWEB 2006. Santiago de Compostela. España. Pags: 14-22.
[21] M. Ruiz, V. Pelechano, O. Pastor. Designing Web Services for Supporting User Tasks: A Model Driven Approach. I Conceptual Modeling of Service-Oriented Software Systems (CoSS 2006) November, Tucson, Arizona (USA) 2006 in conjunction with the 25th International Conference on Conceptual Modeling. LNCS 4231. Pags: 193 – 202.
[22] O. Pastor, J. Fons, S. Abrahao, V. Pelechano. Conceptual Modelling of Web Applications: the OOWS approach. Web Engineering. Eds. Emilia Mendes and Nile Mosley, Springer 2006, pp. 277-302
[23] V. Torres, V. Pelechano, O. Pastor. Building Semantic Web Services Based on a Model Driven Web Engineering Method. I Conceptual Modeling of Service-Oriented Software Systems (CoSS 2006) November, Tucson, Arizona (USA) 2006 in conjunction with the 25th International Conference on Conceptual Modeling. LNCS 4231. Págs. 173-182
[24] V. Torres, R. Quintero, M. Ruiz, V. Pelechano. Towards the Integration of Data and Functionality in Web Applications. A Model Driven Approach. 17th Conference on Advanced Information Systems Engineering Forum (CAiSE Forum 2005) - June, Porto (Portugal), 2005
[25] S. Abrahão, G. Poels. Experimental Evaluation of an Object-Oriented Function Point Measurement Procedure. Information and Software Technology (IST), 2006. Elsevier.

[26] S. Abrahão, G. Poels, O. Pastor. A Functional Size Measurement Method for Object-Oriented Conceptual Schemas: Design and Evaluation Issues, Journal on Software & System Modeling (SoSyM), Vol. 5, No 1, April 2006, 48-7 pp., Springer-Verlag.
[27] N. Condori-Fernández, S. Abrahão, O. Pastor. On the Estimation of Software Functional Size from Requirements Specifications. Journal of Computer Science and Technology (JCST), Special Issue on Advances in Software Metrics and Software Processes. Springer (accepted for publication).
[28] N. Condori-Fernández, O. Pastor. An Empirical Study on the Likelihood of Adoption in Practice of a Size Measurement Procedure for Requirements Specification 6th IEEE International Conference on Quality Software (QSIC 2006), Beijing, China, October 2006.
[29] N. Condori-Fernández, O.Pastor. Re-Assessing the Intention to Use a Measurement Procedure based on COSMIC-FFP, International Conference on Software Process and Product Measurement, MENSURA 2006, Cadiz, Spain, November 2006.
[30] N. Condori-Fernández, O.Pastor. Evaluating the Productivity and Reproducibility of a Measurement Procedure, International Workshop on Quality of Information Systems, Springer LNCS, (QoIS), Tucson-Arizona, USA, November 2006.
[31] S. Abrahão Early Life-Cycle Estimation for Web Projects. The Australian Conference on Software Measurement (ACOSM 2005), co-located with the 2005 Empirical Software Engineering International Week, Organized by ASMA (Australian Software Metrics Conference) & SQA (Software Quality Association), Noosa Heads, Queensland, Australia, 17-18 November 2005.
[32] N. Condori-Fernández, O. Pastor. Verifying the Construction of a Software Model from a Requirements Model, IX Workshop on Requirements Engineering (WER 2006), Río de Janeiro, Brazil, July 2006.
[33] Abrahão S., Poels G. Further Analysis on the Evaluation of a Size Measure for Web Applications. 4th Latin American Web Congress (LA-WEB 2006), Puebla Cholula, Mexico, October 25-27 2006, IEEE Press, pp. 230-240.
[34] N. Condori-Fernández, S. Abrahão, O. Pastor. El Problema de la Duplicidad de Movimientos de Datos en un Procedimiento de Medición. 9º Ibero-american Workshop on Requirements Engineering and Software Environments (IDEAS'2006), La Plata, Argentina, April 2006.
[35] N. Condori-Fernández, J. C. Ortiz, S. Abrahão, Un Marco de Usabilidad para las Aplicaciones Generadas con la Tecnología Oliva Nova Model Execution, VI Congreso Español de Interacción Persona-Ordenador (INTERACCIÓN 2005), Granada, Septiembre 2005.
[36] N. Condori-Fernández, S. Abrahão, O. Pastor. Validación del Diseño de un Procedimiento de Medición de Tamaño Funcional, 8º Ibero-american Workshop on Requirements Engineering and Software Environments (IDEAS'2005), Valparaíso-Chile, May 2005.
[37] J. Muñoz, V. Pelechano. Building a Software Factory for Pervasive Systems Development. The 17th Conference on Advanced Information Systems Engineering (CAiSE'05), Porto (Lisbon). June 2005. SPRINGER. LNCS 3520. Pags: 342 – 356.
[38] J. Muñoz, V. Pelechano, E. Serral. Providing platforms for developing pervasive systems with MDA. An OSGi metamodel. X Jornadas de Ingeniería de Software y Base de Datos (JISBD), Granada (Spain). Septiembre 2005. pags: 19 - 26
[39] J. Muñoz, I. Ruiz, V. Pelechano, C. Cetina. Un framework para la simulación de sistemas pervasivos. Simposio sobre Computación Ubicua e Inteligencia Ambiental (UCAmI'05), Granada (Spain), September 2005. pags: 181 - 190
[40] J. Muñoz, V. Pelechano. MDA vs Factorías Software. II Taller sobre Desarrollo de Software Dirigido por Modelos, MDA y Aplicaciones (DSDM), Granada (Spain). Septiembre 2005. CEUR Workshop Proceedings (vol-157)
[41] J. Muñoz, P. Valderas, V. Pelechano, O. Pastor. Requirements Engineering for Pervasive Systems. A Transformational Approach. 14th IEEE International Requirements Engineering Conference (RE 2006), Minneapolis (USA) 11-15 September 2006.
[42] J. Muñoz, V. Pelechano, E. Serral. Aplicación del Desarrollo Dirigido por Modelos a los Sistemas Pervasivos: Un Caso de Estudio. II Congreso IberoAmericano sobre Computación Ubicua (CICU 2006), Alcalá de Henares (Spain). 7-9 June 2006. pags: 171-178.

[43] J. Muñoz, V. Pelechano, C. Cetina. Implementing a Pervasive Meetings Room: A Model Driven Approach. International Workshop on Ubiquitous Computing (IWUC 2006), Paphos, Cyprus. 23 May 2006 pags. 13 – 20.

[44] J. Muñoz, C. Cetina, E. Serral, V. Pelechano. Un Framework basado en OSGi para el Desarrollo de Sistemas Pervasivos. 9º Workshop Iberoamericano de Ingenieria de Requisitos y Ambientes Software (IDEAS2006), La Plata (Argentina). 24 - 28, Apr 2006. pags: 257 – 270

[45] J. Muñoz, V. Pelechano. Applying Software Factories to Pervasive Systems: A Platform Specific Framework. 8th International Conference on Enterprise Information Systems (ICEIS 2006), Paphos (Cyprus) 23 - 27, May 2006. pags: 337 – 342.

[46] A. Fortier, J. Muñoz, V. Pelechano, G. Rossi, S. Gordillo. Towards an Integration Platform for AmI: A Case Study. Workshop on Object Technology for Ambient Intelligence and Pervasive Computing (OT4AmI 2006), Nantes (France). 4 July 2006.

[47] J. Muñoz, V. Pelechano, C. Cetina. Software Engineering for Pervasive Systems. Applying Models, Frameworks and Transformations. I International Workshop on Software Engineering for Pervasive Services (SEPS 2006), Lyon (France). 29 June 2006.

[48] J. Muñoz, E. Serral, C. Cetina, V. Pelechano. Applying a Model-Driven Method to the Development of a Pervasive Meeting Room. ERCIM News, April 2006 vol. 65, pags: 44-45.

[49] J. Muñoz, V. Pelechano. A Software Factory for Pervasive Systems Development. PERVASIVE INFORMATION SYSTEMS. 2006 (pendiente de publicación) Edited by Panos E. Kourouthanassis and George M. Giagli. Editorial: M.E. Sharpe.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

C4ISR Multimedia System for Emergency Management TIN2004-03588

Manuel Esteve*
Catedrático de Universidad
Departamento de Comunicaciones
Universidad Politécnica de Valencia

Abstract

C4ISR systems (Command, Control, Communications, Computers, Intelligence, Surveillance and Reconnaissance) are one of the most classical examples of critical real-time distributed systems. The origin of these systems falls under the scope of military research. Usually they are embedded systems with particular purpose and ad-hoc designed. The project proposes the migration of the technological and operative concepts presented in the DoD and NATO standards and developed in the C4ISR military systems to the civil environment. In special, emergency management since it has a large social impact. (flood, fire, terrorist attacks, dumping of toxic waste and so on). Additionally, the project proposes the massive inclusion of multimedia flows as a main technological innovation with regard to the military systems. The main project objectives are the following: (i) Generic C4ISR system architecture definition for emergency management. The defined architecture follows the DoD and NATO standards in the operational, system and technological plane and are based on the COTS (Commercial- of-the-self) development principle. (ii) Validation of the proposed concept through the development and assessment of a test application for fire management based on RT-LINUX as a particular case of a generic emergency management system. In order to guarantee the correct working of the distributed real time system and due importance of the communication (protocols, network technologies) another one of the main project objectives will be the analysis and detailed definition of the project communication architecture based on IP and wireless mobile networks. Moreover, from hard and soft requirement integration point of view it is important to emphasize the novel aspect and the added difficulty due to the massive inclusion of multimedia flows.
Keywords: Command and control information systems, emergency management, wireless networks, high quality video streaming, GPS, biometric sensing

1 Project goals

The main project goals fall under the National Plan of Informatics Technologies priority line "Real-time distributed systems. Modelling, development, verification and validation of critical systems. Architectures, platforms and interoperability. Fault tolerance" of the section "Open and Distributed Systems".

* Email: mesteve@dcom.upv.es

The project goals were proposed in order to fulfil the needs of the following scenario:

"A message from 112 emergency number, alerts the firemen officer about the detection of a forest fire with several fronts. The message contains the approximate fire position. Fast, a reconnaissance helicopter equipped with a video camera and an infrared (IR) camera takes off and goes directly to the fire area. In parallel, three terrestrial units drive to the same zone. The helicopter places above the fire area and sends to the commanding officer (who is at the firemen building in the general headquarters) information in digital format, automatically, continuously and in real-time. This information is: (i) GPS reference position; (ii) video images from the area; (iii) IR images from the area; (iv) wind speed and direction. This information is pre-processed and fused at the control computer and help the officer to know the exact size and location of the disaster. Automatically the control computer at the headquarters: (i) extracts from a GIS the map of the disaster area; (ii) isolates and represents the fire fronts onto the map; (iii) sends through a terrestrial wireless or satellite link the map to the tactical computer available at each terrestrial vehicle commanded by a firemen NCO, driving to the fire area; (iv) sends a message to the duty seaplane indicating the exact fire position; (v) sends a message to the traffic management authorities in order to cut roads in the fire area; (vi) fire is serious and it is located close to inhabited areas, then the computer sends a message to the health authorities in order to send paramedics to the area.

By the time each land unit arrives to the fire area, they have received the exact optimal position to install their advanced command posts. Each unit, coordinated by a firemen NCO, would install a wireless network base, which would provide service to three fire extinction brigades formed by a corporal and four firemen. In the brigade, every firemen wears outside and inside temperature sensors; so as vital electromiographic sensing equipment. All of them are equipped with a GPS receiver. Two of the firemen wear a video camera installed over the helmet, and other two wear an IR camera. All the devices worn by each fireman are connected by means of a PAN (Personal Area Network), to an SBC equipped with a wireless transceiver, including also a microphone and an earphone to be used with VoIP. The SBC carried by the corporal, includes a display, which represents in real-time the updated and schematic version of the fire-fighting area assigned to his brigade. In the tactical computer of the first level C2, every received information is processed and the most relevant are retransmitted to the following level C2 system, which fuses them with the all the information received from the tactical C2 computers, together with the information received from the helicopter. The result is, for example, the transmission of a message to the specialized seaplane to indicate the crew where to drop the following water charge. The NCO knows at every time where the fire-fighters under his command are because they are blue points which move over the digital map of the area. If any point turns red because any vital sensor has surpassed the alarm threshold, the NCO gives the vocal order "group and abandon the area!", this order has maximum priority and is sent to every fire-fighter in the brigade. All the communications are IP-supported."

The initial proposed goals of the project, which have been kept during the whole duration of the project, are the following:

1. To define the architecture of a generic C4ISR system for emergency management, following the COTS development philosophy,
   - Application of the DoD standard working framework for the definition of C4ISR systems, to the emergency management systems.
   - Modification of the working framework in order to include multimedia flows with analysis of the implications in the three planes of the architecture definition (operational, systems, and technological).
   - Definition of the architecture operational plane
   - Definition of the architecture systems plane

- Definition of the technological plane
- Definition of the CADM (Core Architecture Data Model) of the C4ISR architecture for emergency management, by means of the data model NATO C2IEDM modification, including and modifying the objects related with emergency management

2. To define the emergency management system communications architecture
   - Definition of the communication hierarchical layers: operational, tactical, field and personal.
   - Definition of the network topology for each hierarchy level and level interconnection
   - Definition of the protocol architecture for each hierarchy level
   - Definition of the technologies for each hierarchy level

3. To validate the defined generic C4ISR architecture
   - Analysis and development of a reduced application for fire-fighting management
   - Systems integration
   - Functional evaluation of the application
   - Performance evaluation of the application

Every goal has been accomplished, except from the performance evaluation of the system through exhaustive and systematic field trials executions. The developed prototype under the SIMAGEM (**SI**stema de **MA**ndo y control para **G**estión de **E**mergencias **M**ultimedia) denomination. Throughout this paper to refer to the full system we will use this name: SIMAGEM.

In the development of the project we have achieved nearly all the foreseen goals in the first two years. We still have another year to continue with the development and we have decided to enlarge these goals with the following tasks:

- To introduce the representation of each brigade members position over a PDA for each brigade chief
- To use WIMAX wireless technology for giving support to the tactical network. This support could be extended to the combat field network when equipment compliant with IEEE 802.16e standard was available with mobility support.
- To introduce vocal communications based on VoIP which will be directly supported on system wireless networks.
- To introduce a system for automatic risk objects designation, by mean on integration of binocular telemetric devices.
- To introduce an automatic system for positioning own forces (Blue Force Tracking), with indoor positioning capabilities for complementing GPS information.
- To achieve international and interagency interoperability based on the Multilateral Interoperability Program (MIP) specifications for Command and Control Information Systems. These specifications include the addition of a common data interface for the system.

2 Success Level Reached in the Project

2.1 Statement of the Project

As we have indicated, in the first two years of the project development we have accomplished nearly all the initial goals, extending them with new additional goals, assumed within the project tasks and work packages. And proposing the follow-up of the project in the 2007 call of the "Plan Nacional de I+D+I", including fixed and mobile wireless sensor networks.

Regarding the stated goals:

- It has been defined a command and control information system architecture, specifically applied to emergency management, whose functionalities are described in a following section.
- It has been defined a communication architecture with four network layers: operational, tactical, field and personal.
- It has been developed a new system prototype: SIMAGEM, and a subjective evaluation of its functionalities has already been performed by expert personnel in tactical emergency management.

The development of the previous goals is stressed in the following sections.

2.2 C4ISR system architecture and functionality

Generally, the aim of military command and control information systems is to elaborate the Common Operational Picture (COP) at a high hierarchical level, i.e, from battalion level and upwards.

Moving this functionality to the civilian emergency management, it is needed to generate the COP at operating units of smaller size, e.g. fire-fighting brigades or specialized rescue teams, usually formed by a chief and a reduced amount of individuals, not surpassing the number of ten individuals.

In such a kind of operations, from the command and control point of view, it is fundamental to acquire a global vision of the situation (Situational Awareness, SA) by means of sharing each unit situation perception. Starting from the perception sharing among units (Shared Awareness), collaboration among them and a better effectiveness in reaching the mission goal are obtained.

Nowadays, this situation sharing is reached through unit's direct vision, maps, and vocal communications using the tactical radio combat network. In general, obtained SA is quite poor. Then, it will be extremely useful a command and control tool which: a) Allows obtaining the COP at fire brigade or rescue team level, virtually locating friend individual units over an operations theatre cartographic database with the suitable scalability at each level; b) Facilitates command and control decisions from one or from many tactical command and control locations, forward or rear; c) Allows self synchronization among fire brigades or rescue teams, shared awareness based. d)

Acquires individualized data from individual units, which therefore will act as sensors and actors simultaneously.

To develop all these objectives, we will start from two hypotheses: a) Availability of a field network capable of delivering digital data with the appropriated bandwidth; b) Automated information acquisition for COP generation. This process must be performed transparently to the individual units (fireman, member of a rescue team), that is, without their intervention in the information elaboration process and without interfering in their tasks.

In these scenarios, the person in charge, so as to obtain the enriched situation awareness, needs many times to "see through his own eyes" the operation scene, to make the proper tactical decisions. This requires the inclusion and fusion of multimedia flows, gathered from different sensors, particularly video flows.

The system architecture consists of various C2 nodes connected among them and to individual nodes with sensor/actuator functionality, by means of a wide capacity (tenths of Mbps) wireless data network, based in IEEE 802.11 WiFi standard in the current functional prototype. It is planned the migration to IEEE 802.16 WiMax.

The functional approach of the system can be summarized in the following points:

- Each individual member carries a SBC (Single Board Computer) which fuses data sensored from different sensors (video or infrared camera, GPS, biometric sensor, telemetric binoculars) and generates a unique information stream per individual element.
- Each individual unit SBC of a brigade is connected to the others by means of a wireless network, which access point is located at this brigade C2 post.
- First level C2 application generates the COP at fire brigade level since it receives data flows from individual units that constitute the brigade. The first level C2 post can manage a single brigade or a group of brigades, and usually will be installed in a tactical communications vehicle.
- C2 node can receive and fuse streams from other non-human sensor sources such as managed aerial vehicles (helicopter flying over operations theatre) or non managed (UAV) or fixed sensors distributed in the operations field (cameras with streaming capacity, presence sensors, etc).
- A brigade/group level C2 node can operate in the following three ways:
    1. Isolated from the others when, due to dimension of the mission, no more C2 levels are required. A unique COP at brigade/group level is generated.
    2. In an self-synchronized manner with other C2 nodes involved in the operation. In order to achieve this mode of operation, full wireless connectivity among involved C2 nodes is required. The COP is generated by the different C2 nodes, by mean of automatic replication of each node tactical database.
    3. Under the hierarchy of a higher level C2 post that coordinates several brigade/group level C2 nodes. Then, groups of brigades COP is automatically generated, at this high level C2 post.

Four screenshots of the current prototype application are included below to get a more graphic view of the ideas explained.

a) Group level, b) Brigades group level, c) Brigade level,
d) Individual element level

2.3 Network architecture

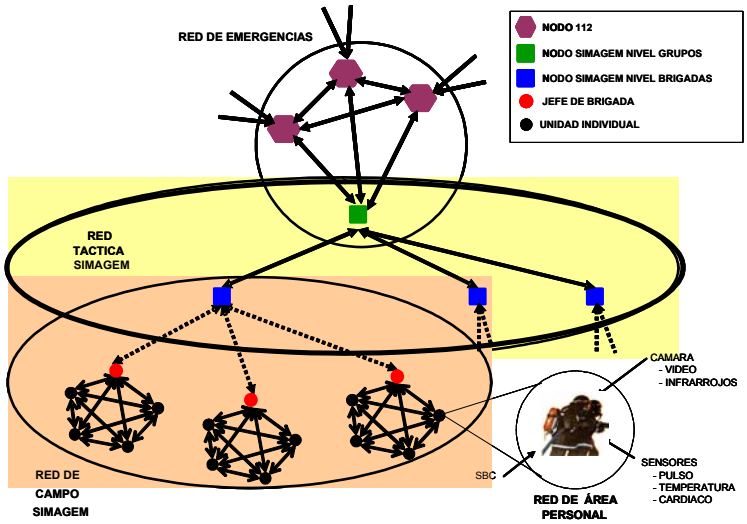
The main components of the SIMAGEM network architecture are the following:

- Personal Area Network, interconnect the sensors of each individual unit (brigade chief and fire-fighter) with his SBC. The PAN is implemented with Bluetooth and coaxial cable.
- SIMAGEM field network. IEEE 802.11 technology for the interconnection between the brigade chief; the individual units and the C2 Headquarters, where the wireless access point would be installed.
- SIMAGEM tactical network, interconnects C2 nodes of brigade group level among them and with the upper level C2 node. Currently the network is deployed using IEEE 802.11.

Some points have to be highlighted to the defined architecture:

- As an improvement of the prototype, we foresee the migration of the SIMAGEM field network to IEEE 802.16 WiMax. Mainly of the links between the C2 nodes of brigades group level and each brigade chief. The reason for this migration is twofold: (i) increase the radio coverage of the WiFi links (less than 1Km) to several Km using WiMax (5 Km without LOS and until 15 Km with LOS, although these values are an estimation and they depend on each vendor); (ii) improvement of indoor propagation, because with WiFi is nearly null.
- Due to the foreseen flexibility for the configuration of SIMAGEM, the C2 node of the brigades group level (or even a C2 node of brigades level if the operation has a reduced size) could be installed remotely with regard to the action scenario, e.g.: in a back headquarter, with the possibility of deploying the tactical network using satellite links, E1 links if available, or if the adequate bandwidth is available the regional or national emergency management radio network.

The C4ISR system network architecture is shown in the following figure:



SIMAGEM network architecture

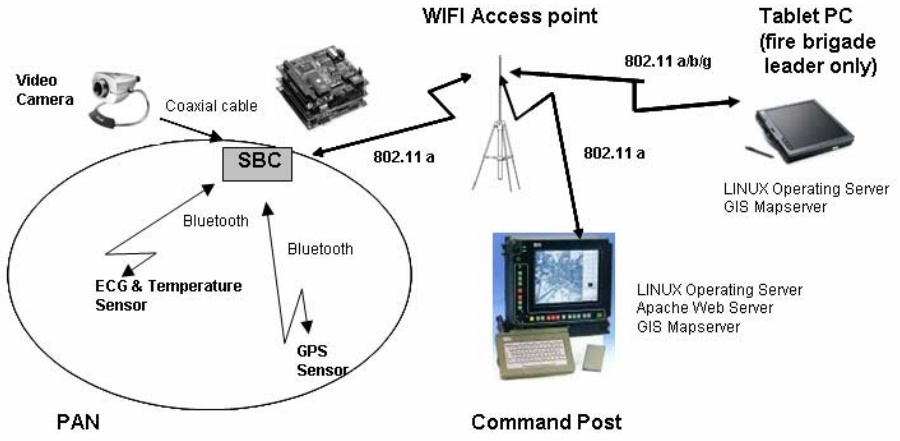
2.4 System prototype

Prototypes of a C2IS node and of the individual equipment for each individual unit, have been developed. Prototypes include the following technologies, all of them COTS and open source:

- Video cameras and grabbing cards.

- Biosensors: electrocardiogram (ECG), pulse and temperature.
- GPS.
- SBC with Linux operating system (distribution with minimal components).
- MPEG-4 codification and streaming software module.
- Data fusion module.
- Bluetooth and coaxial connectivity.
- C2 node computer with Linux operating system.
- Apache web server.
- Mapserver GIS.
- IEEE 802.11 WiFi connectivity, with immediate inclusion of IEEE 802.16 WiMax.
- Tablet PC with Linux, with Mapserver GIS, for fire brigade chief.

All these technologies are summarized in the following figure.



System prototype components

A first version of the system has been evaluated by command and control experts during a field demonstration (see annex with pictures of this demonstration), where video quality and overall system operability subjective aspects were evaluated. Some numeric results from this evaluation are (0 worst, 5 best):

Video quality at command post	4.5
Situation awareness improvement at command post	4
Decision support improvement	4.5
Application usability	4.5

3 Results Indicators

The project results till today have been published in the most renowned international conference in the field of command and control systems for civilian and military applications

- I.Pérez, C.Palau, M.Esteve, F.Carvajal, I.García-Juliá. "C4ISR Multimedia Framework for Situational Awareness Improvement". 10thICCRTS: International Command and Control Research and Technology Symposium. Malean, Virginia. EEUU. June 2005.
- F. Carvajal, I. Pérez, C. Palau, M.Esteve, "Adaptation of C2IEDM for Civil Crisis Management". 2006 International Command and Control Research and Technology Symposium San Diego, California. EEUU. June, 2006.
- Manuel Esteve Israel Pérez, Carlos Palau, Federico Carvajal "Command and Control Information Systems for Little Units: SIMACOP, a Research System to Test New Concepts" 11th ICCRTS: Coalition Command and Control in the Networked Era. Cambridge, UK. September, 2006.

The experience acquired in the previous project, has allowed the inclusion of the research group in three European projects related with command and control systems applied to emergency management. The projects are next briefly described. All of them are funded by EU 6th Framework Program. We have to highlight that in these projects, we are currently working with the main industrial companies of the sector, so as with the most prestigious universities in Europe in the scope of information systems, communications and emergency management.

MARIUS (Mobile Autonomous Reactive Information system for Urgency Situations). PASR Program (Preparatory Action on Security Research).
Project co-ordinated by EADS France, with the participation of Thales, Eurocopter, Selex, Marconi, Amper Programas and Cranfield University. The project began on the first week of March 2006, and its main goal is the development of a aero-transported tactical headquarter, for the management of the first hours of an emergency scenario. The headquarter will integrate and fuse information from different origins: UAV, video and IR from a helicopter, individual terrestrial video cameras); a C2IS for resources management (Zodiaco system from AMPER) and a modulus for decision support. One of the underlying goals of the project is the viability proof of using military technologies in the scope of joint civilian-military applications, within Europe or in detachments outside Europe (peace-keeping missions, catastrophes like tsunamis or earthquakes). UPV is responsible of the integration of terrestrial video cameras system integration, which is a direct result of the project.

DIVYNE (Dinamic Visual Networks). Programa IST (Information Society Technologies).
Project also co-ordinated by EADS France with some partners in common with MARIUS and other companies and universities leading the research in the C2IS field, like Catholic University of Leuven; University of Surrey and EPFL. The project started in September 2006, the main goal of the project is the development of a system allowing the interconnection and joint management from a headquarter and the management of the visual sensor networks of a certain city or a region (mainly video cameras) in case of a catastrophe.

CITRINE (Common Intelligence and Traceability for Rescues and IdentificatioN operations). PASR Program (Preparatory Action on Security Research).
Project co-ordinated by Thales, with other relevant partners like EADS, Selex, Galileo Avionica and Skysoft. The project started in January 2007. The main goals of the project can be summarized in: (i) development of a first version of a modular and scalable system of damages valuation in a catastrophe scenario from the data acquired by different sensors and other sources of human information (messages, reports,…), composing a synthetic picture of the situation that can help in the process of decision making on the catastrophe management; (ii) development of a support tool for decision making for the operational catastrophe management allowing the coordination of the different groups involved by means of task assignment, routes,…

Currently, the following researchers are working in their PhD Thesis in the field of command and control for emergency management, with activities directly related with the described research project. The researchers, the title of the PhD Thesis dissertation and the expected defence date follow:

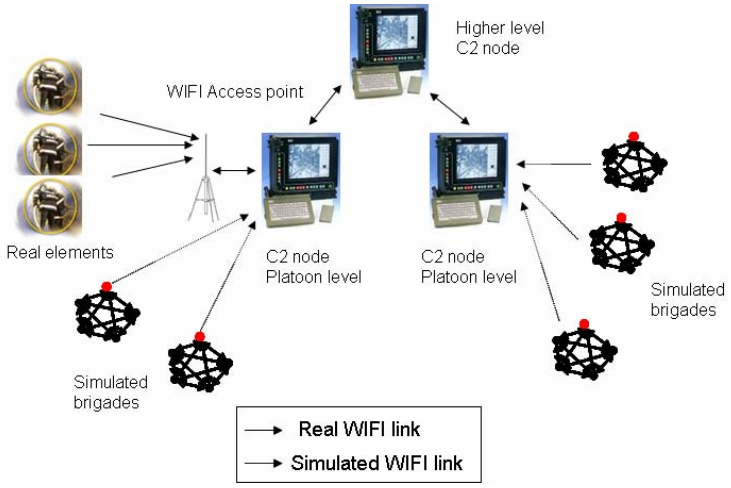
- Israel Pérez Llopis. "Arquitectura C4ISR multimedia architecture for emergency management". July 2007.
- Federico Carvajal Rodrigo. "Adaptation of the tactical data model C2IEDM for command and control information systems applicable to emergency management". December 2007.
- Luis Ernesto Hernández Blanco. "Wireless data networks for command and control information systems for emergency management". July 2008.

4 References

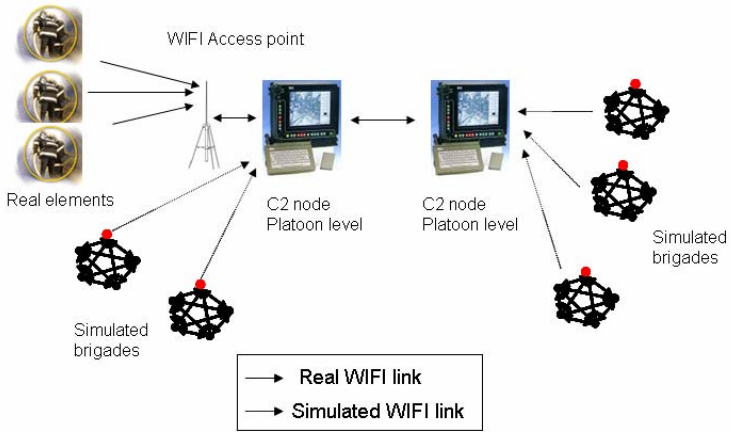
A. Günther, C. Hoene "Measuring Round Trip Times to Determine the Distance between WLAN Nodes", International IFIP-TC6 networking Conference, Waterloo, Canada, 2005
Alberts, Hayes "Understanding Comand and Control", CCRP Publications, 2006
Alberts, Hayes, "Power to the Edge", CCRP Publications, 2003
H. S. Ng, M. L. Sim, C. M. Tan y C. C. Wong. "Wireless technologies for telemedicine", BT technology journal, Vol 24 No 2, April 2006.
M. Donahoo, R. Steckler, "Emergency mobile wireless networks", IEEE MILCOM 2005.
W. Konhauser et al,"Broadband Wireless Access Solutions", Wireless Personal Communications, Issue 37, pags 243 – 259. 2006
I. Opperman, L.Stoica, A. Rabbachin, Z. Shelby "UWB wireless Sensor Networks: UWEN – a practical example" IEEE Communications Magazine, vol. 42, issue 12, December 2004
J.N., Al-Karaki and A.E., Kamal, "Routing techniques in wireless sensor networks: a survey", IEEE Wireless Communications, vol. 11,no. 6, pp.:6-28, December 2004.
I.Pérez, C.Palau, M.Esteve, F.Carvajal, I.García-Juliá. "C4ISR Multimedia Framework for Situational Awareness Improvement". 10thICCRTS. June 2005.
F. Carvajal, I. Pérez, C. Palau, M.Esteve, "Adaptation of C2IEDM for Civil Crisis Management". 2006 International Command and Control Research and Technology Symposium ,June, 2006.
Manuel Esteve Israel Pérez, Carlos Palau, Federico Carvajal "Command and Control Information Systems for Little Units: SIMACOP, a Research System to Test New Concepts" 11th ICCRTS: September, 2006.

Annex:

Configuration and pictures of C4ISR multimedia system field demonstration



Self-synchonized configuration



Hierarchical configuration


WiFi access points at C2 post


WiFi antennas


SBC with WiFi card and video camera mounted on helmet


Application running at C2 post

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Verified systems for reasoning in the Semantic Web
TIN2004–03884

José A. Alonso Jiménez *
Dpto. de Ciencias de la Computación e Inteligencia Artificial
Universidad de Sevilla

Abstract

The main goal of this project is to obtain tools for increasing our confidence in the Semantic Web. More specifically, the proposal is the design of formally verified reasoning systems that can be used for reasoning in the Semantic Web, as well as the development of automatized methodologies for intelligent cleaning of ontologies. In the long term, the purpose is the integration of the solutions obtained for the above problems in a framework where it could be possible to develop trusted intelligent systems for the Semantic Web. This framework would allow not only the development of programs for practical applications, but the possibility of applying formal methods for proving their correctness.

Keywords: Formal methods, Automated Reasoning, Semantic Web.

1 Goals of the project

The main intended goal of the Semantic Web is the possibility of accessing the information on the Web in a way such that an intelligent system were able to process it in trusted way. In the future, in the task of building trusted automated reasoning systems for the Web, it will be necessary, on one hand, to formally verify such systems and, on the other hand, to establish mechanisms for reusing them in generic frameworks where the application of mechanized deductive methods will be simple for the end user. In this way, the logical confidence in the deductive processes would support the confidence in the Semantic Web. There is no doubt in the interest of applying formal methods for obtaining trusted systems.

The specific goals of this project are the following:

1. To design formally verified reasoning systems that could be used in the verification of ontologies, knowledge bases, and digital libraries.

2. To develop automated methodologies for intelligent cleaning of data bases in the Semantic Web, using automated reasoning systems for repairing semantic anomalies.

*Email: jalonso@us.es

3. To integrate formally verified reasoning systems in the development of provers for reasoning with data bases associated with ontologies.

For the development of formally verified applications for the Semantic Web, we intend to apply a methodology that turned out fruitful in other verification efforts carried out by members of the group, with a positive impact in the implementation as well as in the formal proofs of their properties. According to this methodology, an stepwise refinement strategy is applied, beginning with an abstract specification of the application and ending with the last implementation details. More specifically, our goal is to design a generic framework where, without going into implementation details, it could be possible to reason about the essential properties of a prover. By successive refinements, and starting in the generic framework, it would be possible to deal with optimization aspects for the real implementation, as the use of efficient data structures or sophisticated heuristics. A final refinement step would adapt the implementation for its use as reasoning system for the Semantic Web.

2 Level of success reached by the project

The activities carried out up to this moment, can be classified according to two different lines of work: i) the design of formally verified reasoning systems that could be applied for reasoning in the Semantic Web, and ii) the development of automated methodologies for repairing ontologies.

The starting point for the first line of work was the formal verification in ACL2 of a generic framework for developing propositional theorem provers [1]. This generic framework can be instantiated in order to obtain executable (and formally verified) propositional provers in an automated way. Among the possible executable instances, the semantic tableaux method and the Davis–Putnam–Logeman–Loveland are especially remarkable. The second problem we have studied is the trade-off between formal verification effort and the efficiency of the verified implementation; as a case study we have developed a formally verified executable implementation of a quadratic unification algorithm, using a dag representation for first-order terms [12, 13, 17]. The main conclusions of this case study are the following: formally verified efficient algorithms are possible (the ACL2 quadratic algorithm is executed at about 60% of the speed of the corresponding C implementation) but the preconceived idea that algorithms employing more complex data structures or more sophisticated control structures require more verification efforts is confirmed (in this work we needed 214 definitions and 775 theorems in contrast to the 19 definitions and 129 theorems needed in the verification of a naive implementation of unification using a prefix representation of terms that was previously carried out).

The third problem we have studied is the formalization of the $\mathcal{ALC}$ description logic in PVS [5] and the formal verification of the reasoning in $\mathcal{ALC}$ [14]. A proof of the termination of the reasoning in $\mathcal{ALC}$ has leaded us to the formal study of the well–foundedness of the multiset order, developing in PVS the corresponding theory [7].

In the second line of work, we have analyzed and systematized the conceptual anomalies that can appear in an ontology, using the region connection calculus RCC as a metaontology. These results have been presented in the PhD Thesis of A.M. Chávez [3] and in the paper [9]. In addition, in [3] we have studied the reduction of several problems about anomalies in ontologies to problems that require reasoning on lattices, that can be solved using automated reasoning

techniques [4]: the problem of the deficient expressiveness/specification of formal ontologies has been dealt with in [4] (expressiveness) and in [6] (uncertainty due to the deficient specification of elements of the ontology); the problem of ontology fusion has been dealt with in [10, 15]; and the problem of the extension suggested by the data in [16]. In these five papers a logical notion of robust extension (*lattice categoricity*) has been formalized, allowing us to apply automated reasoning techniques to solve the problems.

Finally, in [11], we have analyzed, from the point of view of automated reasoning, the challenges posed by ontology repairing, at the cognitive and logical levels. In addition, we propose an OWL extension for obtaining specialized provers from the certified generic framework [1].

3 Indicators of the results

In the development of the project a PhD thesis [3] has been finished and the following papers has been published:

[1] F.J. Martín, J.A. Alonso, M.J. Hidalgo and J.L. Ruiz: *Formal verification of a generic framework to synthesize SAT-provers.* Journal of Automated Reasoning, Vol. 32, N. 4, 2004, pp. 287-313.

[2] J.A. Alonso, J. Borrego and A.M. Chávez: *Logic Databases and Inconsistency Handling.* Encyclopedia of Databases Technologies and Applications. Hershey, Pa, USA. Idea Group Inc. 2005, pp. 336-340. ISBN: 1-59140-560-2

[3] A.M. Chávez: *Razonamiento mereotopológico automatizado para la depuración de ontologías,* PhD Thesis, University of Seville, 2005. (In spanish).

[4] J. Borrego and A.M. Chávez: *Extension of Ontologies Assisted by Automated Reasoning Systems.* In "Computer Aided Systems Theory (EUROCAST 2005)" (10th International Conference on Computer Aided Systems Theory, Las Palmas de Gran Canaria, Spain, February 7–11, 2005, Revised Selected Papers). Lecture Notes in Computer Science Vol. 3643, Springer, 2005, pp. 247–253.

[5] J.A. Alonso, M.J. Hidalgo, F.J. Martín and J.L. Ruiz: *Formalización de la lógica descriptiva ALC en PVS.* Technical Report of Computational Logic Group, University of Seville, 2005. (In spanish).

[6] J. Borrego and A.M. Chávez: *Controlling Ontology Extension by Uncertain Concepts Through Cognitive Entropy.* Proceedings of the ISWC Workshop on Uncertainty Reasoning for the Semantic Web, 2005, pp. 56-66. `http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-173/paper6.pdf`

[7] J.A. Alonso, M.J. Hidalgo, F.J. Martín and J.L. Ruiz: *Formalización en PVS de la buena fundamentación del orden de los multiconjuntos.* Technical Report of Computational Logic Group, University of Seville, 2005. (In spanish).

[8] F.J. Martín, J.L. Ruiz, J.A. Alonso and M.J. Hidalgo: *Proof Pearl: a Formal Proof of Higman Lemma in ACL2.* In "Theorem Proving in Higher Order Logics" (18th International Conference, TPHOLs 2005, Oxford, UK, August 22–25, 2005. Proceedings). Lecture Notes in Computer Science. Vol. 3603, 2005, pp. 358-372

[9] J. Borrego and A.M. Chávez: *Visual Ontology Cleaning: Cognitive Principles and Applicability.* In "The Semantic Web: Research and Applications" (3rd European Semantic Web Conference, ESWC 2006 Budva, Montenegro, June 11-14, 2006 Proceedings). Lecture Notes in Computer Science, Vol. 4011, 2006, Springer, pp. 317–331.

[10] J. Borrego and A. M. Chávez: *Fusión automatizada de ontologías: Aplicación al razonamiento espacial cualitativo.* Actas del Campus Multidisciplinar en Percepción e Inteligencia, CMPI–2006, Vol. I, pp. 91–102.

[11] J.A. Alonso, J. Borrego, A.M. Chávez and F.J. Martín: *Foundational Challenges in Automated Semantic Web Data and Ontology Cleaning.* IEEE Intelligent Systems. Vol. 21. N. 1, 2006, pp. 42-52

[12] J.L. Ruiz, F.J. Marín, J.A. Alonso and M.J. Hidalgo: *Formal Correctness of a Quadratic Unification Algorithm.* Accepted for publication in the Journal of Automated Reasoning. (Available Online First).

[13] D.A. Greve, M. Kaufmann, P. Manolios, J S. Moore, S. Ray, J.L. Ruiz-Reina, R. Sumners, D. Vroon, and M. Wilding:. *Efficient execution in an automated reasoning environment.* Techical report TR–06–59, November 2006. Department of Computer Sciences, University of Texas at Austin.

[14] J.A. Alonso, J. Borrego, F.J. Martín and J.L. Ruiz: *Semantic Web Verification: Verifying Reasoning in the Logic ALC.* Submitted to ESWC 2007 (4th European Semantic Web Conference, 3-7th, June 2007).

[15] J. Borrego and A. M. Chávez: *Merging Formal Ontologies Assisted by Automated Reasoning Systems.* Under consideration for publication in the Journal of Logic, Language & Information.

[16] J. Borrego and A. M. Chávez: *A Formal Foundation for Knowledge Integration of Deficient Information in the Semantic Web.* Computer Aided Systems Theory (EUROCAST 2007), February 12–16, 2007.

[17] D. A. Greve, M. Kaufmann, P. Manolios, J S. Moore, S. Ray, J.L. Ruiz–Reina, R. Sumners, D. Vroon and M. Wilding: *Efficient Execution in an Automated Reasoning Environment.* Accepted to appear in the Journal of Functional Programming.

Project Monitoring Sessions, 2007
National Programme on Computer Technology (TIN)

Development of Support Services for Linguistic Research over the Internet TIN2004-03988

Octavio Santana Suárez[*]
Main Researcher of the Project
Department of Computer Science and Systems
University of Las Palmas de Gran Canaria

José R. Pérez Aguiar[**]
Research Member of the Project
Department of Computer Science and Systems
University of Las Palmas de Gran Canaria

Francisco Javier Carreras Riudavets
Research Member of the Project
Department of Computer Science and Systems
University of Las Palmas de Gran Canaria

Zenón J. Hernández Figueroa
Research Member of the Project
Department of Computer Science and Systems
University of Las Palmas de Gran Canaria

Juan Carlos Rodríguez del Pino
Research Member of the Project
Department of Computer Science and Systems
University of Las Palmas de Gran Canaria

Margarita Díaz Roca
Research Member of the Project
Department of Computer Science and Systems
University of Las Palmas de Gran Canaria

Juan de Dios Duque Martín de Oliva
Research Member of the Project
Department of Computer Science and Systems
University of Las Palmas de Gran Canaria

Gustavo Rodríguez Rodríguez
Collaborator on the Project
Department of Computer Science and Systems
University of Las Palmas de Gran Canaria

Abstract

The objective of this project is to place a set of remote services and clients at the disposal of the international community over the Internet in order to computationally solve linguistic phenomena of the Spanish language. The implemented services are as follows: a remote service of morphological analysis, a remote service of information on morpholexical relationships and a remote service of functional disambiguation. These services allow access to any authorized remote application by means of the inclusion of the corresponding definition document. Additionally, a client of morphosyntactic analysis of texts and a morpholexical client of information recovery have been developed. Both clients are end-use tools that put at stake the potentiality of services.

Keywords: natural language processing, computational linguistic, morphology, syntax.

[*] E-mail: osantana@dis.ulpgc.es
[**] E-mail : jperez@dis.ulpgc.es

1 Aims of the project

Remote services and their clients are innovative technology, which are based on the use of open standards so as to promote cooperative development by taking full advantage of the potentiality of the Internet. In accordance with point 3.6 of the National Programme on Computer Technology, this project aims at developing:

S1. A remote service of morphological analysis.
This service enables the lemmatization of any word of the Spanish language when identifying its canonical form, grammatical category and the inflection or derivation that produces it. With regard to verbs, the service considers the simple and compound conjugation, enclitic pronouns, the inflection of the participle as a verbal adjective (gender, number) and the diminutive form of the gerund. As far as non-verbal forms are concerned, the service considers the following: gender and number for nouns, adjectives, pronouns and articles; heteronyms due to a change of gender for nouns, superlative degree for adjectives and adverbs, adverbialization of the superlative for adjectives; derivation for nouns, adjectives and adverbs; invariable forms such as prepositions, conjunctions, exclamations, words from other languages and idiomatic expressions or phraseology. Prefixation is taken into consideration when applicable.

S2. A remote service of information on morpholexical relationships.
This service enables the recognition, generation and manipulation of the morpholexical relationships from any given word. It includes the retrieval of all its lexicogenetic information until reaching the primitive, the management and control of the affixes in the treatment of its relationships and regularity in the established relationship. It conveys a global perspective of the behaviour and productivity of Spanish words in the main word-formation processes (suffixation, prefixation, parasynthesis, suppression, regression, zero-modification, apocopation, metathesis and other non-classifiable processes that generate alternative graphical forms).

S3. A remote service of functional disambiguation.
This service offers the grammatical function of every voice in its respective context. It minimizes the possibilities due to its treatment of the syntactic structures.

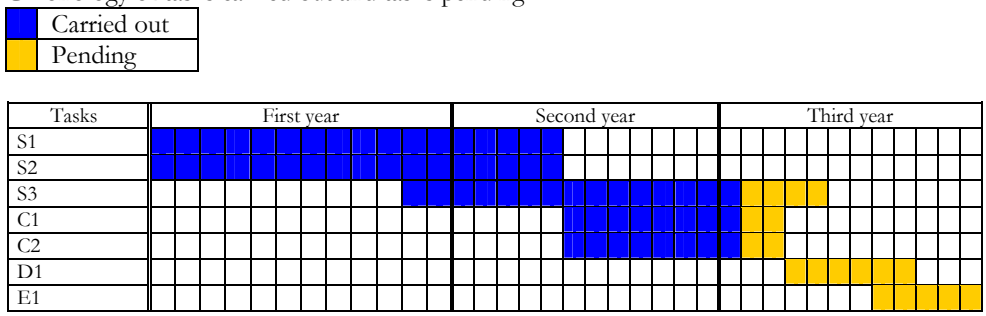
C1. A client of morphosyntactic analysis of texts.
Using the previous services and a user-friendly interface, this client allows users to obtain the morphosyntactic analysis of the chosen text; some statistical measures of its characteristics; the neologisms; and the location of grammatical co-occurrences, verbal periphrasis, lexical collocations and other linguistic phenomena.

C2. A morpholexical client of information recovery.
Using the previous services, this client can locate the documents that meet the specified search requests on the Internet, for instance, specific words affected to a greater or lesser extent by the various existing word transformation mechanisms in the Spanish language, or grammatical characteristics or linguistic phenomena that may appear in the document.

D1. Technical and user documentation.
The technical and user documentation are of utmost importance in order to offer the service or distribute the client to potential users with the assurance that they will be able to use it effectively.

E1. Control of the service or client under real conditions.
Control of the service or client under real conditions. Once the service is offered or the client is distributed to its users, based on the comments they make, a feedback phase focusing on any of the previous tasks will take place. This will entail adjustments and corrections to its operation. Moreover, this phase lasts indefinitely, although, for the purposes of this project, an initial phase of limited duration, during which said feedback will be stronger, is estimated.

Chronology of tasks carried out and tasks pending.



2 Success level achieved in the project

Of the aforementioned objectives, those concerning the remote service of morphological analysis and the remote service of information on morpholexical relationships have been met. Work is currently underway on the remote service of functional disambiguation and on the two clients that have a high percentage of performance. Still pending is the technical and user documentation that was initially going to be prepared at the end of every service but which is now going to be moved to the final year of the project in order to homogenize it and draft it when all services and clients are operating. Accordingly, human resources have been allocated to other more complex tasks in order to complete them within the intended timeline and to efficiently solve the problems encountered.

2.1 Remote service of morphological analysis (S1)

The objective of the morphological service is to support the morphological study of words over the Internet. It is developed for the ".NET" platform as an "ASP.NET" web service that uses C# as the development language in the Microsoft Visual Studio setting. This allows for any authorized remote application to access the service by means of the inclusion of the corresponding definition document in a *Web Services Description Language* (WSDL) format. This can considerably encourage advances in research, since other groups whose work needs to rely on the morphology of the

Spanish language can develop their own applications by using this service, without having to develop their own morphological tools

This service is based on a morphological engine, implemented in C++ as a *Dynamic Link Library* (DLL) of MS-Windows, whose potential is endorsed by the 4,951,802 inflected words taken from all entries in the Dictionary of the Spanish Language (DRAE), the General Dictionary of the Spanish Language (VOX), the Dictionary of Current Spanish Usage (CLAVE SM), the Dictionary of Synonyms and Antonyms (ESPASA CALPE), the Ideological Dictionary of the Spanish Language (JULIO CASARES) and the Dictionary of Contemporary Spanish Usage (MANUEL ALVAR EZQUERRA). Furthermore, in the region of 1240 proper nouns not included as entries in the consulted sources are added. These are related to nationality adjectives and other adjectives and nouns, such as 'follower of' or 'doctrine' —Marxist, Marx's Marxism— among other meanings. In the region of 9000 adjectives derived from verbal participles that have not been included in said sources have also been added. The universe contemplated by the service is made up of 195,743 canonical forms —181,593 non-verbal and 14,150 verbal— that have generated 4,951,802 inflected or derived forms. Moreover, this does not contemplate the extension inherent to the incorporation of prefixes and enclitic pronouns —more than 4000 million words— that the morphological analysis system also includes at a success rate of 100%. With the exception of the recognition of prefixes and enclitic pronouns, it is implemented on the basis of data and not rules. The database system can easily be updated to include possible changes in language or errors resulting from manual processing. The service operates with a word, supplied by a remote client, that is lemmatized in order to return its morphological characteristics by means of a SOAP (*Simple Object Access Protocol*).

2.2 Remote service of information on morpholexical relationships (S2)

As regards the construction philosophy, this service has the same characteristics as the morphological analysis service described in the previous section: an "ASP.NET" service written in C# that is supported on a DLL of morpholexical relationships written in C++.

There is a morpholexical relationship between two Spanish canonical forms when one has been formed from the other by means of a word formation process in the Spanish language: suffixation, prefixation, parasynthesis and others. There is a suffixal morpholexical relationship between two words when one has been formed from the other by adding a suffix; thus its semantic and functional aspect is generally altered —footballer has a semantic and functional relationship with football. There is a prefixal morpholexical relationship between two words when one has been formed from the other by adding a prefix; thus its semantic aspect is slightly altered. Incorporating prefixes does not usually alter the functionality of the root word —antechamber has a semantic and functional relationship with chamber. There is a parasyntetic morpholexical relationship between two words when one has been formed from the other by adding an affix in front of the word and another one behind it —usually a prefix and a suffix—; thus the semantic and functional aspect of the new word varies with respect to the root word —irrationalism has a semantic and functional relationship with rational. Another kind of morpholexical relationship is established if two words have undergone one of the following processes: suppression, regression, zero-modification, apocopation, metathesis and other non-classifiable processes that generate alternative graphical forms. The information engine of morpholexical relationships —implemented according to data— is capable of obtaining information on approximately 70,147 suffixal morpholexical relationships,

11,113 prefixal morpholexical relationships, 3812 parasyntetic relationships and 4694 morpholexical relationships of other kinds.

This service enables selecting between three options that give access to various degrees of information on words that are morpholexically related to some data. The service operates with a word and its grammatical category that can be obtained from the morphological analysis service. The result is a list of words which, according to the chosen option, will only be formed by the nearest ones, those with an intermediate degree of relationship or all those that are in some way related.

2.3 Remote service of functional disambiguation (S3)

As far as the construction philosophy is concerned, this service that is still being developed has the same characteristics as the service described in the previous section: an "ASP.NET" service written in C# that is supported on a DLL of disambiguation and on a DLL of morphological analysis that uses the remote service of morphological analysis.

There is a considerable amount of words in the Spanish language that can have various grammatical functions and, consequently, a textual analysis could produce a great amount of combinations unless the function of each word within the context in which it appears is considered. Functional disambiguation consists of eliminating the results that do not correspond to their function in the text.

This service uses a functional disambiguation method that reduces the size of the answer of the morphological processor into two phases. In the first phase, a functional disambiguation method based on local syntactic structures is applied and those grammatical functions that are incompatible with the immediate context of each word in the phrase are ruled out. In the second phase, a structural functional disambiguation is carried out and the combination of grammatical functions of the phrase that do not produce a valid syntactical representation tree are ruled out.

Studies are currently focusing on whether applying the second phase is worthwhile, since the loss of performance inherent to the application of the second phase does not seem to justify its inclusion even though a disambiguation percentage of 87% can be achieved with the first phase. This figure reaches 96% after applying the second phase.

2.4 Client of morphosyntactic analysis of texts (C1)

The client of morphosyntactic analysis of texts is developed as an "ASP.NET" web application that interacts with the user on the one hand, and with the remote services described above on the other.

This application offers the user two possibilities: writing a text directly in the window for that purpose or extracting it from a file. Once the text to be analysed has been selected, the application can locate recognised and non-recognised words, indicating where they appear and their occurrences; look for co-occurrences of various words within a given radius; find occurrences of known idiomatic expressions; and analyse the occurrence of text segments with a specified minimum degree of repetition.

2.5 Morpholexical client of information recovery (C2)

The objective of this client is make it possible for information to be recovered from the Internet, selected by means of a search pattern that includes inflective and derivative morpholexical characteristics, offering far greater scope than that offered by most search engines whose inflective capacity is reduced.

The client is developed as an MS-Windows application that can be installed onto the user's computer and connected to the previously described services over the Internet. Carrying out a search requires a configuration of two different areas: the field and the pattern of the search. The field is understood to be the set of websites from where the documents to be analysed will be taken; those documents fitting to the specified pattern will be selected, for instance, the search can be focused on documents that fulfil a given pattern, specifying as the field of the search all Spanish universities, just the University of Las Palmas de Gran Canaria or even a department of said University. The pattern of the search can vary from an exact word to the inclusion of a great deal of characteristics based on linguistic phenomena, for instance, documents containing the sequence of words "inflation housing youth" can be located where the words of the sequence can appear inflected or derived in proportions indicated by the user and separated from one another by a certain distance.

3 Results Indicators

Two books on morpholexical relationships have been published:

- *Relaciones morfoléxicas prefijales para el procesamiento del lenguaje natural.*
  Santana, O.; Carreras, F.; Pérez, J.
  Editorial MILETO; ISBN: 84-95282-92-5. Madrid, 2005.
  Pages: 116.
- *Relaciones morfoléxicas parasintéticas para el procesamiento del lenguaje natural.*
  Santana, O.; Carreras, F.; Pérez, J.
  Editorial MILETO; ISBN: 84-95282-96-8. Madrid, 2006.
  Pages: 156.

The following articles have been published in national and international journals:

- NAWeb: un navigateur et analyseur morphologique des pages web pour l'espagnol.
  Santana, O.; Hernández, Z.; Rodríguez, G.
  Cahiers de lexicologie. Revue internationale de lexicologie et de lexicographie, issue 87- 2005-2. ISSN: 0007-9871 (29/43).
  2005.
- Functional Disambiguation Based on Syntactic Structures.
  Santana, O.; Pérez, J.; Losada, L.; Carreras, F.
  Literary and Linguistic Computing, Vol. 21, issue 2, (187/197).
  2006.

The following papers have been presented at international conferences:

- Una aplicación para el procesamiento de la sufijación en español.
  Santana, O.; Carreras, F.; Pérez, J.; Rodríguez, G.

IX Simposio Internacional de Comunicación Social, *Actas*, Vol. II. ISBN: 959-7174-05-7, (623/629).
January, 2005.

- Software Application for Parasynthesis in Spanish Automatic Processing.
  Santana, O.; Carreras, F.; Pérez, J.; Rodríguez, J.C..
  The 2005 International Conference on Machine Learning; Models, Technologies and Applications. MLMTA'05. Proceedings. ISBN: 1-932415-73-4, (46/52).
  June, 2005.
- Spanish Morphosyntactic Disambiguator.
  Santana, O.; Pérez, J.; Losada, L.; Carreras, F.
  The 17th Joint International Conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing. ACH/ALLC 2005. Conference Abstracts. ISBN: 1-55058-308-5, (207/209).
  June, 2005.
- Una Aplicación para el Procesamiento de la Prefijación en Español.
  Santana, O.; Carreras, F.; Pérez, J.; Rodríguez, G.
  4th Multi-Conference on Systems, Cybernetics and Informatics. CISCI 2005. *Memorias*, Vol. II. ISBN: 980-6560-38-8, (322/327).
  July, 2005.
- Parasynthetic Morpholexical Relationships of the Spanish: Lexical Search beyond the Lexicographical Regularity.
  Santana, O.; Carreras, F.; Pérez, J.; Rodríguez, J.
  Proceedings of the IADIS International Conference. Applied Computing. 2006. ISBN: 972-8924-09-7, (627/631).
  February, 2006.

4 Bibliography

[1] Santana, O.; Carreras, F.; Pérez, J.: *Relaciones morfoléxicas sufijales para el procesamiento del lenguaje natural*. Editorial MILETO; ISBN: 84-95282-91-7. Madrid, 2004. (178 pages).

[2] Santana, O.; Carreras, F.; Pérez, J.; Rodríguez, G.: Relaciones morfoléxicas prefijales del español. *Boletín de Lingüística*, Vol. 22. ISSN: 0798-9709, (79/123). July - December, 2004.

[3] Santana, O.; Hernández, Z.; Rodríguez, G.: DAWEB: Un descargador y analizador morfológico de páginas Web. *Procesamiento de Lenguaje Natural*, Revista Nº 30. Ed. SEPLN. ISSN: 1135-5948, (75/87). March, 2003.

[4] Santana, O.; Hernández, Z.; Rodríguez, G.: Morphoanalysis of Spanish Text: Two Applications for Web Pages. *Lecture Notes in Computer Science* 2722. Web Engineering. Ed. Springer-Verlag. ISSN: 0302-9743. ISBN: 3-540-40522-4, (511/514). July, 2003.

[5] Santana, O.; Pérez, J.; Carreras, F.; Duque, J.; Hernández, Z.; Rodríguez, G.: FLANOM: Flexionador y lematizador automático de formas nominales. *Lingüística Española Actual* XXI, 2, 1999. Ed. Arco/Libros, S.L. (253/297).

[6] Santana, O.; Pérez, J.; Carreras, F.; Rodríguez, G.: Suffixal and Prefixal Morpholexical Relationships of the Spanish. *Lecture Notes in Artificial Intelligence*, 3230. Ed. Springer-Verlag. ISSN: 0302-9743, (407/418). October, 2004.

[7] Santana, O.; Pérez, J.; Hernández, Z.; Carreras, F.; Rodríguez, G.: FLAVER: Flexionador y lematizador automático de formas verbales. *Lingüística Española Actual* XIX, 2, 1997. Ed. Arco/Libros, S.L. (229/282).
[8] Santana, O.; Pérez, J.; Losada, L.; Carreras, F.: Bases para la desambiguación estructural de árboles de representación sintáctica. *Procesamiento de Lenguaje Natural*, Revista Nº 32. Ed. SEPLN. ISSN: 1135-5948, (43/65). March, 2004.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

ENLACE: Un entorno basado en agentes para comunidades de aprendizaje (escuela y naturaleza, lugares para aprender, colaborar y experimentar) TIN2004–04232

M. Felisa Verdejo
LTCS Group, Depto de Lenguajes y Sistemas Informáticos
UNED

Abstract

The diversity of knowledge, experience and resources (both real and virtual) that are available in an increasing number of scientific disciplines, together with the wide range of mobile electronic devices, software modelling, simulation and collaboration tools, and distributed and ubiquitous computational technologies, offer an enormous potential for the exploration of new ways to teach and learn. The purpose of the ENLACE project is to explore the design of innovative educational environments, which offer intelligent support for a wide range of learning activities in natural science related domains. From a computer science perspective, the use of agent technology for modelling and providing active support of activities in complex organizations, together with the utilization of the 'semantic Web' to organize and structure shared knowledge and distributed resources, represent the technological baseline needed to achieve the objective of the present project: the modelling and design of open environments with intelligent community learning support where the learning strategies, the collaboration mechanisms and the tools available, offer additional functionality to different work environments, experimental scenarios, and formal levels of social organization: individual, small groups, large groups, and communities. A fundamental part of this investigation is the generation of a semantic portal for learning communities in the area of environmental education, using a participative design methodology and formative evaluation to verify the viability of the proposal.

Keywords: educational networking infrastructure, ubiquitous learning, learning communities

1 Introduction: approach, tasks and chronogram

In the ENLACE project we are exploring the design and implementation of a technological infrastructure, as an educational network platform offering services that will provide ubiquitous web applications for learning activities both inside and outside the classroom in order to improve student learning and teachers' work. A guiding principle in our approach is to inform the technology design with a set of rich scenarios grounded on current learning sciences research on socio-constructivist learning. We are implementing with teachers and experts, scenarios and the

technological functionality for supporting them, to foster active and constructive learning process, meaningful related to different subjects, where the study of a problem is considered through different facets and methods. In our vision students are engaged both in individual and collaborative activities that promote their curiosity by inquiring about open questions, directly exploring and observing their environment, compiling information, analyzing and representing data, building models to explain and predict behaviour, exchanging and discussing their findings, linking their view with real problems faced by professionals and contributing to create a shared memory with personal traces in a learning community. For example, a middle school teacher proposes a question to their students: Why does this bird live in this place/habitat in spring and summer time? This question acts as the thread for a series of activities in differently localized scenarios, including a field trip to a nature park to experience and identify that particular habitat and raises not only topics related to the natural science subjects (flora, fauna, habitat, climate), but to music (motivated by the need to identify the sounds of the birds), to geography (topographic and relief maps), to mathematics (to calculate and represent distances using different measures) motivating also teachers to coordinate themselves and integrate their respective agenda into a larger perspective. The learning workflow includes activities for a long term period (to be carried out either in sequence, parallel, or overlapping in time) in different scenarios (classroom, home, computer lab, field trips…), involving teachers and learners belonging to a school and nature monitors from an association. The educational network should embed the technologies to collect data in a site of interest, supporting its recording and reuse later on, to facilitate the articulation of physical exploration of a site of special interest, with analytical reflection in the school. In order to permit a smooth flow throughout the scenarios of activities, the technological infrastructure should grant a ubiquitous context challenging the integration issue through the school curriculum, location, time, social organization levels, across devices, interoperability, and connectivity. An important component in this educational networked infrastructure is the "Learning Object Repository" (LOR). The LOR integrates data and artefacts created from heterogeneous resources. Artefacts, in this sense, are the products produced by the learners using certain tools on a diversity of devices. The LOR provides an adequate framework for storing, retrieving and re-using individual and group results and by-products, offering group and community navigation tools as well as mechanisms to detect similarities of interests in terms of the produced objects or artefacts . These user-created artefacts are meant to be reused by students in diverse contexts, and using other tools or different devices. For example, using the LOR, users can exchange different annotated and conceptualised pictures to cover views of a tree in different seasons, to complete their measurements, or to aggregate collected data, in order to enrich the joint models of the ecosystems through different observations taken at various sites, in different periods of time, all throughout the year.

The project is organized in tasks. For each task the following description includes the current state of the subtasks and the planned chronogram. There are small adjustments in the actual timing to cope with the feedback of the formative evaluation project cycles.

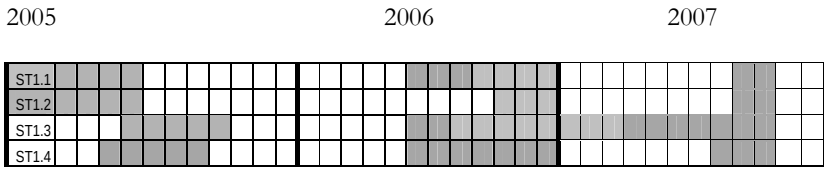
Task1. Model, design and implementation of a portal for virtual communities for learning.

Subtask 1.1 – Develop ontologies for the design of learning scenarios (as planned)

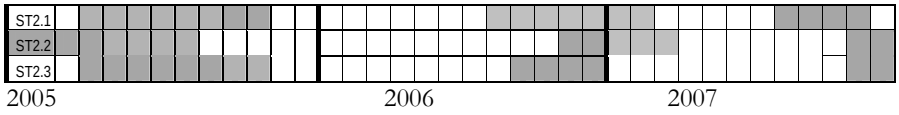
Subtask 1.2 – Define a modelling educative language for learning activities (as planned)

Subtask 1.3 – User modelling and learning trajectories (as planned)

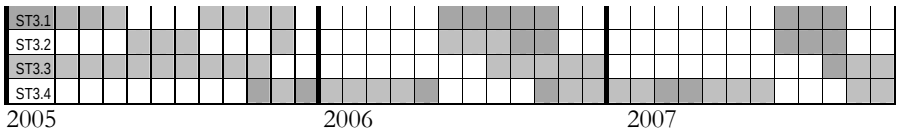
Subtask 1.4 – Adaptation and intelligent support (delayed for 2007, to be synchronized with subtask 2.1)



Task 2. Definition and implementation of the architecture: Agents and resources
Subtask 2.1- Integrate the agent technology with the architecture (in progress, with a small delay)
Subtask 2.2- Specification and development of a Distributed Repository of Objects (finished)
Subtask 2.3 – Explore the reuse and dynamic presentation of resources in different contexts (finished)



Task 3. Generation of a portal for learning communities in the scope of environmental education
Subtask 3.1- Repository with contents and its methodology (in progress as planned)
Subtask 3.2- Selection and adaptation of software tools to users and devices (in progress as planned)
Subtask 3.3- Design of contextual zed activities in real environments (finished)
Subtask 3.4- Define a virtual learning community for a pilot case (in progress as planned)



Task 4. Evaluation and Usability
Subtask 4.1- Formative evaluation (in progress as planned)
Subtask 4.2- Maintenance strategies and international dissemination (in progress as planned)

2 Project results

From a computational point of view, the following software is already operational:
1. The LOR
The LOR service (Figure 1) provides functionality directly accessible from a portal interface, and remotely available via web services. The LOR service offers facilities for managing LO, instances of the LO types that have been defined. The operations are: adding, modifying, deleting or searching objects in the repository.
A LOR is configurable with a set of parameters needed to install and configure the service. Then, additional data allows to personalize and adapt it to the environment where is going to be used. The installation data are: the name of the LOR, the address to be installed, database connection parameters. The configuration data includes information about the users, and the definition of the

types of LO. For all these data, the service offers default values that can be modified during the configuration or during the usage of the service.

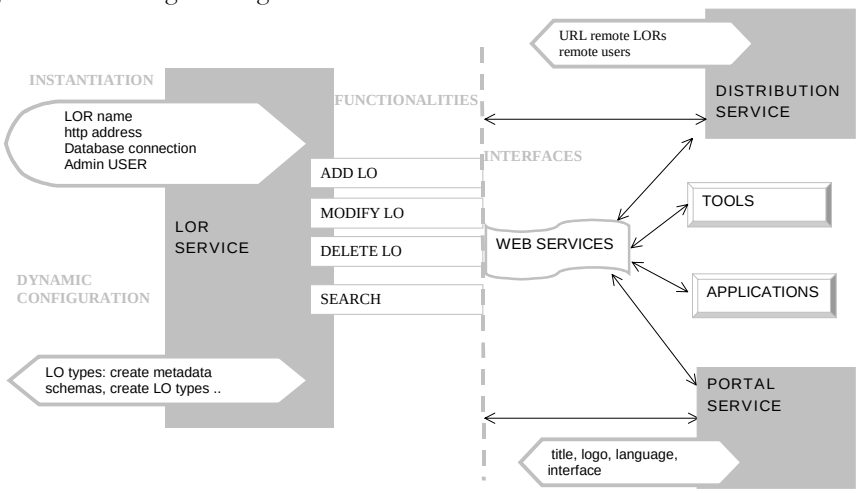


Figure 1. Description of the LOR service and other services available

This repository service, in the current state, offers (1) A persistence mechanism for community sharing and reuse of learning resources (2) Concerns to standards: LOM metadata compatibility and content packaging (3) Interoperability via web services with external tools (4) A layered expressivity: a version relying on metadata, fully automatically deployed. There is a version relying on ontologies with a wider potential for semantic searches but needing ad-hoc management of the ontological modules and the supporting software.

As well as the following featuring characteristics:

- Heterogeneity, i.e. the possibility of storing objects belonging to a variety of Learning Resources: LO types/metadata schema, tools or learning designs defined by application profiles in a declarative way.
- Mechanisms for handling automatic generation of metadata for objects created with external tools, via application profiles for tools.
- An application- a community portal offering: (a) Flexibility for community personalization (workspaces for individuals/groups/communities) (b) Contextual metadata generation for LO, extracted from activity and social models in the metadata version (these models are ontologies in the richer version)
- A service for synchronizing distributed repositories in a network
- Enhanced searching facilities based on metadata, including query by example, relevance feedback, filters and query patterns.

Searching is an important functionality offered by the LOR service. Different modalities of searching are offered based on the constraints expressed in a combination of metadata fields through templates. This functionality can be used by an external tool for finding the necessary objects the user may request to carry out a learning activity with that tool. The retrieved objects can be directly downloaded from the repository to the tool, then handling the objects inside the tool, and lastly reincorporated them to the LOR as a new LO version. In this case the function of the repository is not only to be a persistence mechanism for that tool, but a group memory offering the access, share and exchange of artefacts within a learning community. Searching can be done by

query by example (selecting an object, and using its metadata as a starting template), filters (predefined combination of templates that provide useful and recurrent queries) and query patterns (parameterised combination of templates, that the user can modify and instantiated). Filters and query patterns can be created either by a tool or when configuring the portal.
A main reason for searching is reusing. Thus the role of metadata is to facilitate users and software tools, the query and retrieval of objects, potentially useful for a specific purpose in the learning process. A challenge here is to be able to define metadata capturing the potential interest of an object to be reused. Our approach is to differentiate and enrich the metadata information because it is not the same to describe an object, which is a qualitative model, than an object, which is the result of a simulation, or an object representing experimental data, in each case the number and description of interesting features are different. Further details are available in the publications listed in section 3.

2. The design and implementation of a set of tools: Cards, PDACADSs,Agora, OXO

These tools provide generic functionality for a variety of learning activities, and offer full interoperability with the LOR. A brief description follows:

2.1 *CARDS* is a data collection tool as well as an authoring system for defining models for data collection and the metadata associated to these models. In addition, CARDS can be seen as a metamodeling system for wrapping other tools products. It allows importing and exporting data in different formats, acting as a bridge to the LOR functionalities for those standalone applications not able to generate descriptions for their input/output. Therefore, CARDS is a general purpose tool providing functionality for (1) Creating models for data collection, tests, activities,..(2) Instantiation and differentiation: Teachers, students and monitors are able to create new types of nature cards and activities that combine several of these, by just instantiating and stating parameters from previous ones(3)Aggregation: students can apply transformations and combinations of data to generate new objects. (4) Exchange of objects with the LOR, using web services. (5) Mapping of data to output formats for tools with limited interoperability.

2.2 *PDACARDS* is a stand-alone pda application. It brings solution to the problem of filling observation cards in scenarios where no internet connection is available and therefore there is no access to CARDS web application. When a user logs in, PDACARDS connects automatically with CARDS and downloads the activity, which is are made up from a set of observation cards and a map with the activity localization. From this point, students can take observations without an internet connection during the fieldtrip. The map is interpreted by the "CardMap" tool developed by the members of the U.Málaga participating in the project, which has been integrated within PDACARS. "CardMap" is used to present the map image in the pda screen and to extract the localization coordinates from the pda GPS card. As a result of this, it is possible to determine the position where an observation card was filled. Once the fieldtrip has ended, all observations are sent back to CARDS, which stores them in the LOR, along with metadata that includes the GPS information.

2.3 *Agora* is a web tool developed in the framework of the ENLACE project. It is named after the place where the magistrates and the citizens from ancient Greek polis gathered to discuss any issues, and then voted by raising hands. This tool allows the teacher to design and perform real-time vote sessions in a classroom environment in which some or all of the students are provided with a computer or a handheld; a projector is also required to guide the activity and show the results. The design task is handled by the session manager module, while the real-time performance is carried out by the web player module. There are two different roles considered: the teacher and the student, being the first one the only with access to the session manager. The

purpose of Agora is to give technological support to a wide range of possible discussing and voting scenarios that can come up in an educative environment, from the election of the delegate of the classroom to a collective answer for a test question. Furthermore, the delegate election scenario can be different whether we assume that students will previously put themselves forward for the charge or just consider every student a compulsory candidate. Our approach is to consider two main factors that change between different scenarios: the nature of the candidates and the student participation in the candidate selection. These factors can be modified through the session manager interface. The first issue has been faced by extending the concept of candidate in such a way that practically anything, including a learning object, can be a candidate.

2.4 *OXO* is another tool developed in ENLACE which provides functionality for aggregating "active" icons over a background, to create a new composition object. OXO offers the following functionality: 1- Searching and Filtering objects in the LOR to create a candidate set. 2- Associating objects with symbolic icons to create a palette. 3- Searching and importing an object from the LOR to be used as a background. 4- Selecting icons from the palette and situating them on the background. For example, students can create a symbolic map of the trip and situate the selected icons in the position where they have listened a particular bird: the object associated with the icon can be visualized 5- Finally, when students have positioned all the cards' icons in the background, they can store the composed object (background and icons) as a new object in the LOR.

3. The mechanisms for interoperating the LOR with external tools, have been successfully implemented and tested with different tools and resources such as ModellingSpace and Tree DataBase. (A Botanic database)

4. The networking functionality for the different scenarios

The design of the educational network has been driven to facilitate the flow and transformation of data and artifacts across scenarios, for all the actors involved in the learning activities. The LOR and Cards provide the support for interoperability, together with external tools and other tools developed in ENLACE provide potential for integration in a wide range of scenarios. As an example, figure 2 shows the configuration deployed for the activities described in the introduction, to illustrate our approach: The scenario A is the classroom where the teacher and the students carry out a training activity involving bird song identification, using Agora, a tool supporting presentation, discussion and voting. The scenario B is the classroom where the teacher and the students prepare the activities for the field trip, using Cards and a variety of sources of information. The scenario C is the field trip where students work mainly collecting data with the PDAs in standalone mode. However, there are two periods of connection with the LOR, at the beginning and at the end of the trip, where the cards to be filled are distributed as well as collected when they are filled. The scenario D involves students working in pairs, in the computer lab with a set of tools: Modeling Space, OXOTool, and Cards, interacting with the LOR. Finally scenario E is a group activity in the classroom, to present results and select together a representative sample for the group work. In this case, using Agora interoperating with the LOR to select candidates and store the result. The configuration system is 'intelligent' enough to be able to deploy and initialize the tools for the student's activity taking into account the context of the user and the activity –localization, current working device, tools available– and personalize it according to the situated requirements. The set of specialist agents and the models needed are the main work to be undertaken in the third year of the project.

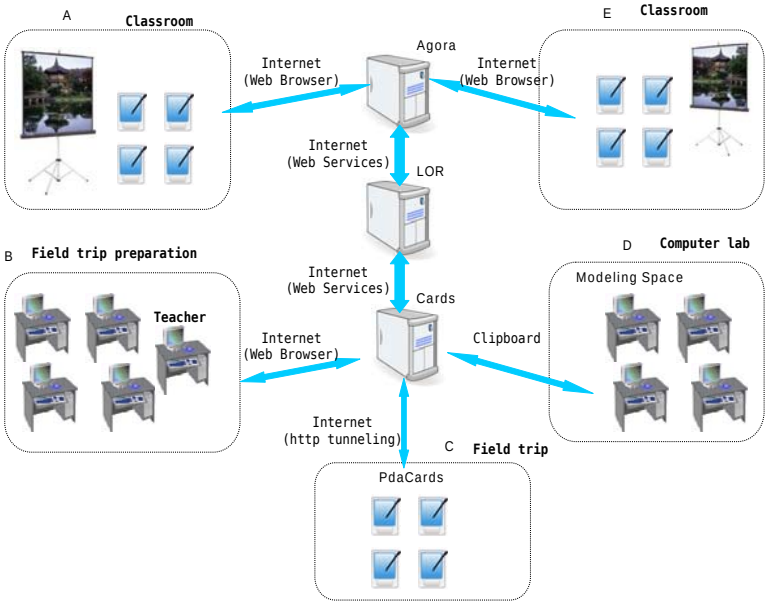


Figure 2. Technological configuration for activity deployment

From a pedagogical point of view, it is worth to mention (1) The design of a wide range of activities integrated with the school curricula, including the definition of models for data collection and knowledge co-construction, as well as a set of qualitative models for a variety of physical phenomena.(2) The production of guidelines and authoring tools for teachers and students (3) The proposal of an evaluation methodology (in progress) to follow and assess the learning processes.
These tasks are carried out with the participation of Marta Millán, a teacher that has spent a sabbatical year (2006) with us; and the 12/13 years classes of students from Instituto Diego Velazquez, that have been involved in the pilots of the formative evaluation cycle. As well as the contribution of Federico García and three volunteers monitors from SEO-Birdlife who have provided their expertise and contributed to the organization of the activities in the three nature fieldtrips carried out with students in El Monte del Pardo.
From a practical point of view, several trials have been carried out in order to have an accurate estimation of duration of PDAs batteries using different functionalities (such as wifi cards, GPS,…) in order to tune the configuration of the network to be operative for the requirements of the outdoor activities.

3 Indicators

- Five Phd students are involved in the Project, in different stages (3 have already obtained the DEA, 2 of them are working for their Phd thesis while the other two are curently working for the DEA).
- Eleven papers have been published: two articles in journals, 7 in Proceedings of relevant international events and 2 in national Conferences. The list follows:

1-Verdejo M.F., Celorrio C. (2007)"A Multi-Agent Based system for activity configuration and personalization in a pervasive learning framework". Proceedings of the 3rd IEEE International Workshop on PervasivE Learning (PerEL 2007), New York, USA. (to appear)
2-Verdejo M.F, Celorrio C., Lorenzo E.J., Ruiz A., Sastre T. (2007) "Sustaining learning activity flow in a framework for ubiquitous learning". To appear in Beyond Mobile Learning Workshop. Kaleidoscope CSCL Alpine Rendez-Vous
3-Verdejo M.F, Celorrio C., Lorenzo E. (2006) "Improving Learning Object Description Mechanisms to support an integrated framework for ubiquitous learning scenarios" . Proceedings of the 4th IEEE International Workshop on Wireless, Mobile and Ubiquitous Technologies in Education (WMUTE 2006)
4-Mayorga J.I., Barros B., Celorrio C., Verdejo M.F. "Accessing a learning object repository through a semantic layer". Proceedings of the ECDL 2006 Workshop on Learning obect repositories as digital libraries: current challenges.
5-Verdejo M.F., Celorrio C., Lorenzo E.J., Satre T (2006) "An educational networking infrastructure supporting ubiquitous learning for school students". Proceedings of the 6th IEEE International Conference on Advanced Learning Technologies. 2006
6-Vélez. J.; Barros, B.; Verdejo, M.F., (2006) "A Framework to define Web Based Communities". International Journal of Web Based Communities (IJWBC). Vol 2, N.3 Pgs 339-359
7- Celorrio C., Verdejo M.F,. Barros B. (2006). "Un Modelo de distribución de repositorios para un portal de comunidades de aprendizaje colaborativo". Revista Iberoamericana de Informática Educativa, N.3, pg 27-42. ISSN 1699-4574.
8-Vélez. J.; Mayorga, J.I.; Barros, B.; Verdejo, M.F. "A Metamodel for defining and managing Web Based Communities". Proceedings of the Conference on Web Based Communities 2005
9-Mayorga, J.I., Barros, B., Celorrio, C., Verdejo, M.F. (2005). An Ontology-driven portal for a collaborative learning community. 12th International Conference on Artificial Intelligence in Education, Frontiers in Artificial Intelligence and Applications, vol. 125, pags. 872-874, IOS Press, ISBN 1-58603-530-4
10- J.L. Mayorga, C. Celorrio, E.J. Lorenzo, J. Vélez, B. Barros, M.F. Verdejo. (2005) Una aplicación semántica de apoyo a Comunidades Virtuales de Aprendizaje Colaborativo . XI Conferencia de la Asociación Española para la Inteligencia Artificial. Actas del taller "Técnicas de la Inteligencia Artificial Aplicadas a la Educación". Santiago de Compostela 2005
11- C. Celorrio, M.F. Verdejo, B. Barros (2005). "Una Aproximación a la Distribución de Repositorios de Objetos de Aprendizaje basada en Servicios Web".VI Congreso Nacional de Informática Educativa. SINTICE'2005 (ADIE)

- Three invited presentations have been carried out in the following events
  1. Inquiry learning, building bridges to practice. Kaleidoscope workshop. 29-31 May 2006. University of Twente. Netherlands
  2. ONLINE EDUCA May2006: Kaleidoscope session (Nuevas y brillantes ideas en el ámbito de las tecnologías y del aprendizaje informal Moderadora: L Montandon, Atos Origin).
  3. 21st September 2006, Madrid. Workshop : e-COMODE: 2nd International Workshop on Collaborative Modelling Environments implementation in Schools
- Participation in the FP6 Kaleidoscope network of excellence (http://www.noe-kaleidoscope.org)
  - through the SIG groups: AI&ED, CSCL and Inquiry Learning.
  - Contribution to the Alpine Rendez-vous (http://craftwww.epfl.ch/events/alpine/) workshop 2: Beyond Mobile Learning: Innovating Rather Than Replicating Existing Learning Scenarios
- Organization of a workshop in the 2006 European Conference on Digital Libraries Workshop : Learning Object Repositories as Digital Libraries: Current challenges.(http://ltcs.uned.es/wslo2006/)
- Participation in the Spanish semantic web network. TSI2006-26928

- The system is currently in use in "Instituto Diego Velazquez" de Torrelodones

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Foundations of Automated Deduction and Satisfiability, iDEAS TIN2004-04343

María Luisa Bonet Carbonell *
Universidad Politecnica de Catalunya

Abstract

The project is about searching for new algorithms to show the satisfiability of Boolean formulas. It is also about studying which proof systems are the most efficient to show their unsatisfiability. Both directions have aplications in the fields of circuit verification, program verification with model checking, automated deduction, and others.

The satisfiability problem is a particular case of the constraint satisfaction problem. The decisional form of both problems is NP-Complete. Constraint satisfaction problems appear in a large variety of practical situations, such as scheduling, temporal reasoning, machine vision, etc., so their study is important. We will study this topic both from the point of view of verifying satisfiability, and from the point of view of showing the unsatisfiability of constraints via proof systems. For that we will need to define proof systems that will work with domains more general than 0/1. This is a new and promising approach, because it brings to the field of constraint satisfaction problems the techniques from proof complexity and finite model theory that had not been used before in this context.

Keywords: Constraint Satisfaction, Satisfiability, Automatizability, Proof Systems.

1 Project Objectives

Obj. 1 *Proving Lower Bounds for Propositional Proof Systems.*

Obj. 2 *Automated Deduction. Analisys of the Theoretical Limitations*

Obj. 3 *Study and Comparison of Different Measures of Complexity for Propositional Proof Systems*

Obj. 4 *Study of Structural Properties of SAT and CSP Instances*

Task 4.1 Try to find structural properties of boolean formulas that allow us to detect tractable cases of SAT.

*Email: bonet@lsi.upc.edu

Task 4.2 We will study the topic of phase transition of random instances of SAT. The point of phase transition is the ratio number of variables divided by number of clauses that separates the satisfiable formulas from the unsatisfiable ones. We want to study properties of this random formulas related to the point of phase transition.

Task 4.3 We will work on the tractable cases of the problem of morphism. This is part of the general objective of obtaining a complete characterization of the tractable cases of the CSP problem.

Obj. 5 *Definition of Proof Systems for CSPs and Study of the Connection with Finite Model Theory*

Task 5.1 Definition of Proof Systems for CSPs. This proof systems would allow to refute the satisfiability of constraints. Our theoretical results would give insights for future tools to refute CSPs.

Task 5.2 Study the relationship between Finite Model Theory and Constraint Satisfaction Problems (CSPs).

Tar.	Centro	Investigadores	1^o año	2^o año	3^o año
T 1.1	UPC	Atserias, Bonet, Postdoc 1			
T 1.2	UPC	Esteban, Galesi			
T 1.3	UPC	Atserias, Esteban, Galesi			
T 2.1	UPC, CSIC	Bonet, Levy, Villaret			
T 2.2	UPC, U. Toronto	Bonet, Pitassi, Postdoc 1			
T 2.3	UPC	Atserias, Bonet			
T 2.4	UPC, CSIC	Bonet, Levy, Villaret			
T 3.1	UPC	Esteban, Galesi			
T 3.2	UPC	Atserias, Esteban, Bonet			
T 4.1	UPC	Lozano, Galesi			
T 4.2	UPC, U. Toronto	Atserias, Pitassi			
T 4.3	UPF	Dalmau, Postdoc 2			
T 5.1	UPC, Rice U.	Atserias, Dalmau, Vardi			
T 5.2	UPC, UPF, Rice U.	Dalmau, Postdoc 2, Vardi			

2 Level of Success Attained

2.1 Results on Objetive 1

We obtained some results on the task 1.2 that appear in the publication [24]. In it the authors prove lower bounds for the size and the space of refutations for the propositional proof system

Res(k) tree-like. Res(k) is an extension of the Resolution proof system where we don't only work with clauses, but also we can use disjunctions of conjunctions of up to k literals. Also in the article the authors show that the proof systems Res(k) for $k \geq 1$ form a strict hierarchy. This means that we can exponentially separate Res(k) from Res(k-1), and therefore the second system has to refute some clases of formulas with exponentially more clauses than the first.

2.2 Results on Objetive 2

In the article number [5] we study different notions of automatizability for various propositional proof systems. Also we determine which of these notions are equivalent and which aren't. Furthermore we see which properties should fulfill the system of Resolution to be weakly automatizable.

2.3 Results on Objetive 3

Members of our group (among others) see [15] are writting a manuscript about a new measure of formulas, called the Strahler number. One way of defining this measure for a formula is the minimal complete binary tree that can be embeded into a refutation of the formula. This measure also corresponds with the minimal tree-like space of writing a refutation of a formula. We show how the Strahler number is a measure of how structured a formula is, and a small Strahler means that the formula is easy to refute and viceversa. Small Strahler usually means real life instance while high Strahler is related to random instances and hard combinatorial principles (crafted instances). We have developed an algorithm to obtain refutations based on our new measure that could be competitive for real life instances.

2.4 Results on Objetive 4

Paper [25] is about task T.4.1. In this work, the authors combine graph theory and linear algebra to study SAT problems of low "linear algebra complexity", i.e. formulas with bounded hermitian rank. They show polynomial time SAT algorithms for classes of formulas with hermitian rank at most one. They do this by applying methods from hypergraph transversal theory.

Task T.4.2 proposed the study of properties of random instances for distributions with fase transition. The most important result that we obtained in this topic is the article [3]. In this publication the authors show that the standart distribution on 3-CNF formulas is indistinguishable from the planted distribution respect to the properties expresable in first order logic. To obtain this result, the authors needed a detailed study of the Shelah-Spencer results. Adapting these results to the planted distribution is not at all trivial technically. On the contrary, the results are highly difficult.

In task 4.3 great progress has been made. The articles [9, 13, 18, 17, 21, 16, 19, 20, 22, 23] deal with posible tractable problems of constraint satisfaction (CSP). Some of these articles study methods like gausian elimination that have helped to detect the class of paraprimal algebras that has been shown tractable. Also, we have studied notions of consistency for quantified CSPs and CSPs of bounded arity.

2.5 Results on Objetive 5

In task T.5.1 we proposed to define proof systems for CSPs. Our proposal (article [7]) consists in defining instances of the constraint satisfaction problem as a relational database, and as a consequence to use the relational algebra operations or joint, projection and weakening to derive new restrictions. In this article, the authors show that a CSP instance is unsatisfiable if and only if one can derive the empty relation with the relational algebra operations. Next the authors propose the concept of width of a CSP refutation and they characterize it using Ehrenfeucht-Fraissé games. This article manages to connect the area of constraint satisfaction problems with the area of finite model theory, as we also wished to do in this task.

Also in the same article the authors propose a new proof system where the lines of the proof are OBDDs, and the rules are again join, projection and weakening. The articles contains proofs of the fact that this system simulates Resolution and Cutting Planes with polynomial coeficients, and also that the system admits monotone interpolation. It would be desirable that this new proof system would have practical applications the same that Resolution has. In fact in the conference SAT 2004 Pan and Vardi published an experimental study.

Recently, Jan Krajicek from the Academy of Sciences of the Czech Republic, communicated us a solution to one of the open problems proposed in the article. Cavgnetto and himself have proposed a method to build propositional unsatisfiable formulas for which, for every ordering of the variables, the minimum OBDD refutation is exponential. The result is based on [7] using an additional idea. The fact that such prestigious researcher as Krajicek has worked on open problems that our group formulated is an indication of our success.

Under the task T.5.2 several papers have been written. The first is [4]. In this work the authors work with Tseitin tautologies, well known in the area of proof complexity, to show width lower bounds on Datalog programs to solve certain CSPs. This work presents a simplification and improvement to a former result of Feder and Vardi.

The second paper we want to mention for this task is [2]. This article received the "Best Newcomer Award" of the conference ICDT 2005. In this article the authors propose the adaptation of an algorithm of Beame and Pitassi to show how to authomatize tree-like resolution to the context of CSPs. The adaptation is not as direct as it seems, and the article leaves some open problems that we know some foreign researchers are working on.

Also for the task 5.2, other results have been obtained. In [10, 1] our group has defined refutational systems (like Resolution) that our sound and complete for the optimization versions of SAT and CSP, called Max-SAT and Weighted CSP. This problems can be defined as: given a CNF formula or a constraint satisfaction problem, find an interpretation that satisfies the maximum number of clauses or restrictions. These variants of SAT and CSP are very interesting both from the point of view of theory as well as applications. There are many practical problems that require optimization techniques. At present we are in the process of implementing software based on our results. We are doing this in conjunction with the Lerida group, given that they have a cluster and we have colaborated with them on the theoretical results.

3 Indicators of Success

First we want to mention that a Ph.D. thesis has been written during this period. This is the one of Mateu Villaret, directed by Jordi Levy (both in the project). Their joint work has given

rise to many publications as can be seen in the next section.

Our results have had a big impact in the community. In the prevous text we have already given some indication of that, showing how our work opened new directions and how other people have tried or answered questions that we proposed in our work. Also some articles of us like [1], have been invited to be published as book chapters. So an extension of [1] will be published in the book "Current Trends in Constraint Programming". Also, the paper [2] obtained an award for the best paper written by somebody of outside the field of the conference.

But the main indication of the quality of our work are our publications. Most of them appear in first rate conferences and journals. We have published in confernces like LICS, IJCAI, SAT, CP, RTA, etc., and in journals like SIAM journal on Computing, Artificial Intelligence, Journal of Computational Biology, Information and Computation, Journal of Combinatorics, etc.

Below, in the references section, we give the list of publications obtained by members of our group in the last 2 years. We have kept in the document 4 papers that belong to the year 2004 because after writting the proposal of this project we started working on it right away.

Notice that we have obtained 31 publications for a project of 6 EDPs.

References

[1] C. Ansótegui, M. L. Bonet, J. Levy, and F. Manyà. The logic behind weighted csp. In *Proc. of the Twentieth International Joint Conference on Artificial Intelligence, IJCAI-07*, pages 32–37. AAAI Press, 2007.

[2] A. Atserias. Conjunctive query evaluation by search-tree revisited. In *10th International Conference on Database Theory (ICDT)*, volume 3363 of *Lecture Notes in Computer Science*, pages 53–67. Springer-Verlag, 2005.

[3] A. Atserias. Definability on a random 3-cnf formula. In *20th IEEE Symposium on Logic in Computer Science (LICS)*, pages 106–115, 2005. Long version accepted for publication in European Journal of Combinatorics.

[4] A. Atserias. On digraph coloring problems and treewidth duality. In *20th IEEE Symposium on Logic in Computer Science (LICS)*, pages 106–115, 2005. Long version accepted for publication in European Journal of Combinatorics.

[5] A. Atserias and M. L. Bonet. On the automatizability of resolution and related propositional proof systems. *Information and Computation*, 189(2):182–201, 2004.

[6] A. Atserias, A. Bulatov, and V. Dalmau. On the power of k-consistency. in preparation, work announced at Dagstuhl workshop on Complexity of Constraints, 2006.

[7] A. Atserias, P. G. Kolaitis, and M. Vardi. Constraint propagation as a proof system. In *10th International Conference on Principles and Practice of Constraint Programming*, volume 3258 of *Lecture Notes in Computer Science*, pages 77–91. Springer-Verlag, 2004.

[8] Balcazar, Bifet, and A. Lozano. Intersection algorithms and a closure operator on unordered trees. Workshop on Mining and Learning with Graphs, MLG 2006.

[9] M. Bodirsky and V. Dalmau. Datalog and constraint satisfaction with infinite templates. In *STACS*, pages 646–659, 2006.

[10] M. L. Bonet, J. Levy, and F. Manyà. A complete calculus for max-sat. In *Proc. of the 9th International Conference on Theory and Applications of Satisfiability Testing, SAT'06*, volume 4121 of *Lecture Notes in Computer Science*, pages 240–251. Springer-Verlag, 2006.

[11] M. L. Bonet, J. Levy, and F. Manyà. Resolution for max-sat. *Journal of Artificial Intelligence*, to appear.

[12] M. L. Bonet, K. StJohn, R. Mahindru, and N. Amenta. Approximating subtree distances between phylogenies. *Journal of Computational Biology*, 13(8), 2006.

[13] A. A. Bulatov and V. Dalmau. A simple algorithm for mal'tsev constraints. *SIAM J. Comput.*, 36(1):16–27, 2006.

[14] J. Buresh-Oppenheim, N. Galesi, S. Hoory, A. Magen, and T. Pitassi. Rank bounds and integrality gaps for cutting planes procedures. *Theory of Computing*, 2:65–90, 2006.

[15] J. L. Carlos Ansotegui, Maria Luisa Bonet and F. Manya. Measure of how structured a formula is. Manuscript.

[16] H. Chen and V. Dalmau. Beyond hypertree width: Decomposition methods without decompositions. In *CP*, pages 167–181, 2005.

[17] H. Chen and V. Dalmau. From pebble games to tractability: An ambidextrous consistency algorithm for quantified constraint satisfaction. In *CSL*, pages 232–247, 2005.

[18] V. Dalmau. Generalized majority-minority operations are tractable. In *LICS*, pages 438–447, 2005.

[19] V. Dalmau. Linear datalog and bounded path duality of relational structures. *Logical Methods in Computer Science*, 1(1), 2005.

[20] V. Dalmau. A new tractable class of constraint satisfaction problems. *Ann. Math. Artif. Intell.*, 44(1-2):61–85, 2005.

[21] V. Dalmau, R. Gavaldà, P. Tesson, and D. Thérien. Tractable clones of polynomials over semigroups. In *CP*, pages 196–210, 2005.

[22] V. Dalmau and A. Krokhin. Majority constraints have bounded pathwidth duality. *European Journal on Combinatorics*, to appear.

[23] V. Dalmau, A. Krokhin, and B. Larose. Retractions onto series parallel posets. *Discrete Mathematics*, to appear.

[24] J. L. Esteban, N. Galesi, and J. Messner. On the complexity of resolution with bounded conjunctions. *Theoretical Computer Science*, 321(2-3):347–370, 2004.

[25] N. Galesi and O. Kullmann. Polynomial time sat decision, hypergraph transversals and the hermitian rank. In *SAT Selected Papers*, pages 89–104, 2004.

[26] N. Galesi and N. Thapen. Resolution and pebbling games. In *SAT*, LNCS, pages 76–90, 2005.

[27] J. Levy, J. Niehren, and M. Villaret. Well-nested context unification. In *Proc. of the 20th Int. Conf. on Automated Deduction, CADE-20*, volume 3632 of *Lecture Notes in Artificial Intelligence*, pages 149–163. Springer-Verlag, 2005.

[28] J. Levy, M. Schmidt-Schauss, and M. Villaret. Bounded second-order unification is np-complete. In *Proc. of the 17th International Conference on Rewriting Techniques and Applications, RTA'06*, volume 4098 of *Lecture Notes in Computer Science*, pages 400–414. Springer-Verlag, 2006.

[29] J. Levy, M. Schmidt-Schauss, and M. Villaret. Stratified context unification is np-complete. In *Proc. of the 3rd International Joint Conference on Automated Reasoning, IJCAR'06*, volume 4130 of *Lecture Notes in Computer Science*, pages 82–96. Springer-Verlag, 2006.

[30] J. Niehren and M. Villaret. Describing lambda-terms in context unification. In *Proceedings of the 5th. Int Conf. on Logical Aspects of Computational Linguistics, LACL'05*, volume 3492 of *Lecture Notes in Artificial Intelligence*, pages 221–237, Bordeaux, France, 2005. Springer-Verlag.

[31] M. Villaret. *On Some Variants of Second-Order Unification*. PhD thesis, UPC, 2005.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Mechanisms of Multisensory and temporal Integration in Cognition and Intelligent Systems
TIN2004-04363-C03-01

Gustavo Deco*
Universitat Pompeu Fabra
Barcelona

Salavador Soto**
Universitat de Barcelona
Barcelona

Francisco de Borja Rodríguez***
Universidad Autónoma de Madrid
Madrid

Abstract

Developing paradigms to advance in the design of new interfaces leads current researchers to look at the principles of information processing evolved by living organisms in their interaction with the environment. One of the most critical and impressing capabilities of intelligent organisms is their ability to manage, in an adaptive way, the rich and continuous flow of information present in the environment. In order to achieve this, the brain must incorporate mechanisms able to synthesize all the fragments of information throughout time and across the different sensory modalities (vision, audition, haptics,...). The aim of this coordinated project is to study the mechanisms underlying such cross-modal and cross-temporal integration at different levels of analysis, in order to provide a more robust and unified account of such processes. The different levels of analysis incorporated throughout the three subprojects extend from the level of synaptic and single-cell activity, to the patterns of activity in whole brain areas at the cortical level, and to the manifestations of these integration mechanisms in behaviour and cognition. Our general goals are: (i) to provide novel and deeper empirical knowledge on multisensory integration processes in perception (using speech and motion perception as our case studies); (ii) to provide new and deeper knowledge on temporal integration (within the framework of the delayed response tasks); (iii) to obtain empirical data on temporal and multisensory integration by means of functional Magnetinc Resonance Imaging (fMRI) and modelling; and (iv) to contrast formal models provided by a unifying theoretical neurodynamic framework with the empirical data. The multidisciplinary nature of this project is critical in order to obtain an adequate account of the processes underlying multisensory and temporal integration by helping to constrain the explanations given at one level of analysis with the knowledge generated by the other levels.

Keywords: Cognitive Neuroscience, Computacional Neuroscience, Decisión-Making.

* Emaill : gustavo.deco@upf.edu
** Email: ssoto@ub.edu
*** Email: F.Rodriguez@uam.es

1 Aims of the Project

In recent years a considerable amount of knowledge has been gathered concerning how the human brain operates and its relationship with different cognitive functions. This enterprise, although far from being accomplished, has unveiled the underlying organizational structure of the cortical system and different cognitive functions. Understanding the fundamental principles underlying higher brain functions requires the integration of different levels of experimental investigation in cognitive neuroscience (from the operation of single neurons and neuroanatomy, neurophysiology, neuroimaging and neuropsychology to behaviour) via a unifying theoretical framework that captures the neural dynamics inherent in the computation of cognitive processes. The main aim of the current project is to elucidate the neural, synaptic, and cortical mechanisms underlying multisensor and temporal integration in cognition by linking the dynamics at the cellular (neuronal) , large-scale network, and behavioural level. To achieve this aim, we investigated and integrated the different experimental and theoretical neuroscience levels in three different coordinated subprojects:

I.: Neurodynamical Mechanisms Underlying Temporal and Multisensorial Integration (Universitat Pompeu Fabra, UPF, IP: Gustavo Deco)

The level of description in a theoretical framework should be accurate enough to allow the relevant neuronal and synaptic mechanisms to be properly taken into account. However, the description should also be simple enough, so that we can readily infer by abstraction the relevant principles substantiating perception and cognition. A mere reproduction of phenomena of a complex system, like the brain, just by simulating the same kind of artificial complex system, is most of the time not useful, because there are usually no explicit underlying first principles. The aim of this subproject is to develop a unifying theoretical framework that captures the neural dynamics inherent in the computation of cognitive processes. In particular, we focus on the processing underlying temporal and multisensorial integration of information. During the execution of this project, we actively contributed to this enterprise by developing a theoretical framework that fulfils these requirements. This theoretical framework was obtained by developing explicit mathematical neurodynamical models of brain function based on the level of neuronal spiking and synaptic activity. Analysis of networks of neurons each implemented at the integrate-and-fire neuronal level (including non-linearities) enables the study of many aspects of brain function, from the spiking activity of single neurons and the effects of pharmacological agents on synaptic currents, through fMRI and neuropsychological findings, to be integrated, and many aspects of cognitive function to be modelled and understood. In particular, we used this computational neuroscience approach to produce a unified theory of attention and working memory, multisensor integration, and how these processes are influenced by rewards to influence decision making.

II.: Behavioral Measures of Temporal and Multisensorial Integration
(Universitat de Barcelona, UB, IP: Salvador Soto)

Humans have an impressive array of sensory systems, each encoding a specific type of information about the environment. Yet, our perception of the world is not fragmented into separate sensory experiences, but rather is a unitary one. The brain integrates information across sensory modalities in order to obtain accurate and coherent representations of the external environment that can be used to guide adaptive behaviour. The aim of this subproject is to address the behavioural/psychophysical correlates of multisensory integration in three particular domains of interest. (a) Perceiving speech; (b) Perceiving motion of objects; (c) When deploying selective attention. These lines of research are, in turn, integrated in the main objective of the present Coordinated Project, namely the description of the cerebral mechanisms responsible for integrating information in time asnd acros sensory modalities at different (complementary) levels of analysis. In order to achieve this main objective, we have also addressed a fourth perceptual integration domain from a combined computational (UPF) and psychophysical (UB) perspectives. This combined approach should be regarded as the model for future integrated work in other domains of cognition and perception for further work in the project as well as for future coordinated projects between our groups; (d) Decision making in a tactile discrimination task

The methodological approach is psychophysical (experimental psychology) and neuropsyhological (tests on neurological patients).

III.: Spatio-temporal Codification in Biological Rhythms
(Universidad Autónoma de Madrid, IP: Francisco B. Rodríguez)

Sensory perception is coded by sequences of spikes in the nervous system forming a part of the neural code. There are two main parameters to analyse the neural code: the neuronal identity of spikes and the timing of spikes. From this point of view we can discuss spatial, temporal and spatio-temporal code. The main goal of this project is the study and analysis of several mechanisms involved in the neural coding processes and their implications in multisensory integration and sensory motor transformation in neural systems. The methodology used for this study is the neuronal modelling at different scales of descriptions using the nonlinear dynamics formalism, information theory and graph theory.

2 Scientific Results

We were able to enhance our knowledge on the neuronal and cortical mechanisms underlying multisensor and temporal integration of information in cognition. We describe here briefly the main achievements:

I.: Neurodynamical Mechanisms Underlying Temporal and Multisensorial Integration (UPF)

We placed a particular emphasis on the following question: How are different partial representations, held in cortical areas, integrated to form a coherent stream of perception, cognition, and action?. Instead of a central coordinating brain structure, a massive recurrent connectivity between cortical brain areas is found, which forms the white matter that occupies the largest fraction of the brain volume. This suggests that integration of partial representations held in different cortical areas might be integrated by mutual cross talk, mediated by inter-areal neural fibers. Based on this view and further neurophysiological evidence, it has been hypothesized that each cortical area is capable of representing a set of alternative hypotheses encoded in the activities of alternative cell assemblies. Representations of different conflicting hypotheses inside each area compete with each other for activity and for being represented. However, each area represents only a part of the environment or internal state. In order to arrive at a coherent global representation, different cortical areas bias each others' internal representations by communicating, their current states to other areas, thereby favoring certain sets of local hypotheses over others. For example, the representation of conflicting hypotheses in one area might be resolved by a bias given towards one of the two hypotheses from another area, as obtained from this other area's local view. By recurrently biasing each other's competitive internal dynamics, the global neocortical system dynamically arrives at a global representation in which each area's state is maximally consistent with those of the other areas. This view has been referred to as the biased competition hypothesis. In addition to this competition-centered view, we have formulated a cooperation-centered picture of brain operation, where global representations find their neural correlate in assemblies of co-activated neurons. Co-activation is achieved by stronger than average mutual connections between the members of each assembly. Reverberatory communication between the members of the assembly then leads to persistent activation of the member neurons and gives rise to a representation extended in time. In summary, the theoretical and computational framework of biased competition and cooperation offers a unifying principle for neurocognitive modeling of higher neocortical functions (see Deco and Rolls, 2005c).

Cognitive behaviour requires complex context-dependent processing of information that emerges from the links between attentional perceptual processes, working memory and reward-based evaluation of the performed actions. Indeed, cognitive flexibility, requires associative cortical areas that mediate the coordination between working memory, attention, expected rewards, and the alteration of behaviour if their reinforcers are not obtained as expected. Our computational neuroscience theoretical framework was able to show how an attentional state held in short term memory in the prefrontal cortex can by top-down processing influence ventral and dorsal stream cortical areas using biased competition to account for many aspects of visual attention. Attention then appears as an emergent effect related to the dynamical evolution of the whole network. This formulation incorporates spiking and synaptic dynamics which enable to simulate and explain in a unifying framework visual attention in a variety of tasks and at different cognitive neuroscience experimental measurement levels (see Deco and Rolls, 2005a, Deco and Rolls, 2005b). Furthermore, we also extended the analysis to the case of perceptual learning, in which attentional mechanisms are not only driven by external biasing top-down signals associated by the definition of the task, but are also hardwired in the synaptic structure, after learning (based on a biologically

realistic Hebbian plasticity rule). In other words, by experience, the system learns the features dimensions that are relevant for the optimal solution of the task, i.e. for behaving properly. This hardwiring emerges from a "learn-to-attend" process (Szabo, Deco, Fusi, Del Giudice, Mattia, and Stetter, 2006).

In the context of working memory and reward processing, we were further able to model in a unifying form attentional and memory effects in the prefrontal cortex, integrating single-cell and fMRI data, and different paradigms in the framework of biased competition. In particular, we were able to succesfully implement a model of the neurodynamics underlying: 1) Visual Working memory, in the context of the „What" and „Where" paradigm and the Context-Dependent Delay Match to Sample task (including the model of single-cells, of the global behaviour of the cortical network as evidenced by fMRI, and the effects of pharmacological manipulations, that are relevant for clinical diagnosis) (Loh and Deco, 2005, Deco and Rolls, 2005c, 2005d); 2) Attentional Filtering, which represents a particularly strong attentional effect, in which the context gates sensory input in a all-or-none fashion. Moreover, attentional filtering might be part of a neural correlate of inattentional blindness, which is the unability of humans to recover any information from unattended sensory stimuli. Inattentional blindness is thought to be part of an important cognitive mechanism, namely that of focusing or "concentrating" on a task to be performed (Szabo, Almeida, Deco and Stetter, 2005); 3) Selective Working Memory, which allows the formation of working memory, where task-relevant information is maintained in mind over a delay period while task-irrelevant information is filtered out through an attentional mechanism (Almeida, Deco and Stetter, 2004); 4) Task-Switching in the Wisconsin-Paradigm, we also studied a computational model of working memory load during a task-switching paradigm (Wisconsin card sorting). Earlier models involving semi-realistic neural networks focussed on the existence of 'rule coding clusters' and on perseverative subject errors. The current model improves the biophysical realism in order to generate predictions for metabolic load (functional MRI) and for lesion effects. Comparison with the actual fMRI data showed an impressive, almost quantitative agreement with experimental data (Stemme, Deco, Busch, and Schneider, 2005). The models also directly address how bottom-up and top-down processes interact in visual cognition, and show how some apparently serial processes reflect the operation of interacting parallel distributed systems. It is also possible to show how within the prefrontal cortex attentional bias can influence the mapping of sensory inputs to motor outputs, and thus play an important role in decision making. Furthermore, the absence of expected rewards can switch the attentional bias signal, and thus rapidly and flexibly alter cognitive performance (Deco and Rolls, 2005c).

We extended also the framework for the consideration of nonstationary transient phenomena. In particular, we studied a neurodynamical model of cross-modal and cross-temporal associations. We have shown that a network of integrate-and-fire neurons can generate spiking activity with realistic dynamics during the delay period of a paired associates task. In particular, the activity of the model resembles reported data from single-cell recordings in the prefrontal cortex integrating multisensor ial and temporal integration (Deco, Ledberg, Almeida, and Fuster, 2005).

The temporal integration of information can be thoroughsly investigated in the context of decision-making. In many situations, the behaviour of animals and humans involves deciding between possible actions according to the sensory or contextual information available. Recently, neuroscientists have begun to study the neurophysiological correlates of decision-making (see for example: Platt & Glimcher, 1999; Roitman and Shadlen 2002; Romo et al. 2002, 2004, Huk and

Shadlen, 2005). In these studies it is common to use experimental paradigms where a subject has to make a decision between two possible options. Accordingly, in this work we have, in a first approach, focused on constructing a model for binary decision-making (Wang, 2002). The model consists of a set of interconnected excitatory and inhibitory neurons. The excitatory neurons are structured into populations. Two excitatory populations receive inputs that carry information about each of the two possible precepts or contexts. Then, each possible decision corresponds to high activity of one of the populations. Such a network is said to show a competitive behaviour, in the sense that the activities of the two excitatory populations inhibit each other. Another important property of the network used is its stochasticity. The inputs each neuron receives are stochastic, what causes the network to have a stochastic behaviour. This property means, that given particular information, the model will neither always reach the same decision, nor always reach a decision after the same amount of time. This property enables our model to account for reaction time distributions and error rates observed in decision-making experiments. We have developed the decision-making model in association with two experimental paradigms used to investigate the neuronal correlates of decision-making. We have constrained our model using a simplified version of one of the experimental paradigms employed in Platt and Glimcher (1999) to study the neural correlates of decision-making in parietal cortex of macaques. The model, that we developed, accounts for the experimental findings observed by Platt and Glimcher and hence might start shedding light on how basic reward-biased decision making can be implemented (see Martí et al. 2006). The model proposed for binary decisions was also analysed in the context of the comparison between vibrotactile stimuli applied to the fingertips of macaques with different frequencies. This experimental paradigm was used in Romo et al. (2004) to study the neuronal correlates of a perceptual decision in the VPC. The experimental task consists of the applications of a first stimulus with a certain frequency f1, followed by a delay period where the subject must keep f1 in mind. After this delay period a second stimulus with frequency f2 is applied, after which the subject as to respond, by pushing one of two buttons, indicating whether $f1>f2$ or $f1<f2$. Romo and co-workers found that some neurons in VPC responded strongly only during the comparison period of the task, and their activity correlated with their choice between the possibilities $f1>f2$ and $f1<f2$. The activities of these neurons are hypothesized to reflect the decision-making process leading to the monkey final choice of pushing the button corresponding to one of the possible options. Hence, these are the neurons whose behaviour we try to explain with our model. In particular, we designed a neurodynamical model to account for the experimental neurophysiological data of Romo et al. (2004) that can account for the most relevant characteristics of the decision-related neural activity (Deco and Rolls, 2006). The behavior of the model predicts that the Weber's law will underlie perceptual discrimination. This is not a trivial prediction, given that contrary to most other demonstrations, the Weber's law is predicted in a domain different from stimulus intensity, namely frequency. In close collaboration, with the subproject II. (see next subsection) we performed behavioral tests of vibrotatile discrimination in humans and propose a computational explanation of perceptual discrimination which accounts naturally for the emergence of Weber's law. The human behavioral data complements the current neurophysiological evidence, constraining even more the underlying neurodynamics and computational principles involved in perceptual discrimination.

In summary, our theoretical framework enables the study of mechanisms of temporal and multisensorial integration of information involved in brain function, such as attention, memory,

reward processing, and decision-making. We believe that this kind of analysis is fundamental for a deep understanding of how the brain performs complex tasks.

II.: Behavioral Measures of Temporal and Multisensorial Integration (UB)

The level of success, which is assessed against the proportion of the work plan achieved so far in this subproject, and the publications generated, can be qualified as excellent. The following list describes briefly the aspects of the work plan that have been achieved so far. The ones marked with an asterisk (*) have resulted in publications (as many as asterisks) which are already published or accepted for publication. The ones marked with a plus sign (+) have been submitted for publication waiting for editorial decision, or in the process of writing up. Finally, the ones marked with a hash sign (#) are works in progress where the data collection has already began. The parts of the original work plan that are not listed below are either in the design stage, or else have suffered some type of deviation or delay.

Experimental series 1.1. SUPER-ADDITIVITY IN MULTISENSORY INTEGRATION

Study 1.1.3 Garner interference with syllables. (* *)

Experimental series 1.2. AUTOMATICITY IN MULTISENSORY INTEGRATION OF SPEECH

Study 1.2.1. Garner interference task. (*)

Study 1.2.2. Attentional load experiments. (* +)

Study 1.2.3. Mismatch Negativity. The electrophysiological recordings have been completed and the results are currently being written. It is expected to submit the paper within next moth (+).

Experimental Series 1.3. MULTISENSORY CONTRIBUTIONS TO THE PERCEPTION OF MOTION

Study 1.3.1. The manuscript corresponding to the Exploration of bi-stable audio-visual combinations has been recently send for publication to an international journal (Proc. of the Roy. Soc. Section B) (+).

Study 1.3.2. There line of work corresponding to the exploration of audio-visual, audio-tactile, and visuo-tactile combinations has generated several sub-projects of its own. This has been caused by thee number of different experiments and questions that has generated. Some of these sub-projects have been now finalised, others are in progress, and yet others will probably be the object of future work beyond the scope of this particular project. (* * #).

Experimental series 1.4. IS THE MULTISENBSORY INTEGRATION OF MOTION AN AUTOMATIC PROCESS?

Study 1.4.1. The role of response biases. (* *).

Study 1.4.2. The role of attention. (* *)

Experimental series 1.5. STUDYING MULTISENSORY INTEGRATION IN HEMINEGLECT AND EXTINCTION PATIENTS.

Study 1.5.3. Additional studies with patients

(a) We have studied spatial attentional deficits using tasks of temporal order judgments (*)

(b) We have also studied attentional orienting in space through the eyegaze of others (+)

It is important to highlight that, in general, most of the goals planned at the beginning of the project have been achieved or are in the process of being achieved. There are some aspects, however, that we have needed to delay or even to re-consider because of technical difficulties of even because of temporal constraints (as the timings may have been underestimated). This is the case, for example, of goal 1.1.1. (gating task). In the case of goal 1.1.2. (shadowing task), there has been an initial pilot testing but the project will be pursued in the future.

Collaborative Experimental Series

It is important to stress that this project has involved the collaboration between research approaches of the different component groups. One example of that is the research line on Perceptual Discrimination of Vibrotactile Frequencies, which Prof. Gustavo Deco's group and ourselves are carrying out currently. This project has now generated one manuscript which is being readied for its submission to a scientific journal. This research has allowed us to address a particular cognitive process (decision making) using a behavioral and a biologically plausible computational model. The results are in fact quite relevant because the behavioural data conforms to a non-trivial prediction if the computational model, namely that decision making processes will follow the weber's law (usually applied to sensory discrimination tasks using intensities.)

III.: Spatio-temporal Codification in Biological Rhythms (UAM)

The following items describe briefly the goals of the work plan that have been accomplished.

Spatio-temporal sequences in functional clusters and its implications in memory mechanisms:

The causality of events captured by the living brain, together with the storage and recall of sequential memory drive many researches to address the issue of temporal neuronal coding. We have analysed an optimal neural system that maps stimuli into unique sequences of activations of fundamental atoms or functional clusters (FCs). In this framework, we studied the capacity achievable by the system, minimal completion time and complexity in terms of the number of parallel sequences. We showed that the maximum capacity of the system is achieved without using parallel sequences at the expense of long completion times. However, when the capacity value is fixed, the largest possible number of parallel sequences is optimal because it requires short completion times. The complexity measure adds two important points: (i) the largest complexity of

the system is achieved without parallel sequences, and (ii) the capacity estimation is a good estimation of the complexity of the system (Rodriguez and Huerta, 2004).

Other approach is to study how the information capacity in an attractor neural network depends of the network topology. A wide range of networks, including those with small-world topology, can be modelled by the connectivity ratio and randomness of the links. Both learning and attractor abilities of a neural network can be measured by the mutual information (MI) as a function of the load and the overlap between patterns and retrieval states. In this work, we use MI to search for the optimal topology with regard to the storage and attractor properties of the network in an Amari-Hopfield model. We found that while an optimal storage implies an extremely diluted topology, a large basin of attraction leads to moderate levels of connectivity. This optimal topology is related to the clustering and path length of the network. We have also built a diagram for the dynamical phases with random or local initial overlap and show that very diluted networks lose their attractor ability (Domingez et al. 2004, 2006, 2007, Kostadin et al. 2004).

Spatio-temporal patterns with sub-threshold dynamic: control and motor coordination:

A motor action often involves the coordination of several motor synergies and requires flexible adjustment of the ongoing execution based on feedback multisensory signals. To elucidate the neural mechanisms underlying the construction and selection of motor synergies, we study a well known system: prey-capture in anurans (Corbacho et al. 2005a, 2005b).

Other aspect which is analyzed is the modelling of spiking-bursting neurons in assemblies (Aguirre et al. 2005a, 2005b, 2006). The spiking-bursting mode is used to generate the code in motor coordination. Numerical studies of the collective behaviour of ensembles rely on neuron models that describe the neuron behaviour on the basis of differential equations. These models require a high computational effort to reproduce the neuronal membrane potential behaviour. Well known conductance models based on differential equations are the Hodking-Huxley (HH) model or the Hindmarsh-Rose (HR) model. Such models contain variables of multiple time scales ranging from one millisecond to hundreds of milliseconds to describe slow intrinsic process and therefore require a number of floating point operations that range from 1200 operations in the case of the HH model to 70 operations in the HR model to simulate a single neuron in the network for 1 ms. Recently, some models have been proposed to solve the drawbacks of the differential models. These new models are built over a phenomenological basis and are implemented by the use of iterative two--dimensional maps that can present similar neuro-computational properties that the differential models. These models present a low computational effort that makes possible the simulation of big ensembles of coupled neurons during relatively long periods of time. We have designed a two--dimensional continuous piecewise linear map that models spiking-bursting neural behaviour. The map is constructed under phenomenological assumptions and mimics the dynamics of oscillations observed in real neurons. The behaviour of coupled neurons is also investigated, in particular the degree of synchronization of a pair of coupled maps is studied for different coupling strengths.

Mechanisms of identification from signal origin, neural signatures:
Recent experiments have revealed the existence of neural signatures in the activity of individual cells of the pyloric central pattern generator (CPG) of crustacean. The neural signatures consist of cell-specific spike timings in the bursting activity of the neurons. The role of these intraburst neural fingerprints is still unclear. It has been reported previously that some muscles can reflect small changes in the spike timings of the neurons that innervate them. However, it is unclear to what extent neural signatures contribute to the command message that the muscles receive from the motoneurons. It is also unknown whether the signatures have any functional meaning for the neurons that belong to the same CPG or to other interconnected CPGs. We use realistic neural models to study the ability of single cells and small circuits to recognize individual neural signatures. We show that model cells and circuits can respond distinctly to the incoming neural fingerprints in addition to the properties of the slow depolarizing waves. Our results suggest that neural signatures can be a general mechanism of spiking–bursting cells to implement multicoding (Tristan et al. 2004, Latorre et al. 2004, 2006, 2007, Campos et al. 2007).

Other research areas (Image Super-Resolution and Dynamic-clamp):
In addiction there have been efforts in other research areas. Among then we have developed super-resolution algorithms based on neuronal computation and artificial computation. First, an optimal distance-based weighted interpolation of the image sequence is performed using a new neural architecture, hybrid of a multi-layer perceptron and a probabilistic neural network, trained on synthetic image data. Secondly, a linear filter is applied with coefficients learned to restore residual interpolation artifacts in addition to low-resolution blurring, providing noticeable improvements over lens-detector Wiener restorations. Our method has been evaluated on real visible and IR sequences with widely different contents, providing significantly better results that a two-step method with high computational requirements. Results were similar or better than those of a maximum-a-posteriori estimator, with a reduction in processing time by a factor of almost 300. This paves the way to high-quality, quasi-real time applications of super-resolution techniques (Miravet and Rodriguez 2005, 2006, Kostadin et al. 2005).

Other line of research is the developing of a new software in Real Time (ADDCLAMP) to use in advanced neurophysiology experiments. This software will allow study the spatio-temporal codification in the biological rhythms. Traditional techniques to stimulate neurons in Neuroscience include current injection using several protocols. In most cases, although neurons are able to react to any stimulus in the physiological range, it is difficult to assess to what extent the response is a natural output to the processing of the input or just an awkward reaction to a foreign signal. In experiments that try to study the precise temporal relationships between the stimulus and the output pattern, it is crucial to use realistic stimulation protocols. Dynamic-clamp is a recent method in electrophysiology to mimic the presence of ionic or synaptic conductances in a cell membrane through the injection of a controlled current waveform. We have developed a set of advanced dynamic-clamp protocols (ADDCLAMP) for realistic stimulation of cells that allow from the addition of single and multiple ionic or synaptic conductances, to the reconfiguration of circuits and bidirectional communication of living cells with model neurons including plasticity mechanisms (Muñiz et al. 2005a, 2005b, Campos et al. 2005).

3 Outcomes

I.: Neurodynamical Mechanisms Underlying Temporal & Multisensorial Integration (UPF)

Publications:
19 Papers published (+9 submitted or in press) in international (ISI) journals.

Attention, and Memory
R. Almeida, G. Deco, and M. Stetter, European Journal of Neuroscience, 20, 2789-2803, 2004.
G. Deco, E. Rolls, and B. Horwitz, Journal of Cognitive Neuroscience, 16, 683-701, 2004.
G. Deco and E. Rolls, Journal of Neurophysiology, 94, 295-313, 2005a.
G. Deco and E. Rolls, Progress in Neurobiology, 76, 236-256, 2005b.
G. Deco and J. Zihl, Visual Cognition, 14, 1006-1024, 2006.
M. Szabo, R. Almeida, G. Deco and M. Stetter, European Journal of Neuroscience, 19, 1969-1977, 2004.
M. Szabo, G. Deco, S. Fusi, P. Del Giudice, M. Mattia, and M. Stetter, Biological Cybernetics, 94, 351-365, 2006.

Cross-Modal, Cross-Temporal Integration
G. Deco, A. Ledberg, R. Almeida, and J. Fuster, Experimental Brain Research, 166, 325-336, 2005.
G. Deco and E. Rolls, Cerebral Cortex, 15, 15-30, 2005c.
G. Deco, E. Rolls, Journal of Cognitive Neuroscience, 17, 294-307, 2005d.
E.Mavritsaki, D. Heinke, G. Humphreys and G. Deco, Journal of Physiology, 100, 110-124, 2006.
E. Rolls and G. Deco, Neural Networks, 19, 1383-1394, 2006.
A. Stemme, G. Deco, A. Busch, and W. Schneider, NeuroImage, 26, 454-470, 2005.

Decision-Making
M. Loh and G. Deco, European Journal of Neuroscience, 22, 2927-2936, 2005.
D. Martí, G.Deco, P. Del Giudice, and M. Mattia, Neurocomputing, 69, 1175-1178, 2006.
G. Deco and E. Rolls, European Journal of Neuroscience, 24, 901-916, 2006.

Others
R. Andrzejak, A. Ledberg, and G. Deco, New Journal of Physics, 8, 6 , 2006.
F. Koepke, M. Loh, A. Costa, and G. Deco, Neurocomputing, 69, 1317-1321, 2006.
G. Deco, Pharmacopsychiatry, 39 Suppl 1, S65-S67, 2006.

Phd and Master Theses
1 PhD Student has defended his project successfully in the context of these research lines.
- Anja Stemme
3 Master Theses have been defended successfully in the context of these research lines

International projects
We have obtained Funding from two European Community grants:
- European Project „EmCap“
  „Emergent Cognition Through Active Perception“
  FP6-2002-IST, 013123-2
- European Project “DiM”
  „Neural Decision in Motion“
  FP6-IST-4

We have obtained Funding from the prestigious German Volkswagen-Foundation:
„Audiovisual Processing of Speech and Non-Speech Oral Gestures"

II.: Behavioral Measures of Temporal and Multisensorial Integration (UB)

Publications:
19 Papers published (+7 submitted) in international (ISI) journals.

Multisensorial Attention
Navarra, J., et al. (In press). Neurosci. Letters
Soto-Faraco, S & Alsius, A. (In press). Neuroreport
Sinnett, C., Spence, C., & Soto-Faraco, S. (in press). Percept.&Psychophys.
Sinnett, C., Juncadella, M. , Rafal, R., Azañón, E. & Soto-Faraco, S. (In press). Neuropsychologia, 2006 May 8; [Epub ahead of print]
Sinnett, S., Costa, A. & Soto-Faraco, S. (2006). Quarterly Journal of Experimental Psychology – Section A, 59(8), 1425–1442
Soto-Faraco, S., Sinnett, C., Alsius, A. & Kingstone, A. (2005). Bulletin & Review, 12(6):1024-31

Multisensorial Integration of Movement
Sanabria, D., Soto-Faraco, S. & Spence, C. (in press). J. of Exp. Psych.: Hum. Percept. & Perform.
Sanabria, D., Spence, C., & Soto-Faraco, S. (In press). Cognition, 2006 Feb 13; [Epub ahead of print]
Sanabria, D., Soto-Faraco, S., Chan, J., & Spence, C. (2005). Neuroscience Letters, 377, 59-64
Sanabria, D., Soto-Faraco, S., & Spence, C. (2005). Experimental Brain Research, 166(3-4), 548-558
Sanabria, D., Soto-Faraco, S., & Spence, C. (2005). Experimental Brain Research, 165(4), 505-514

Multisensorial Integration of Speech
Soto-Faraco, S., Navarra, J., Voloumanos, A., Sebastián-Gallés, N., Weikum, & Werker, J. F. (In Press). Perception and Psychophysic, 2006 May.
Navarra, J., Zampini, M., Soto-Faraco, S., Humphreys, W. & Spence, C. (2005). Cognitive Brain Research, 25, 499-507
Navarra, J. & Soto-Faraco, S. (In Press). Psychological Research, 2005 Dec 14; [DOI: 10.1007/s00426-005-0031-5
Alsius, A., Navarra, J., Campbell, R. & Soto-Faraco, S. (2005). Current Biology, 15(9), 839-843
Navarra, J. & Soto-Faraco, S. (2005). Journal of Exp. Psychology: Human Percept. & Performance, 31(5), 912-918

Others
Soto-Faraco, S., Foxe, J. J. & Wallace, M. T. (2005). Experimental Brain Research, 166(3-4), 287-288
Toro, J. M., Sinnett, S. & Soto-Faraco, S. (2005). Cognition, 97(2), B25-B34
Soto-Faraco, S. (2005). European Journal of Cognitive Psychology, 17(6), 882-885

Phd theses
2 PhD Students have defended their projects successfully in the context of these research lines.
- Jordi Navarra Ordoño
- Scott Sinnett

International projects
We have obtained Funding from two European Community grants:
- STREP project from the FP6-NEST Program 'Measuring the Impossible': Measuring of Naturalness.
- EUROCORES project from the ESF.

III.: Spatio-temporal Codification in Biological Rhythms
(Universidad Autónoma de Madrid , IP: Francisco B. Rodríguez)

Publications:

17 Papers in international (ISI) journals and other contributions.

Memory mechanisms

Rodríguez FB, Huerta R. Neural Networks 17(7): 963-73, 2004
D. Dominguez, K. Koroutchev, E. Serrano, F. B. Rodríguez. Lect. Notes Comput. Sc. 3173: 14-19, 2004
K. Koroutchev, D. R. C. Dominguez, E. Serrano, F. B. Rodríguez. Lect. Notes Comput. Sc. 3173: 20-25, 2004
D. Dominguez, K. Koroutchev, E. Serrano, F.B. Rodriguez. Research in Computer Science. 21:39-48, ISSN: 1665-9899, 2006.
D. Dominguez, K. Koroutchev, E. Serrano, F.B. Rodriguez. Neural Computation. In press.

Control and motor coordination

Corbacho F, Nishikawa KC, Weerasuriya A, Liaw JS, Arbib MA. Biological Cybernetics 93(6): 391-409, 2005a.
Corbacho F, Nishikawa KC, Weerasuriya A, Liaw JS, Arbib MA. Biological Cybernetics 93(6): 410-25, 2005b.
C. Aguirre, D. Campos, P. Pascual, E. Serrano. Lect. Notes Comput. Sc. 3696: 103-108, 2005a.
C. Aguirre, D. Campos, P. Pascual, E. Serrano. Lect. Notes Comput. Sc. 3512: 130-135, 2005b.
C. Aguirre, D. Campos, P. Pascual, E. Serrano. Neurocomputing. 89:1116-1119, 2006.

Neural signatures

R. Latorre, F.B.Rodríguez, P. Varona. Neurocomputing 58-60: 535-540, 2004
A. Tristán, F. de Borja Rodríguez, E. Serrano, P. Varona. Neurocomputing 58-60: 41-46, 2004.
R. Latorre, F.B. Rodriguez, P. Varona. Biological Cybernetics 95(2): 169-183, 2006.
R. Latorre, F.B. Rodriguez, P. Varona. Neurocomputing. In press.
D. Campos, C. Aguirre, E. Serrano, F.B. Rodriguez, G.G. de Polavieja, P. Varona. Neurocomputing. In press.

Others

C. Miravet and F.B. Rodríguez. Lect. Notes Comput. Sc. 3696: 499-505, 2005.
K. Koroutchev, C. Miravet, and F. B. Rodríguez. Image Processing and Computer Vision. Madrid (SPAIN).March 30-April 1, 2005.
C. Miravet, F. B. Rodríguez. Image and Vision Computing, In press 2007.
C. Muñiz, S. Arganda, F.B. Rodríguez, G.G. de Polavieja, P. Varona. Lect. Notes Comput Sc. 3561: 95-105, 2005a.
C. Muniz, S. Arganda, F.B. Rodriguez, G.G. de Polavieja, P. Varona. Society for Neuroscience Abs. 2005, 569.21, 2005b.

Pricit 2006 project

Network Biociencias (UAM, CSIC Materiales, CSIC Instituto Ramón y Cajal y Hospital General de Alcorcón) OLFACTOSENSE (S-SEM – 0255-2006: Análisis de la percepción olfativa para el diseño de sensores de utilidad clínica e industrial) por la Consejería de Educación y Ciencia de la Comunidad de Madrid, en colaboración con el Ministerio de Industria).

Personal Formation

Carlos Miravet (Ph.D. student)
Carlos Muñiz (Ph.D. student, DEA en junio 2005)
Timour Aliev (Ph.D. student)

4 References

A.C. Huk and M.N. Shadlen (2005) Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. J. Neurosci., 25(45):10420-36.

M.L. Platt and P.W. Glimcher (1999) Neural correlates of decision variables in the parietal cortex, Nature 400 (6741):233-238.

J.D. Roitman and M.N. Shadlen (2002) Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. J. Neurosci., 22(21):9475-89.

R. Romo, A. Hernandez, A. Zainos, L. Lemus, and C. D. Brody (2002) Neuronal correlates of decision-making in secondary somatosensory cortex. Nat Neurosci, 5:1217-1225.

R. Romo, A. Hernandez and A. Zainos (2004) Neuronal correlates of a perceptual decision in ventral premotor cortex. Neuron, 41:165-173.

X.J. Wang (2002) Probabilistic decision making by slow reverberation in cortical circuits. Neuron, 36: 955-968.

Jornadas de Seguimiento de Proyectos, 2007, 2007
Programa Nacional de Tecnologías Informáticas

AUTOMATIC STRUCTURE RECOGNITION IN SATELLITE IMAGES BY MEANS OF HYBRID NEUROSYMBOLIC TECHNIQUES. APPLICATION TO OCEAN AND BURNED SURFACES DETECTION. TIN2004-05346

Manuel Cantón Garbín *
Dept of Computer Science and A.I.
University of Almería

Abstract

In this project we propose to continue the research activities developed during the last seven years in the automatic interpretation of AVHRR infrared satellite images and their extension to radar (ASAR), optics (MERIS) and infrared (AATSR) images from ESA ENVISAT satellite. Our team is receiving free ENVISAT data from AOID342 ESA Project. The use of new neurosymbolic techniques applied to satellite images has produced excellent results in the identification of complex ocean structures (upwellings, filaments, gyres, wakes).The rule based expert systems are also producing good results in this task. In this project we propose as novelty the use of Bayesian nets to extract the significant knowledge from the descriptors database and the automatic input of this information to a neurofuzzy net that will identify the relevant features in ENVISAT images. Adding, we have developed a methodology for estimating burned areas and Co2 emissions from wildfires at continental scale using low resolution satellite images. The selected target was the tropical Africa region for the reference year of 1990. For that region, we also have studied the possible relationship between large fires ocurrences and global atmospheric phenomena like "El Niño" as a basis for forecasting fire risk.
We hope also that the use of these new neurosimbolic techniques applied to satellite images can help to improve results about relationships between burnt areas, CO2 estimation and global atmospheric phenomena like El Niño.
Keywords: Image processing, knowledge based systems, pattern classification, remote sensing, hybrid system, graphic expert system,burned area algorithm, NOAA-AVHRR,"El Niño".

*Email: mcanton@ual.es

1 Goals of the Project

This report is organized according to the main research line automatic structure recognition in satellite images. This line is divided into two applications: recognition of oceanic structures and identification of burned surfaces. The specific goals of each research line are summarized in this section.

1.1 Automatic interpretation of oceanic satellite images

We focus on the developing the following specifics goals:

a) The design of a system that fulfill an iterative segmentation region builder based in intelligent threshold detection. For it has been implemented a graphics expert system.

b) The design of a methodology of feature selection that select an optimal or sub optimal feature subset from a set of candidate features. We study the use of Bayesian networks as a mechanism for feature selection in a system for automatic analysis of oceanic satellite images.

c) The design, implementation and evaluation of hybrid systems that improve the automatic interpretation of ocean satellite images. We have done a comparative study of different classifiers: Graphic Expert System (GES), ANN-based Symbolic Processing Element (SPE), Bayesian Network, ANN - Radial Base Function and Fuzzy System, Neuro-Fuzzy System and Fuzzy Lattice Neurocomputing Model. We wish to know the utility of hybrid systems in the recognition of oceanic structures.

1.2 Burned surface detection and CO2 emissions

The international scientific community recognizes the long-term monitoring of biomass burning as important for global climate change, vegetation disturbance and land cover change research on the Earth's surface [1]. Biomass burning combusts Earth's vegetation (in forests, savannas and agricultural lands) and occurs over huge areas of the Earth's surface [2]. Global estimates of biomass burning are thus required in order to provide exact figures of the gas fluxes derived from this source. Although high spatial resolution satellite images may offer a more detailed view of land surfaces, their limited area coverage and temporal sampling have restricted their use to local research rather than global monitoring. Low spatial resolution images provide an invaluable source for the detection of burned areas in vegetation cover (scars) at global scale along time [3].

The specific goals of this part of the project are:

a) Design an automated burned area algorithm (BAA) applicable at continental scale sufficiently robust to accommodate the global variation in burned scar signals from the low resolution satellite images data set: "Global daily National Oceanic and Atmospheric Administration-National Aeronautics and Space Administration (NOAA-NASA) Pathfinder AVHRR 8 km land dataset" [4]. The NOAA/NASA Pathfinder AVHRR 8-km Land (PAL) dataset provides an invaluable source of data for researching global changes on the Earth's surface due to its long time and spatial scales [5].

b) Modelling the temporal behaviour of above-ground biomass using monthly maximum Normalized Difference Vegetation Index (NDVI) composites and biomass density measurements for the main seasonal vegetation classes in the study region [6]. The amount of above-ground burned biomass and therefore CO2 emissions can be estimated from burned land cover area, above-ground biomass density, burn efficiency and emission factor of trace gas by land cover class.

c) Design a software chain for processing the PAL datasets using de BAA algorithm in order to get monthly burned scar maps from the entire period considered (1981-2000) for the tropical Africa region.

d) Modelling the Burned area time series using a Seasonal AutoRegressive Integrated Moving-Average (SARIMA) model for forecasting and studing its relationship with global effects atmospheric like "El Niño" [7].

2 Level of Success

2.1 Automatic interpretation of oceanic satellite images

In this project, the automatic analysis system of AVHRR images has been designed to detect and label the oceanic structures of interest that can be observed by an oceanographer in a satellite image. All the main goals related with the study of different hybrid classifiers has been reached and sometimes surpassed. On the other hand, has been impossible to work with images from the ENVISAT satellite. After three years with changes in the way of distribution, bad organization and many time consumed ordering data, even personally at the data distribution office of the European Space Agency (ESA) in Frascati (Italy), we have now only a set of ten AATSR images with low resolution that are inadequate for our purpouses. So we concentrated in AVHRR images for working and testing our systems.

The system includes an automatic cloud mask system developed using a neural approach, an iterative segmentation region builder based in intelligent threshold detection and a set of different classifiers that we compare in this work and that are described ahead.

2.1.1 Overall structure of the system

Figure 1 represents the overall structure of the system that has been developed in this work.

In a first step, the raw image is processed by standard models to perform the radiometric and geometrical correction and land masking. These are well known techniques also used when the image analysis is performed by human experts.

The second step attempts to detect automatically cloud pixels that are opaque to the ocean radiance data measured in the AVHRR infrared and visible scenes and create a 0´s mask for these areas [8].

The next task, segmentation, divides the image into meaningful oceanic regions. The nature of ocean dynamics makes this fundamental process very difficult, so we have designed a new iterative knowledge-driven method to perform this part of the process [9]. After segmentation we select the type of features or descriptors to describe or represent the segmented regions.

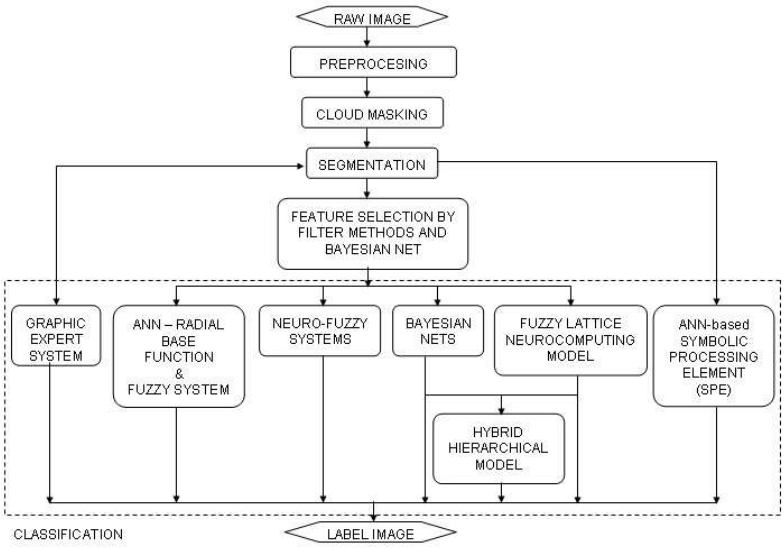


Figure 1: Automatic ocean structure recognition system.

In the feature selection step, we select an optimal or sub-optimal feature subset from the set of features previously chose.

In the last step of the ocean region recognition, each region produced during the segmentation and represented with a selected group of descriptors is analyzed and, if the recognition is positive, it is labeled with the identifier of the matching structure (upwelling, cold or warm eddy or wake). We have implemented a redundant recognition subsystem, with different classifiers:

a) Radial base function network (RBFN) and Sugeno fuzzy systems (FS) (Cooperative neurofuzzy system).

b) Hybrid neurofuzzy system: Adaptative network based fuzzy inference system (ANFIS), neurofuzzy function approximator (NEFPROX) and NEFCLASS.

c) Fuzzy lattice neural networks (FLN).

d) Bayesian networks (BNs) (Naïve Bayes and Tree Augmented Naïve Bayes).

e) ANN-based symbolic processing elements (SPE).

f) Graphic expert system. (GES)

g) Hybrid hierarchical model (HHM).

One of the most important goals of this work has been to compare the efficiency of the different classifiers in the ocean mesoscalar structures recognition.

2.1.2 Segmentation

Segmentation is a key task for any automatic image analysis system. The final results will be good only if high-quality segmentation is achieved. But, as explained in [9], AVHRR images of the ocean are very difficult to segment due to the high variability of the pixel values, yielding poor segmentation results with conventional techniques that rely merely on gradients or textures [10] [11]. The method that we proposed in [9] uses isothermal lines that have been proven to produce good segmentation in this kind of images. This, like any other threshold segmentation, has the inherent drawback of threshold selection, but we have solved this by applying the knowledge obtained in the recognition phase, thus creating an iterative procedure. The initial threshold to segment each structure can be set to a fixed empirical value or to the water-pixel mean. This is then used for segmentation and results are passed to the classifier that labels (upwelling, cold gyre, warm gyre, wake) all the image´s regions. Then, the threshold can be raised or lowered and segmentation repeated with a new threshold.

The new region is then compared with the previous by means of the GES, that perform the recognition of the regions every time that the thereshold change, and the system determines whether the threshold change was favorable, must be reversed, or the task ended at the last value. This method yields good segmentation of AVHRR images, producing compact regions and reducing the oversegmentation resulting from other methods tested like watersheds [10] and Canny edge detectors [11].

2.1.3 Feature Selection

The most common way to proceed with feature selection is to define criteria for measuring the goodness of a set of features [12], and then use a search algorithm that finds an optimal or sub-optimal set of features [13] [14]. We applied the filtering methods with good results. Filtering methods use heuristics based on general characteristics of the data rather than a learning algorithm to evaluate the merit of feature subsets. The best filtering methods used has been correlation based features selection (CFS) [15]. CFS measures correlation between features. The features used are:

- Simple features: area, perimeter, density, volume, equivalent diameter.
- Bounding ellipse: centroid, major and minor axis, orientation, eccentricity, irradiance.
- Bounding box: height, width, area.
- Level of gray: min, max and mean level of gray, standard deviation, barycenter level of grays.
- Inertia moments: Hu's moments, Maitra's moments, Zernike's moments and tensorial moments.
- Etc.

CFS reduce the number of features from 80 to only 16. In order to evaluate the efficiency of feature selection, we have made the cross-validation with Naive Bayes (simple bayesian classifier).

2.1.4 Graphic Expert System

This system is like an oceanographer, collecting the knowledge about the ocean and representing it as production rules. The system filters the information that arrives at the GES in order to obtain only the information relevant to recognition of the ocean structures present in the image. In order to train this component, every ocean feature of interest has to be defined by a human expert using numerical or symbolic descriptors. The hard job here is to figure out the imprecise, sometimes intuitive, deduction scheme that human experts use to perform the task. The inexistence of an exact model for each feature of interest in our problem can lead to gaps or inconsistencies in the knowledge that the expert provides the system with. This is a common problem in expert systems so some techniques have arisen to manage such imprecise or incomplete knowledge (bayesian networks or fuzzy methods). In our system the solution adopted has been to consider that hypotheses are not absolutely proven when the conditions of any of its rules are true. Instead, something plausible is added to it and accumulated through every step of the iterative segmentation recognition cycle.

2.1.5 Hybrid neuro-symbolic systems

We have studied different hybrid classifiers. These are:

a) The relationship between RBFN and FS (zero-order Sugeno fuzzy system) is called a cooperative model. A cooperative neurofuzzy system can be considered as a preprocessor where artificial neural network learning mechanism determines the fuzzy system membership functions or fuzzy rules, from the training data [16].

b) NEFCLASS [17]is a neuro fuzzy classification system derived from the generic model of a 3-layer fuzzy perceptron (it can be viewed as a special 3-layer feed-forward neural network). The nodes in this network use T-norms or T-conorms instead of the activation functions common to neural networks. The first layer is associated with the input features, the hidden layer represents the fuzzy rules (each node is associated to one rule) and the third layer represents output features (in our case the classes). Fuzzy sets are encoded as connection weights. The inputs have real values. The nodes of the output layer use a T-conorm to accumulate the activations of rules, where the T-conorm is usually the maximum. The winner class is determined by the output node with high value.

c) ANFIS [18] implements a zero-order Sugeno fuzzy system and has five layer. The first hidden layer is for fuzzification of the input variables and T-norm operators are deployed in the second hidden layer to compute the rule antecedent part. The third hidden layer normalizes the rule strengths followed by the fourth hidden layer where the consequent parameters of the rule are determined. Output layer computes the overall input as the addition of all incoming signals. ANFIS uses backpropagation learning to determine premise parameters and least mean square estimation to determine consequent parameters.

d) NEFPROX [19] extend NEFCLASS using a general approach for function approximation by means of a neuro-fuzzy model based on supervised learning that use a Mamdani fuzzy system. The main difference between both systems is that NEFCLASS models do not use membership functions on the consequent.

e) FLN [20] emerges as a connectionist paradigm in the framework of fuzzy lattices, which have the advantages of the capacity to deal rigorously with different types of data, such as numerical and linguistic data, value intervals , missing and don't care data. A fuzzy lattice is a conventional lattice L such that the ordering relation has been extended to all elements in L in a fuzzy degree of truth. The *sigma*-FLNMAP (Sigma Fuzzy Lattice Neural MAP) classifier has been introduced in [21]. The *sigma*-FLNMAP classifier is a relationship between two *sigma*-FLN models: *sigma*-FLNa and *sigma*-FLNb. The two models are employed to produce a *sigma*-FLNMAP classifier. In particular, the *sigma*-FLNa clusters the input data, *sigma*-FLNb has the classes (category labels), whereas an intermediate MAP field is used to associate clusters in *sigma*-FLNa with clusters in *sigma*-FLNb. That is, the MAP field assigns a class (in *sigma*-FLNb) to a data cluster (in *sigma*-FLNa). The rules extracted with *sigma*-FLNMAP are represented as hyperboxes.

f) We have tested two simple Bayesian classifiers [22]: Naïve Bayes and Tree Augmented Naïve Bayes Classifier (TAN). Naive-Bayes is oriented to classification, and is based on the assumption that all the features are conditionally independent when the value of the class is known. This assumption implies that the structure of the network is rather simple, since only the arcs in the network links the class with each of the features, and there are no arcs among the feature variables. Naive Bayes is easy to construct and has surprisingly good performance in classification, even though the conditional independence assumption is rarely true in real world applications. A more effective and straightforward way to improve naive Bayes is to extend its structure to represent dependencies among attributes. TAN model is a restricted family of BNs in which the class has no parents nodes and the parents of each feature are the values of the class and at most another feature.

This study has proposed as the best solution for the automatic ocean structure recognition, an hybrid hierarchical model, which has three classifiers: two Bayesian Nets (Naïve Bayes) and one Fuzzy Lattice Neural Networks (*sigma*-FLNMAP). This model has a hierarchical structure of cluster classes. That improves or maintain the accuracy rate of other classifiers used and reduce the number of rules managed by the system.

The figure 2 shows the identifiers of different oceanic structures:

- Upwellings - (ID: 2, 3 and 4)
  - 2 identifies the upwellings between Cape Jubi and Cape Bojador.
  - 3 identifies upwellings southward Cape Bojador.
  - 4 identifies the upwellings in both coast regions.
- Wakes - (ID: 5, 6, 7, 8, 9 and 10).
  - 5 identifies El Hierro's Wake.
  - 6 identifies Gran Canaria's Wake.
  - 7 identifies La Palma's Wake.
  - 8 identifies La Gomera's Wake.

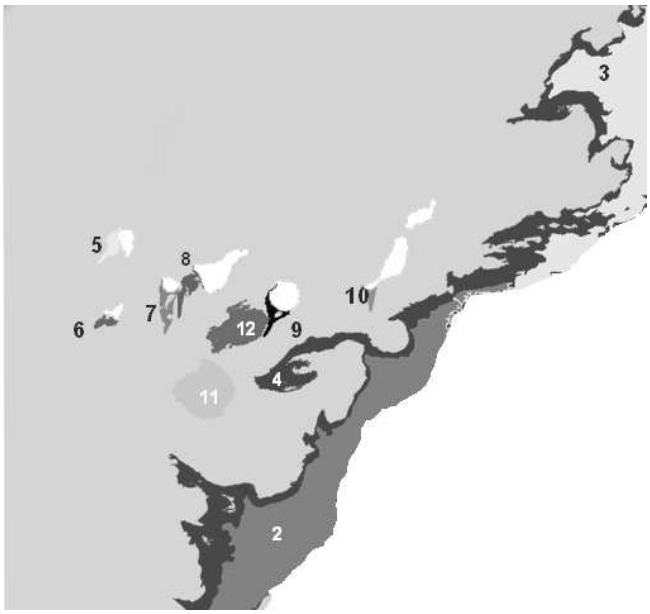


Figure 2: The identifiers of oceanic structures.

  - 9 identifies Tenerife's Wake.
  - 10 identifies Fuerteventura's Wake.
- Anticyclonic eddy - (ID 11).
- Cyclonic eddy - (ID 12).
- No structure - (ID 0).

The classifiers produce good results in the ocean structure recognition, as shown in table 1. We performed the classifications with 10-fold cross-validation. We have worked with 1000 cases (upwellings: 590 cases, wakes: 180 cases, anticyclonic eddies: 10 cases, cyclonic eddies: 40 cases and misclassified regions: 180 cases). The number of classes used are 12 (3 classes of upwellings, 2 classes of eddies, 6 classes of wakes and 1 class of misclassified regions).

The best classifications were provided by NEFCLASS, *sigma*-FLNMAP, GES and the Hybrid Hierarchical Model. It's important to emphasize the work of neuro fuzzy systems, which improve the classification results and enable knowledge to be acquired by means of fuzzy rules (understandable to the user). Of course, the results of the GES are one of the best, with only 50 rules compared with the 367 of NEFCLASS. The worst classification was obtained by RBFN and FS. We think that this is so because the training of the RBFN is limited by the equivalence conditions.

The *sigma*-FLNMAP classifier learns in one run through the training data, furthermore, there is only one parameter to be tuned (the vigilance parameter *sigma*). Vigilance *sigma* received several values, of which $sigma = 0.85$ yielded the best results. Classification performance on the dataset was 92.29

Table 1: Comparative results of classification

classifier	N° of Features	N° of Rules	Accuracy rate
RBFN-FS	14	100	69.13
NEFCLASS	16	367	96.49
NEFPROX	16	420	89.62
ANFIS	14	407	83.63
sigma-FLNMAP	15	156	92.29
Naïve Bayes	16	-	89.68
TAN	14	-	87.08
S.P.E.	50	48	80.00
Hybrid Hierarchical Model	1st level: 11 2nd level: 5 and 25	2nd level: 8	94.93
G.E.S.	18	50	95.00

In the first level, the classification performance of the hybrid hierarchical model on the testing data set was 95.79In the second level, the classification performance was 93.06FLN-MAP obtained an accuracy rate of 95.96The global classification accuracy rate of the hybrid hierarchical model was 94.93This model improves the classification of the rest of the classifiers reducing the number of rules and increasing the accuracy rate. The best classification was produced by NEFCLASS but it uses a very high number of fuzzy rules (about 370 rules) and the linguistic labels for each features is about 100, whereas the hybrid hierarchical model uses 8 rules and 8 linguistic labels (one interval for each one of the rules). Therefore, the efficiency of NEFCLASS is a 1.5(100 linguistic labels and 370 rules) worse than that (8 rules and 8 linguistic labels). The classification performance of GES was 95.00but this system has been built by human experts, that is to say, the learnnig process is manual and not automatic like the neuro-fuzzy system.

2.2 Burned areas analysis

2.2.1 Burned area time series analysis

A 18-year time series of monthly NOAA-AVHRR Pathfinder Land burned area was analyzed for the region of tropical Africa, from July 1981 to June 1999. The period from July 1993 to June 1995 was not included due to missing and outlier data. Stability of the time series was addressed for the input variables in the burned area algorithm, reflectance and temperature channels. Chronograms show a significant interannual seasonal variation in both (being more apparent in the southern region) possibly as a consequence of the influence of the El Niño Southern Oscillation (ENSO) phenomenon-induced Sea Surface Temperature (SST) anomalies in the Pacific Ocean (annual average SST-3.4 and delayed one year and a half BS (burn surface) anomalies were found in the considered period [23]).

2.2.2 Modelling burned areas and SOI Time Series as the basis for forecasting wildfire risk.

Remote sensing data provides valuable systematic and consistent information for environmental time series analysis such as burned area, but special effort should be put in checking time series stability. We addressed the evolution in burned area over 18 years generated from the Daily Tile PAL 8 km data set and how burned area could be forecasted when this data set is combined with additional information.

SARIMA time series modeling was demonstrated to be a powerful tool for this purpose providing results with statistical significance [24]. Lagged SOI was a good predictor variable in the model, proving that El Niño events are factors that drive burned area for tropical Africa. Burned area forecasting could be implemented into wildfire risk assessment integrated with other information on wildfire ignition and wildfire spread probability such as vegetation moisture content or vegetation structure. It could also be applied, together with records on pre-existent vegetation, to estimate the amount of future biomass burned or even to model future atmospheric emissions considering the burning efficiency.
The SARIMA model identified an autoregressive regular term with one month lag and an autoregressive 12 month seasonal term with one season (12 months) component. The cross-correlation between Southern Oscillation Index (SOI) and burned area was significant using a 20-month lag. SOI time series with 20-month lag was identified as predictor variable for inclusion in the SARIMA model. Results showed that the SARIMA model with this predictor did a better job, improving both fitting and forecasting accuracy, and therefore showing the potential relationship between SOI and burned area. Forecasting was estimated by considering only the first 16 years of the monthly burned area time series, from July 1981 to June 1997. The prediction for the following 24 months (from July 1997 to June 1999) was within the 95indicating that the forecast was a valid characterization of the modeled process.

3 Result Indicators

The main indicators are classified in the following categories: formation of students, publications, spanish projects, international and spanish collaborations.

3.1 Formation of Students

3.1.1 PhD. Thesis

a) D. José Antonio Piedra Fernández, member of Project and associate professor in Almería University, finished his doctoral thesis titled "Automatic interpretation of satellite images by means of neurofuzzy system" with Dr. Manuel Cantón as advisor. He presented it by in July 2005.

b) D. Francisco Guindos Rojas, member of this project and TEU at Almería University finished his PhD "Intelligent segmentation and recognition of ocean structures in AVHRR satellite images". He is writing the text. Planned defense for April 2007.

3.1.2 Final degree projects

a) D. David Yétor Sánchez finished his project titled "Application of j2me technology to the image processing by means of mobile devices". He presented it by in January 2004.

b) D. Fernando Pérez Rodríguez finished his project titled "The client server environment for the remote access to camera by means of mobile technology". He presented it by in March 2004.

c) D. Alberto Molina Martínez finished his project titled "Integration and development of a recognition system of satellite images". He presented it by in March 2005.

d) D. José Alberto Gea Pérez finished his project titled "Object-Oriented Distributed Mobile technology to business environments. Hospital application". He presented it by in January 2006.

e) D. Victor Andrés Andrés finished his project titled "Image capture with mobile devices applied to the expert work of vehicles in a client server environment". He presented it by in March 2006.

f) D. Victor Escudero Cabrera finished his project titled "video-monitoring intelligent system integrated in mobile technology and webcams". He presented it by in September 2006.

g) D. Diego Raúl López Garrido finished his project titled "Mail system focused on the spam identification". He presented it by in January 2007.

h) D. Luis Manuel Domene Miralles finished his project titled "Artificial vision mobile system to blinds". He presented it by in January 2007.

3.2 Publications

3.2.1 Paper in international journals

- MORENO J.A., CANTÓN M. (2004) Estimating burned area for tropical africa for the year 1990 with the NOAA-NASA pathfinder AVHRR 8KM land dataset. International Journal of Remote Sensing; 25(17), 3389-3410.

- MORENO J.A., BARÓN J., CANTÓN M. (2005) Estimating above-ground burned biomass and co2 emissions for tropical Africa for the year 1990 with the NOAA-NASA pathfinder avhrr 8 km land dataset. International Journal of Remote Sensing; 26(11), 2407-2422.

- MORENO J.A., BARÓN J., DEL SAGRADO J. (2006) Burned data time series for tropical Africa (1981-2000) with the daily tile NOAA-NASA pathfinder AVHRR 8-km land dataset. International Journal of Remote Sensing; 27(6),1119-1134.

- RIAÑO D., MORENO J.A., BARÓN J. AND USTIN S.L. (2006) Burned area surface forecasting using past burned area surface records and southern oscillation index for tropical Africa (1981-1999). Remote Sensing of Environment. (in press).

- PIEDRA J.A., CANTÓN M., GUINDOS F. Automatic analysis of satellite images using an hybrid hierarchical model. IEEE Trans. On Geoscience and Remote Sensing. (Submitted).
- PIEDRA J.A., CANTÓN M., GUINDOS F., MOLINA A. (2005) Automatic recognition of ocean structures from satellite images by means of hybrid systems. Computer aided system theory, 3643:373-378.
- GARCIA L., LUQUE A., TEJERA A., BERGASA O. and CANTÓN M. Analysis of Canary island induced eddies during the year 1998 observed with infrared and altimeter data. Submitted to Deep Sea Research Part I.

3.2.2 Future papers from the PhD. Thesis of GES (F. Guindos)

- Artifact Reduction in Watershed Segmentation of Low-Variability Images.
- Expert System for Detection of Mesoscale Ocean Structures.
- Result Validation of Knowledge Based Detection of Mesoscale Ocean. Structures.

3.2.3 Papers in national journals

- PIEDRA J.A., CANTÓN M, GUINDOS F. Interpretación automáticas de imágenes de satélite mediante sistemas neurodifusos. Revista española de Teledetección. In press
- PIEDRA J.A., CANTÓN M., GUINDOS F., MOLINA A. Análisis automático de imágenes oceánicas de satélite mediante sistemas basados en conocimiento. Revista española de Teledetección. In press

3.2.4 Book chapters

- CANTÓN M. Satellites observations of oceans in relation with Global Change. Springer Verlag. In press. May 2007.
- PIEDRA J.A., CANTÓN M., GUINDOS, F. Applications of Fuzzy Lattice Neurocomputing for Pattern Recognition in Ocean Satéllite Images Computational Intelligence Based on Lattice Theory, Springer-Verlag, Heidelberg, Germany. In press, summer 2007.

3.2.5 Papers in international an national conferences

- MORENO J.A., RIAÑO D., BARÓN J., USTIN S.L. (2005) Burned area forecasting using NOAA-NASA pathifinder AVHRR 8km land data set for the tropical Africa (1981-2000) as the basis for forecasting wildfire risk. 5th international workshop on remote sensing and GIS applications to forest fire management: fire effects assessment. Zaragoza (Spain), 223-227.
- PIEDRA J.A.; CANTÓN M ; CASTRO J.L.; GUINDOS, F. Automatic analysis o of satellite images using bayesian nets and fuzzy lattice neurocomputing model. information processing and management of uncertainty. In knowledge based Systems (IPMU 2006), Paris 02/07/2006

- PIEDRA J.A., CANTÓN M., GUINDOS, F. Reconocimiento automático de estructuras oceánicas mediante sistemas basados en conocimiento. Red nacional de teledetección marina. Instituto de ciencias marinas de andalucía y el centro andaluz de ciencia y tecnología marinas. . 28/11/2005 - 29/11/2005

- PIEDRA J.A., CANTÓN M., GUINDOS F. AUutomatic recognition of ocean structures from satellite images by means of hybrid neurofuzzy systems and fuzzy lattice neurocomputing model. European Space Agency (ESA). ESA-EUSC 2005: Image Information Mining - Theory and Application to Earth Observation. ESRIN, Frascati, Italia, 05/10/2005 - 07/10/2005.

- PIEDRA J.A., SALMERÓN A., CANTÓN M., GUINDOS F. and RUIZ D. Reduction of irrelevant features in ocean satellite images by means of ´ bayesian networks. VI Jornadas de Transferencia Tecnológica de Inteligencia Artificial.I Congreso Español de informática. Palacio de congresos y exposicioes de Granada, 13/09/2005 - 16/09/2005, 133-140.

- PIEDRA J.A., CANTÓN M., GUINDOS F., MOLINA A. (2005) Automatic recognition of ocean structures from satellite images by means of hybrid systems.TENTH International Conference on computer aided systems theory, EUROCAST 2005.Las Palmas de Gran Canaria, España, 07/02/2005 - 11/02/2005.

- GUINDOS F.; CANTÓN M.; PIEDRA J.A. Ocean features recognition in avhrr images by means of bayesian nets and expert systems. 3RD International Workshop on Pattern Recognition in Remote Sensing. Kingston University, 27/08/2004, 1-8.

- GUINDOS F.; CANTON M.; TORRES J.A.; PERALTA M.; PIEDRA J.A.; MOLINA A. Automatic recognition of ocean sructures from satellite images by means of neural nets and expert systems. Proceedings of ESAEUSC 2004 theory and applications of knowledgedriven image information mining with focus on earth observation. ESA-EUSC Madrid, ESPAÑA, 17/03/2004 18/ 03/2004, 1-5.

- Participation in the CDTI meetings about Space and EU-ESA GMES programs in relation with the VII EU Program.

3.3 Spanish Projects

- Integración de información multiespectral e hiperespectral para la selección y caracterización de indicadores de cambio en ecosistemas áridos. bases para un programa de seguimiento del p.n. cabo de gata-níjar.(INCAMAR)", REF. CYCIT- CGL2004-02063. Ministerio de Ciencia y Tecnología. 13/12/2004 - 12/12/2007.

- Red nacional de Teledetección Marina (RETEMAR).Centro Público I+D.Programa Nacional del Espacio.Investigadores y tecnólogos de 15 centros de investigación (OPIS, UNIVERSIDADES Y EMPRESAS).

3.4 International and Spanish Collaborations

- Analysis of mesoscale ocean features in nw africa coastal transition zone and canary islands region. European Space Agency (ESA). AOID-342.
- New relation established recently with Prof. V. Karbulassos from Tessalonica University (Greece), developer of the Fuzzy Lattices. Testing the possibilities for a proposal in the VII EU Research Program.
- Strong relation with Dr. J.M. Corchado from Salamanca University, specialist in intelligent systems. Another possible partner for the EU proposal.

References

[1] Hao, W.M. and Liu, M.H. (1994), Spatial and temporal distribution of tropical biomass burning, Global Biochemical Cycles. 8: 495-503.

[2] Levine, J.S. (1991). Global biomass burning. Atmospheric, Climatic, and Biospheric Implications, The MIT Press, Cambridge, Massachusetts.

[3] Justice, C.O., Kendall, J.D., Dowty, P.R. and Scholes, R.J. (1996), Satellite remote sensing of fires during the SAFARI campaign using NOAA advanced very high resolution radiometer, Journal of Geophysical Research-Atmospheres. 101(D19): 23851-23863.

[4] Barbosa, P.M., Grégoire, J.M. and Pereira, J.M.C. (1999), An algorithm for extracting burned areas from time series of AVHRR GAC data applied at a continental scale, Remote Sensing of Environment. 69: 253-263.

[5] Agbu, P.A. and James, M.E. (1994). The NOOA/NASA Pathfinder AVHRR Land Data Set User's Manual, Goddard distributed Active Archive Center, NASA, Goddard Space Flight Center, Greenbelt (USA).

[6] De Fries, R.S., Hansen, M., Townshend, J.R.G. and Sohlberg, R. (1998), Global land cover classifications at 8 km spatial resolution: the use of training data derived from Landsat imagery in decision tree classifiers, International Journal of Remote Sensing. 19(16): 3141-3168.

[7] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994). Time series analysis: forecasting and control, Prentice Hall, Upper Saddle River, N.J.

[8] Torres JA, Guindos F, Peralta M and Cantón M (2003) An Automatic Cloud-Masking System Using Backpro Neural Nets for AVHRR Scenes. IEEE Transactions On Geoscience and Remote Sensing, 41: 826–831.

[9] Guindos F, Piedra JA and Cantón M (2004) Ocean features recognition in AVHRR images by means of bayesian net and expert system. 3rd International Workshop on Pattern Recognition in Remote Sensing, Kingston University, United Kingdom, 11: 1–4.

[10] Beucher S (1990) Segmentation Tools in Mathematical Morphology. SPIE-1350 Image Algebra and Morphological Image Processing, 70–84.

[11] Canny J (1983) Finding Edges and Lines in Images. Artificial Intelligence Lab., M.I.T., Cambridge. MA. Tech. Rep. 720.

[12] Yi Lu Murphey and Hong Guo (2000) Automatic Feature Selection - an hybrid statistical approach. Proc. Int. Conf. Pattern Recognition, Barcelona, Spain, 3–8.

[13] Langley P and Sage S (1994) Induction of selective Bayesian classifiers. In Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence, Morgan-Kaufmann, 399–406.

[14] Yamagata Y and Oguma H (1997) Bayesian feature selection for classifying multi-temporal SAR and TM Data. Int. Geoscience and Remote Sensing Symposium, Proc. IGARSS1997, 2: 978–980.

[15] Hall MA (1999) Correlation-based Feature Selection for Machine Learning. Ph.D diss. Hamilton, NZ: Waikato University, Department of Computer Science.

[16] Jang R. and Sun C. (1993) Functional equivalence between radial basis function networks and fuzzy inference systems. IEEE Trans. on Neural Networks, 4: 156–159.

[17] Nauck D. and Kruse R. (1995) NEFCLASS - A neuro-fuzzy approach for the classification of data. In Proc. ACM Symp. Applied Computing, Nashville, TN.

[18] Roger Jang, "ANFIS Adaptative Network based Fuzzy Inference System", IEEE Transactions On Systems, Man and Cybernetics, May/June 1993, vol 23 (3).

[19] Nauck D. and Kruse R. (1998) A neuro-fuzzy approach to obtain interpretable fuzzy systems for function approximation. Fuzzy Systems Proceedings, IEEE World Congress on Computational Intelligence, 2: 1106–1111.

[20] Kaburlasos V. G. and Petridis V. (2000) Fuzzy lattice neurocomputing (FLN) models. Neural Networks, 13: 1145–1170.

[21] Petridis V. and Kaburlasos V. G. (1999) Learning in the framework of fuzzy lattices. IEEE Transactions on Fuzzy Systems, 7: 422–440.

[22] Duda R.O. and Hart P.E. (2001) Pattern classification. 2nd Edition John Wiley and Sons, New York.

[23] Riaño D., Moreno J.A., Barón J. and Ustin S.L. (2006) Burned area surface forecasting using past burned area surface records and southern oscillation index for tropical africa (1981-1999). Remote Sensing of Environment. (In press).

[24] Chu, P.S., Yan, W.P. and Fujioka, F. (2002), Fire-climate relationships and long-lead seasonal wildfire prediction for Hawaii, International Journal of Wildland Fire. 11(1): 25-31.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

A Methodology for Analysis and Design of Hybrid Systems: Integration of Software and Knowledge Modeling Techniques TIN2004-05694

Samuel Túnez[*]
Universidad de Almería

José Palma[1]
Universidad de Murcia

Abstract

The main goal of this project is the definition of a methodology allowing the development of hybrid software systems. Hybrid systems are those including knowledge-based components and standard information systems. Knowledge-based components try to solve problem aspects where requirements are poor structured and subjective, and there are uncertainty, incompleteness and inconsistency too. Standard information systems represent a valid solution for problem aspects with objective requirements, with a quantitative or qualitative result. The proposed methodology must support the development on the whole of these components, getting the methods, tools and techniques of other methodologies, called "support methodologies". They have shown their validity for developing each of these two kind of components. Furthermore, we want to develop a tool that guides the application of the methodology. Methodologies currently available do not solve this integration successfully. Applying software engineering methods it not possible to model knowledge, and the knowledge engineering methods can not be used if there is not knowledge. These kinds of shortcomings have appeared during the development of decision support systems in agriculture carried out by the research group. Furthermore, we have proposed the validation of this methodology applying it in the development of a Decision Support System for the Integrated Surveillance in Agriculture (SAVIA), that help us to make decisions in the crop monitoring and to keep the Integrated Production quality standard.

Keywords: Software Engineering and Knowledge Engineering Integration. Knowledge-based Systems. Software Development Methodology. Decision Support Systems in Agriculture.

1 Project goals

The specific goals of the project are as follow:
1. Design of a methodology (InSCo-met) for developing hybrid software systems. This goal is specified with the definition of a development process, describing the activities, models and modeling tools needed, as well as the generated documentation and the control activities. The

[*] Email: stunez@ual.es
[1] Email: jpalma@dif.um.es

development process is divided into two abstraction levels, the conceptual and computational levels. The specification of the methodology starts from well-known and widely tested methodologies of SE (Knowledge Engineering) and SE (Software Engineering), focusing our efforts on their integration (InSCo).

2. Development of supporting tool for the InSCo methodology. Due to the features of the hybrid systems (large, poorly structured, imprecise and high risked), the defined methodology must be supported by a application guiding and making the development process easier.

3. Development of an hybrid software system using this methodology and supporting tool. In order to accomplish this goal, the "Sistema de Asesoramiento para Vigilancia Integrada en Agricultura" (SAVIA) system will be developed; it tries to solve the problem of managing the Integrated Production quality mark, starting with tomato, courgette, pepper and table grape.

4. Evaluation of InSCo methodology and developed SAVIA system. This is a twofold objective; on the one hand, the evaluation of the methodology defined in goal 1, putting it into practice developing the system included in goal 3; and on the other hand, the test with real cases of the system developed in goal 3, preparing it for subsequent implantation.

1.2. Project Schedule

Figure 1 shows the project schedule, the level of achievement of the tasks and the milestones reached. The project started on December 2004. At the moment of writing this report it is located 11 months left.

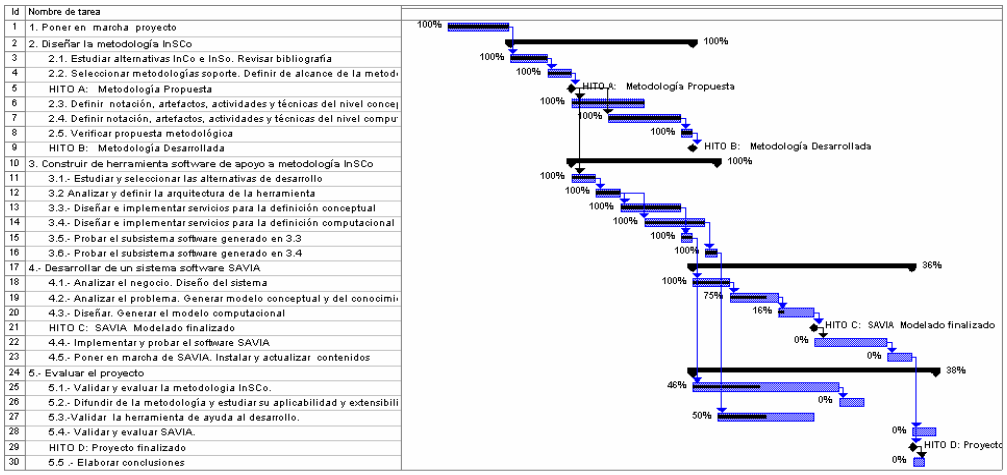


Figure 1. Project Tasks and Scheduling

2 Success level achieved in the project

This section describes the goals achieved, problems found, solutions proposed and research results. The goal of evaluating InSCo methodology and SAVIA system will be achieved in the project time left.

2.1. Design of the InSCo-met methodology

The main goal of this project is the definition of the InSCo-met methodology supporting the development process of hybrid software systems. Applying this methodology, two main tasks must be performed: the conceptual modeling and the computational modeling from conceptual models. The main result obtained is the methodology process model as a convergence point between SE and KE. As a result of conceptual modeling activity, we confirmed the necessity of designing a new analysis model, which allows the whole description of non-knowledge-based and knowledge-based requirements. Nevertheless, dynamics aspects of the process model are not finalized yet because of the difficulties found in the link between conceptual and computational aspects, in particular, the definition of the form and content of the analysis model. A process of application of the Model Driven Architecture (MDA) in the development of hybrid software systems has been defined, allowing to obtain computational models from conceptual ones. This process is focused in the development of knowledge-based components integrated in hybrid software systems.

Result 1.- Selection of supporting methodologies.

Designing a new methodology for hybrid software systems implies the study of existing methodologies and the selection of suitable supporting methodologies, in order to avoid an important problem: the price of change. The solution means to extend or improve well-known methodologies with whatever they lack, or to integrate and converge methodologies from both SE and KE. After studying the most used methodologies in both fields, the selected supporting methodologies are: Structured, Rational Unified Process, and CommonKADS. The first and second are selected because they cover the most used approaches for software systems development nowadays; CommonKADS is chosen because it is one of the most detailed and extended methodologies in Europe for knowledge-based systems development. Convergent topics between those methodologies have been found, such as common tools and notations. Also, the creation of computational models from conceptual models is supported by the MDA, which defines an approach to software development based on modeling and allows the definition of transformations between models, making easier the definition of the structured process.

Result 2.- InSCo-met process model.

The main result obtained is the definition of a process model for hybrid software systems development. We consider it as a new convergent point between SE and KE, and the first step towards a common methodology.

The InSCO-met process model defines the activities, artifacts and tools needed in hybrid software systems development. It uses an activity-oriented approach and defines a set of basic activities or workflows. These basic activities are divided into some activities or tasks, and they obtain and use some artifacts, standing out the models as fundamental results in hybrid system development. Each model defines the system from its specific point of view and abstraction level.

The InSCo-met process model differentiates between two main model types and modeling activities: conceptual modeling, close to the problem world, and computational modeling, close to the solution world. This differentiation has caused a non-desired separation between the conceptual and computational parts, similar to the classical separation between specification and design phases in general software development.

Result 3.- InSCo-met Workflows

InSCo-met activities and models have been classified into four packages. The computational level is placed near the software solution, and the other three conceptual level packages belonging to the

problem world. The conceptual level provides a unified point of view of knowledge-based and non-knowledge-based software components. Their activities are focused in the analysis of the environment where the system will work: "Business" package groups the activities and models defining the problem and the context of the software solution; "Client oriented development" and "systems oriented development" packages has been defined separately in order to minimize the connexions with the computational models. In the last package, the knowledge of the hybrid software system is injected. Now, the specification obtained is ready to be translated to the software application.

Result 4.- Necessity of Analysis Model
Conceptual models are the connection between the real world and the software system to be developed. A Hybrid software system has several subsystems, and everyone can be specified using a different conceptual modeling language. Some of them use SE language, such as UML, to describe the client requirements. Other conceptual models are described using a KE language as CML (Conceptual Modeling Language). These heterogeneous conceptual models must be joined into an unified description of the whole system. This model, called the Analysis Model, is a complete description of the knowledge and the requirements. It is easier to be maintained and simplifies the system structure. This model is a basic element in hybrid software system development, but its necessity has been appeared in other methodological proposal developed in order to build a bayesian network based system, submitted and pending of acceptance.

Result 5.- Towards the computational level
In order to obtain the computational models from conceptual models, we proposes the application of MDA approach to software development, focusing in the use of this approach in knowledge-based components. As MDA proposes, a metamodel for the platform independent modeling language has been specified: the CML metamodel describes the language for knowledge modeling. Several implementation platforms are considered as implementation technologies of knowledge-based components, such as Jess, the most used rule representation language and inference engine for Java-based rule systems, and G2, a commercial and powerful tool for knowledge based systems implementation. A metamodel for designing platform specific models (PSM) in Jess has been defined. The major and current difficulty is the definition of the transformation from CML models to Jess platform specific models.

2.2. Development of an InSCo-met supporting tool

In order to put InSCo-met in practice, we need software tools that allow the collaboration among clients, users, software and knowledge engineers, designers and testers. This is the way we will be able to deal with the development of solid software in all the aspects related to hybrid software systems. However, we have found many difficulties to define the Analysis Model causing this objective has not been completed yet. For this reason, we have focused on InSCo-met tool, and we are currently working on the Analysis Model definition.

Several commercial alternatives, such as IRqA, Rational Rose or Eclipse, were considered in order to apply our methodological proposal. However, they were discarded because there is no Web version available for them, difficulting the access to clients and users. In spite of this, it is possible to share information with these tools by means of standard formats as XMI (XML Metadata Interchange) allowing interoperability among. At this point, it is important to remark that we are having some difficulties translating the different models into XMI format, due to the different versions of this standard.

Result 6. InSCo-met tool
This tool is the most important technological result reached by now. InSCo-Requisite is a Web application developed using the J2EE platform (Java 2 Enterprise Edition), applying a MVC (Model View Controller) pattern. This pattern implementation is performed by means of Jakarta Struts, an open-source and multiplatform framework. Although Jakarta Struts is now an independent project, it was previously developed as a part of the Jakarta Project (Apache Software Foundation). The DMBS used is Oracle. Finally, in order to have a Web application properly working, we need a Web Server, whose main function is to manage all the requests showing the appropriate data to the user. In our case, we chose Jakarta Tomcat, because it gives support to JSP (Java Server Pages) and servlets, an essential feature for our application. InSCo-Requisite tool is available at http://www.dkse.ual.es/insco/.

2.3. Development of SAVIA system

This goal means the development of a decision support system for integrated surveillance in agricultural crop production (SAVIA). The system is based on the Integrated Production quality protocol, which regulates the production of tomato, courgette, pepper, and grape crops. The business analysis has been performed with the collaboration of agricultural technician experts. However, this task is taking more time than planned, because of the difficulties on knowledge extraction in the agricultural domain, which is a complex environment affected by many changeable market requirements. On the other side, the non-knowledge-based part of the system has been already developed, and it is widely and successfully used by agricultural technicians, as we mention in section 3.3.

Result 7.- Phytosanitary Information System
Phytosanitary Information management in SAVIA system has become more important than it was initially expected, due to the lack of similar systems in the plant protection context. Several functionalities have been developed for the Web http://www.juntadeandalucia.es/agriculturaypesca/ (links: Agricultura - Plagas en los cultivos y Red de alerta e Información Fitosanitaria). These applications provide technicians with useful tools to generate adapted reports, in order to make decisions in plant protection control. The most important reports the application provides are: Phytosanitary Products list, Phytosanitary Products registered for different allowed applications, Phytosanitary Product Card from its identifier, Authorised Products for Pest Control, Biological Products for Pest Control, Effects of Active Material on Biological Products, LMR (Maximum Limit of Residues) of a active substance on the different vegetal products.

2.4. InSCo-met evaluation

Once the B milestone (methodology developed) was reached, the evaluation process started. The evaluation tasks showed the difficulties and problems mentioned previously. Besides the methodology, the InSCO-Requisite tool is being applied in the first business analysis stages of SAVIA system, and tested at the same time by the users.

3 Achievement indicators

3.1 Training accomplishments

Currently, there are 5 people under research training.

Three PhD Theses in progress, expected to be defended in 2007:

- InSCo. A Process Model for the Integration of Software Engineering and Knowledge Engineering. Phd. student: Isabel María del Águila Cano.
- Software Technology applied to the Design of Knowledge-Based Systems. Phd. student: José Joaquín Cañadas Martínez.
- Fuzzy Temporal Patterns. Phd. student: Francisco Gabriel Guil Reyes.

Two DEAs, one of them defended in 2006 and the other will be defended in 2007:

- Requirement Management on Knowledge Engineering. Researching student: Francisco Javier Orellana Zubieta.
- Sistema de Gestión de Información Fitosanitaria: Optimización de los Registros de Productos. Researching student: Martín Miguel Acebedo Vaz.

3.2 Publications

The following table shows the number of publications related to the project, listed in the References section.

International Journals	4	[2] [4] [7] [9]
National Journals	1	[11]
National Book Chapters	1	[12]
International Conferences	3	[3] [5] [10]
National Conferences	3	[1] [6] [8]

3.3 Technology transfer

Several companies and organizations have taken part on the technology transfer performed during the project course. They showed an interest in the results, as the "Plan de Difusión y Explotación de resultados" of the "Memoria Científico-Técnica" reflects, and their participation as EPO. This report was recently submitted to the Subdirección General de Gestión Económica y Fondos Estructurales Comunitarios. The main results of this transfer are:

a) Requirement Engineering on Software and Knowledge Engineering:

A collaboration agreement was carried out by the University of Almería and the company TCP Sistemas e Ingeniería (http://www.tcpsi.com). This company is the developer of IRqA, a CASE tool for requirement engineering. The main activity of this company is to offer advanced solutions of software and communications development. The results of this agreement have been:

- The conference "Las necesidades de los usuarios como base para la obtención de calidad: Herramientas para la gestión integral de requisitos", held in the University of Almería the 4th February 2005. It was aimed at directors and managers of computer departments of public organizations.
- The course: "Ingeniería de Requisitos: Software de Calidad". It was held in the University of Almeria on June 2005. This course aimed at the computer engineers community and offered a methodological and practical perspective for developing quality software.

b) Decision support systems in agriculture:

From 2004 to 2006, three collaboration agreements with the Dirección General de la Producción Agraria de la Consejería de Agricultura y Pesca de la Junta de Andalucía, were established, funded

with 496.254 €. These agreements have implied the development of several Web applications located at http://www.juntadeandalucia.es/agriculturaypesca/. In 2007, we are currently transferring these applications to the Web portal Besana (http://www.portalbesana.es), which is funded by eleven banks and D.A.P, Empresa Pública para el Desarrollo Agrario y Pesquero de Andalucía, S.A. The last one offers news and technical information about the agriculture, livestock, fishing and food sectors.

Finally, the conference "Aplicación de las tecnologías de la información y las comunicaciones en la agricultura almeriense" was held in the University of Almería on 29th and 30th of November in 2006. It was organised by the Instituto de Estudios Almerienses, in collaboration with the Consejería de Agricultura y Pesca e Innovación, Ciencia y Empresa de la Junta de Andalucía, the University of Almería and PHYTOMA – Spain. This conference was coordinated by Samuel Túnez Rodríguez.

3.4 Participation in Projects

- *Gestión de información de Sanidad Vegetal mediante Internet.* Sponsored by the Agencia Española de Cooperación internacional – AECI. Participants: University of Almería (España) y Escuela Agrícola Panamericana ZAMORANO (Honduras). Duration: from March 2005 to February 2006. Fund amount: 14.500 €. Head researcher: Samuel Túnez Rodríguez.
- *Comunidad Iberoamericana de Conocimiento sobre Agricultura Protegida.* Sponsored by CYTED. Participants: Universidad de Almería (España) and other 11 groups. Duration from 2003 to 2006. Fund amount: 90.000 (US $). Head researcher: José Fernando Bienvenido Barcena.
- *Un sistema de alerta e información fitosanitaria, georreferenciado y accesible mediante Internet* (Referencia: P06-TIC-02411). Sponsored by the 2006 Incentivos a Proyectos de Investigación de Excelencia en Equipos de Investigación Andaluces. Duration from 2007 to 2010. Fund amount: 123.636,30 €. Head researcher: Samuel Túnez Rodríguez.

3.5 Collaborations with International and National Groups

During the course of the Project we have established collaborations and with the following groups:

- The participation in Knowledge Engineering and Software Engineering Workshop (KESE-2005), allowed us to establish contacts with members of the Department of Applied Informatics and Artificial Intelligence, at the University Würzburg in Germany. A member of our research group has applied for a grant from the Programa de Movilidad del profesorado de la Junta de Andalucía 2007 offers. In addition, along with this group, contact has been established with groups in other countries within the European Union, in order to apply a COST action to coordinate this project with other projects already existing in Europe.
- A member of the research group has ended a stay of one month at the department of Languages and Computer Science at the Madrid Polytechnic University (research field SE and KE). Her goal consisted in making the first steps towards the definition of the architecture of the suggested methodology.
- The work developed in the field of the MDA, has permitted the cooperation with the Software Engineering group at the Murcia University, more specifically, with the group headed by Dr. Jesús García Molina, whose work is focused on applying MDA in the development of Web applications. This cooperation allowed a member of our research group to stay five months at the Murcia University.
- The Data Analysis group of the University of Almeria was interested in one of the results of this project and we are currently cooperating towards the definition of a process model for the

development of Bayesians networks based systems. A first result, "A process model for Bayesian networks based systems development", was submitted to the European Journal of Information Systems, and it is pending for acceptance.

- In the field of decision support systems for agriculture, there have been established different collaborations with institutions and university groups from Central and South America. Our group has been responsible for a project AECI and has taken part in a project CYTED.

4 References

[1] *Aplicación de MDA a la Ingeniería del Conocimiento*. José Joaquín Cañadas Martínez, Samuel Túnez Rodríguez, José Tomás Palma Méndez. Actas del I Taller sobre Desarrollo de Software Dirigido por Modelos, MDA y Aplicaciones (DSDM04), pg. 33-42. Málaga, España. Octubre, 2004.
[2] *Decision Support System for Pest Management using Biological Control Organisms*. Martín Miguel Acebedo Vaz, M. Navarro, José R. Díaz Álvarez. Acta Horticulturae 659, pg. 287-294. 2004.
[3] *Herramientas informáticas de Ayuda a la Toma de Decisiones y Fertirriego*. Manuel Erena, Martín Acebedo, José A.Gil, José J. Matarín, Samuel Túnez. IV Congreso CIDAPA, pg. 23-26 Bogotá, Noviembre 2004.
[4] *A WEB-CASE Tool Prototype for Hybrid Software Development*. Francisco Javier Orellana Zubieta, Fco. Gabriel Guil Reyes, Isabel Mª del Águila Cano, Samuel Túnez Rodríguez. Lecture Notes in Computer Science 3643, pg. 217-222. 2005.
[5] *From Knowledge level to Symbolic level. A metamodeling approach*. José Joaquín Cañadas Martínez, José Tomás Palma Méndez, Samuel Túnez Rodríguez. Proceedings of Knowledge Engineering and Software Engineering Workshop, pg. 13-23. Koblenz, Alemania. Septiembre 2005.
[6] *Un Sistema de ayuda a la toma de decisiones para control fitosanitario en agricultura*. José Joaquín Cañadas Martínez, Isabel Mª Del Águila Cano, Alfonso José Bosch Arán, Samuel Túnez Rodríguez. Actas de la XI Conferencia de la Asociación Española para la Inteligencia Artificial, pg. 391-400. Santiago de Compostela, España. Noviembre 2005.
[7] *An Iterative Method For Mining Frequent Temporal Patterns*. Fco. Gabriel Guil Reyes, Antonio Bautista Bailón Morillas, Alfonso Jose Bosch Arán, Roque Marín Morales. Lecture Notes In Computer Science 3643, pg.189-198. 2005.
[8] *Evaluación de expresiones temporales a partir de secuencias frecuentes*. Fco. Gabriel Guil Reyes, Alfonso Jose Bosch Arán, Roque Marín Morales. Actas de la XI Conferencia de la Asociación Española para la Inteligencia Artificial, pg. 41-50. Santiago de Compostela, España. Noviembre 2005
[9] *A tractable subclass of fuzzy constraint networks*. Alfonso bosch, Francisco Guil, Roque Marín. Lecture Notes In Computer Science 3643, pg.269-274. 2005.
[10] *Towards a Methodology for Hybrid Systems Software Development*. Isabel María del Águila, Joaquín Cañadas, José Palma, Samuel Túnez. Proceedings of Eighteenth International Conference on Software Engineering and Knowledge Engineering, pg 188-195. San Francisco, USA. Julio, 2006.
[11] *Ofertas de Información Fitosanitaria en la Web de la Consegería de Agricultura y Pesca de la Junta de Andalucía*. Jesús Gonzáles, Martín Acebedo, Samuel Túnez. PHYTOMA España, 184, pg. 22-25. 2006 (Iª Jornadas sobre Aplicación de las Tecnologías de la Información y las Comunicaciones al Sector Agrícola Almeriense).
[12] *Modelo de Conocimiento en un Sistema de Asesoramiento Experto en Producción Integrada en Agricultura*. Isabel Mª del Águila, Samuel Túnez. La Gestión del Conocimiento en las Organizaciones. Ed. Enrique Paniagua. Universidad de Murcia (Pendiente de publicación).

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

DEFINITION AND INSTANTIATION OF AN INTEGRATED DATA MINING PROCESS TIN2004-05873

Javier Segovia Pérez*
Universidad Politécnica de Madrid

Abstract

In practice, CRISP-DM is the most commonly used data mining process in both industry and academia. CRISP-DM has one major weakness: it is at once a process model, methodology and lifecycle. Therefore it lacks definition and detail. Also, the data mining process is completely decoupled from the software engineering process, even though its results have a definite impact on this process. This methodological deficiency is one of the main reasons why many data mining projects are not completed or why, if they are, they fail to meet customer expectations and are not used. This project aims to mitigate the above problems. To do this, it sets out to: (i) define and integrate a data mining process with a software process, examining what tasks they have in common, their respective inputs and outputs and CRISP-DM's weaknesses, (ii) develop a process instance tailored to CRISP-DM and unify this instance with the Unified Process (RUP) and (iii) validate and transfer the technology to real cases.

1 Project objectives

The specific project objectives are as follows:

1. **Define and Integrate a Data Mining Process with a Software Process**. Based on an established software process, like IEEE Std. 1074 or ISO 12207, and CRISP-DM, the aim is to create an integrated process by carefully examining the tasks they have in common, their respective inputs and outputs and CRISP-DM's weaknesses. This study will include ideas or tasks from other processes enacted in related fields like customer relationship management (CRM) or knowledge engineering.
2. **Develop a Process Instance**. This objective aims to tailor the above integrated process to a particular software development paradigm, i.e. object orientation based on the Unified Process (RUP). The key goal is to incorporate and extend RUP techniques across the entire integrated process, stressing the software-DM relationship, the connection between software development tasks and pure DM tasks via inputs/outputs,

* Email: fsegovia@fi.upm.es

and project management. Techniques borrowed from other fields more akin to the goals of a DM project, like knowledge engineering or CRM, respectively, will be tailored and used for tasks where RUP techniques are unsuitable or insufficient for DM goals, e.g. for requirements elicitation or business modeling.

3. **Validate and Transfer Technology**. The achievement of the above goals will be entirely subject to their validation and practical application. Apart from providing experts and experience to work on the above research goals, some companies have volunteered to collaborate on pilot testing during the project's third year. Additionally, as they have a vested interest in applying the project results to their own DM projects, they will be the first to test the technology transfer to their human team.

To achieve these objectives, the project has been divided into the following four phases (the task decomposition is detailed in the original proposal):

- Phase 1. Define an integrated and generic data mining process
- Phase 2: Define an instance of an integrated and generic data mining process or methodology
- Phase 3: Integrate this process with a software development methodology
- Phase 4: Validate on pilot projects

The project chronogram divided by phases is as follows:

		Jan'05-Jun'05	Jul'05-Dec'05	Jan'06-Jun'06	Jul'06-Dec'06	Jan'07-Jun'07	Jul'07-Dec'07
Phase 1	Define an integrated and generic data mining process						
Phase 2	Define an instance of an integrated and generic data mining process or methodology						
Phase 3	Integrate this process with a software development methodology						
Phase 4	Validate on pilot projects						

Phases 1, 2 and 3 have been completed successfully, and Phase 4, Validate on pilot projects, is to kick off this year. Phase 4 will be run in conjunction with G2 Marketing Intelligence, a member of the Grey Global Group, which acquired MDS Boole, the company that has been collaborating on the project from the start. There are also possibilities of running tests with or transferring technology to Ingenuity Solutions Bhd. This company is based in Kuala Lumpur, Malaysia, and has worked on several projects with the research group.

2 Project success level

This project is based on the idea that the problems to be solved in the field of data mining nowadays are acquiring the dimensions of engineering problems. Therefore the processes to be applied should cover all the activities and tasks required in an engineering process, tasks that CRISP-DM might not be considering. The submitted project proposal was inspired by software engineering research conducted in recent years on the standardization of SE development processes by learning from other engineering disciplines and applying the experience of software developers. It intended to borrow from these ideas to establish a complete process model for DM

that would improve and add to CRISP-DM. To do this, we counted on the invaluable advice of International Software Engineering Research Network members.

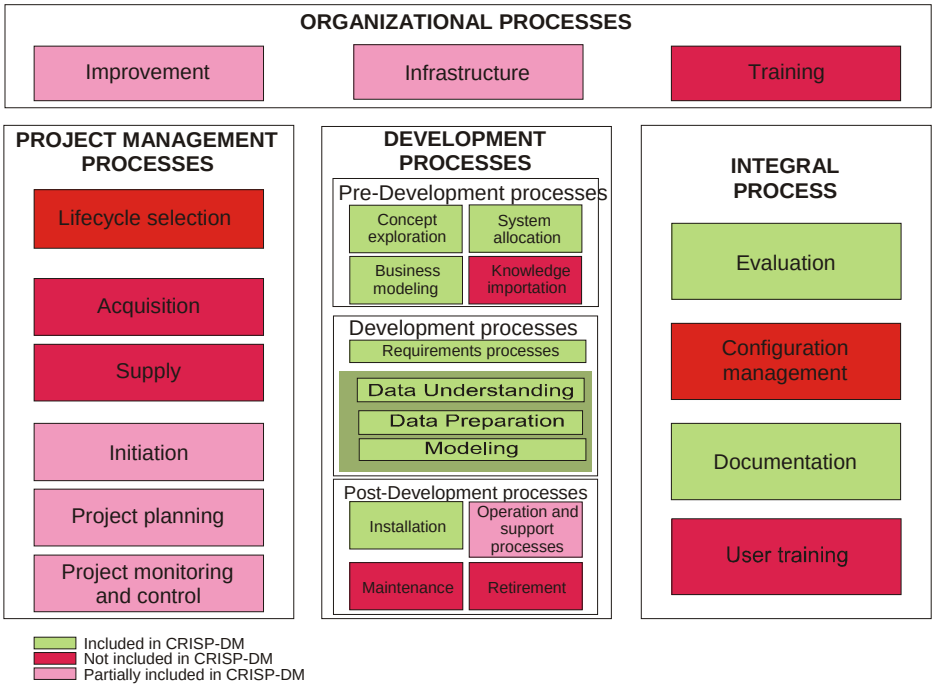
Having completed phases 1 to 3, which cover the theoretical and conceptual tasks of creating the process, the general impression is that the proposal was rather good. On the one hand, software engineering standards demonstrate that CRISP-DM did not consider or accounted only vaguely for some processes, especially management-related processes, at the same time as it dealt with other processes, such as business modeling, incorrectly. And, on the other hand, many of the processes in software engineering standards can, if necessary, be tailored to CRISP-DM without too much difficulty. These aspects are detailed in the following.

Additionally, the consortium that developed the CRISP-DM standard [1], a European industrial consortium formed in the 90s to undertake an original ESPRIT project and led by SPSS and Teradata, has been reviewing the standard's first version, 1.0, since mid-2006 and has set up a Special Interest Group to do this. Our project development team has joined this SIG. This has been useful as a preliminary review of the project results. The SIG has held three workshops, two in 2006, and a third on 18 January 2007 in London, at which our research group presented the current project results. The first workshop conclusion is that the problems that we had detected in CRISP-DM as part of the project are precisely the problems that the consortium intends to put right in the new versions. This is an important measure of the project's success because while our analysis is based on theoretical research, comparing data mining and software engineering standards and models, their analysis is a wholly practical exercise, based on the experience they and hundreds of customers all over the world have gathered from using CRISP-DM. And both analyses arrive at the same conclusions. The second finding is that while the consortium believes that some of the solutions borrowed from software engineering that we propose would be difficult to adopt (due primarily to the preconceived idea that developing software is a very structured and not very agile business, whereas data mining is quite the opposite), they have taken others into account and intend to add them either to version 2.0, scheduled for 2007, or the planned version 3.0. This is another indication of the success of the project.

2.1 Phase 1. Define and Integrate a Data Mining Process with a Software Process

To define the generic process we reviewed both processes and methodologies directly related to data mining or CRM, like KDD Process, SEMMA, Two Crows, 5 A's, 6-σ, CRM Catalyst, Data Mining Industrial Engineering, or Market ConsulTeks℠'s proposal to unify RUP and CRISP-DM, and the software engineering process model standards IEEE Std. 1074 and ISO 12207.

The first difficulty we encountered was that the data mining field makes no distinction between process model, methodology and lifecycle; they are all fused into one. This meant that the comparisons aimed at finding a correct definition were quite complicated. In the end it was decided to set up a common development process framework based on the two software engineering standards, IEEE Std. 1074 and ISO 12207, to accommodate the tasks and processes established by CRISP-DM. The figure below summarizes the result.



Clearly, CRISP-DM either fails to define or does not properly consider most of the project management processes, integral processes and organizational processes.

The second difficulty derived from data mining's conceptual fusion of process and methodology that we came across is that the tasks belonging to CRISP-DM phases are misplaced in completely different processes. For example, the first phase in CRISP-DM, business understanding, includes the following tasks:

- Determine business objectives, which is related to Business Modeling and belongs to Development Processes.
- Situation assessment, which is related to Infrastructure and belongs to Organizational Processes, is related to Requirement Processes and belongs to Development Processes, and is related to Project Planning and belongs to Project Management processes.
- Determine DM goals, which is related to Requirement Processes and belongs to Development Processes.
- Produce project plan, which is related to Project Planning and belongs to Project Management processes.

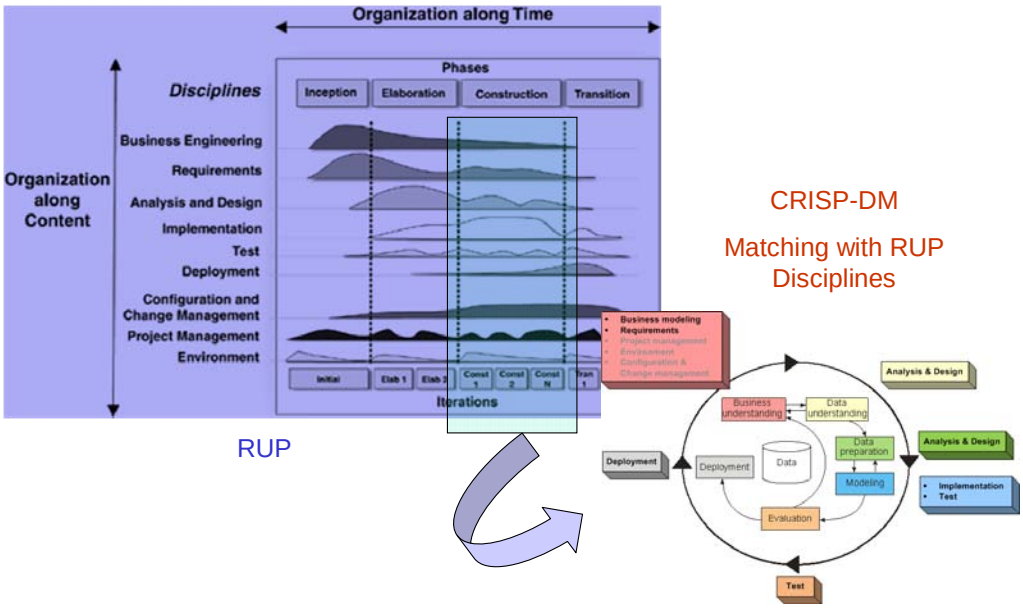
Even so, we have been able to map CRISP-DM tasks and activities more or less correctly to the generic and integrated process framework that we have defined and identify any others that CRISP-DM fails to cover either totally or partially. This served as input for the next stage of the project. As it has been created on the basis of SE standards, the framework has the advantage of providing for a future integration of data mining project activities with software development project activities.

2.2 Phase 2. Define an instance of an integrated and generic data mining process or methodology

Having defined the process, we then had to define a methodology and a lifecycle. As already mentioned, CRISP-DM is a mixture of all three things. Therefore, part of the work involved figuring out what was what and be able to put it to the best possible use.

As mentioned in the original objectives, we examined RUP [2] as an underlying methodology and lifecycle, taking into account that, for use in data mining, it would have to be suitable for both well-defined and complex processes and other short processes with non-existent requirements, such as what are known as "Exploratory Projects". Projects like these, which are very common in data mining, probably do not go beyond RUP's Inception stage. RUP appears to be well suited for data mining projects, because it covers almost all development process activities, mandates iterative and incremental development, is intended for a team size of two or more, with no upper limit, and its project management is risk oriented. Additionally, RUP is also used to build business intelligence systems using Data Warehouse, Data Marts and reporting technologies. This is evidence for the fact that it could be suitable for an environment where software is not developed. But perhaps the most important concern for our project is that requirements are expressed as Use Cases and Features, serving as guidance throughout the process. This is something that was missing from CRISP-DM.

The figure below is a comparison between RUP and CRISP-DM tasks and lifecycles.



The overall findings after tailoring RUP to CRISP-DM are as follows:

- The processes and tasks not covered by CRISP-DM can be covered and organized according to RUP disciplines.
- The CRISP-DM lifecycle is iterative but not incremental, overlooking the fact that the objectives and intensity of the activities to be carried out can vary from one project phase to another, and RUP's distribution is very well suited for this.
- The Agile Unified Process is the RUP version that should be used for most data mining projects [3].

Other very important findings are:

- RUP supports a formal specification: UML. UML is a common language for customers, end users, and developers. This is of vital importance for data mining projects where there is continual contact with the customer and a (non-existent) common language should be used.
- RUP defines roles, skills and artifacts to be used in each task, another of the points missing from CRISP-DM.

Special attention should be attached to RUP's Business Modeling (BM) task, as repeatedly pointed out at CRISP-DM workshops. CRISP-DM's key problem is perhaps how to link data mining results with the ROI (Return on Investment) of the business where it is to be used. The question is, in other words, how to link the Deployment phase with the Business Understanding phase. This is crucial for data mining projects because the average customer finds it difficult to envisage how a data mining project can benefit his or her business (this does not apply to software because everyone is very well acquainted with its benefits). We put this lack of vision is down to the fact that, unlike the RUP Business modeling phase objectives, the CRISP-DM Business Understanding (BU) objectives are not very useful for this purpose (the very name denotes the difference, understanding is one thing and modeling another). One of the BM objectives is to understand the structure and the dynamics of the organization in which a system is to be deployed (the target organization). BU accounts for this point, but, more importantly, BM also covers the following objectives not included in BU:

  - To ensure that customers, end users, and developers have a common understanding of the target organization and the project
  - To understand current problems in the target organization and identify potential improvements, which is the seed of a data mining project for that organization.
  - To understand how the new system would affect the way customers conduct their business and its potential benefits, which is the link we were looking for between the final ROI and the deployment process, a link that is missing from CRISP-DM and which it dearly needs.
  - To derive the system requirements (data mining goals) from business use cases (business goals), a procedure that is not clarified in CRISP-DM.

In view of this, we addressed the following points with the aim of defining an instance of an integrated and generic and data mining process or methodology:

- Define RUP as a methodological and lifecycle framework integrating CRISP, providing for an iterative and incremental cycle useful for both exploratory and well-defined projects, and reorder CRISP-DM tasks accordingly.
- Tailor RUP management tasks to CRISP-DM within the methodology.

- Switch the business understanding (business-specific) task for the business modeling task.
- Choose a business process framework [4] to guide the business modeling process. This framework is necessary because business modeling is more important than in software development.
- Define the business use cases to drive project development. This is a conceptual difference from a software development project which is driven by use cases defined in the Requirements phase.
- Define data mining use cases as the building blocks for Requirements.
- Define data mining paradigms (clustering, classification, dependency modeling, deviation detection, sequence analysis, etc.) as the basis for Design (the later Implementation process selects techniques for each paradigm).
- Tailor RUP artifacts and roles to all of the above, and use UML especially for Business Modeling, Requirements, and Design, creating new elements if necessary.
- Design an agile version of the process.

Additionally, we also reviewed the software Configuration and Change Management techniques, as no such techniques exist in CRISP-DM (the SAS tool alone covers some aspects). We looked at IEEE Std. 828-1998, IEEE Std. 1042-1987 and MIL-HDBK-61 Std., as well as most Revision Control Software tools, and we listed the data mining project elements that should be subject to such a control. The RUP tools that support the configuration and change management processes, such as Rational® ClearCase® and Rational® ClearQuest®, which are used in conjunction with Unified Change Management (UCM), remain to be analysed.

2.3 Integrate this process with a software development methodology

By having defined an instance of a generic data mining process or methodology integrated with the RUP process the data mining process is automatically integrated with the software development methodology. The underlying idea is that, after business modeling, a decision is taken on which business cases call for previous iterations covering the data mining project, and the results are added as requirements at the software project requirements stage.

3 Results indicators

As already mentioned, the project team has joined the CRISP-DM consortium's Update SIG, and explained and discussed the project results at the last workshop that was held in January 2007 in London.

The team is also participating in KDubiq (Knowledge Discovery in Ubiquitous Environments). KDubiq is the first Coordination Action (CA) for Ubiquitous Knowledge Discovery, funded by the European Union under IST (Information Society Technology) and FET Open (Future and Emerging Technologies) within the 6th Framework Programme. Ernestina Menasalvas coordinated one of the working groups defining the priorities for the 7th Framework Programme related to data mining processes.

The team is a member of the Spanish Data Mining and Learning Network.

The following table lists the results planned in the project proposal.

	Publication type	Submission date (project runs from Year 1 to 3)	Quantity
Planned achievements	Congress papers	Year 2-4	7-9
	JCR journal publications	Year 2-4	4
	Congress tutorials	Year 2-3	1-2
	PhD dissertations	Year 3-4	2
	Inclusion of the results in the UPM's Expert in Business Intelligence course	Year 3	1
Likely achievements depending on the extent of the results	Proposal for editing a handbook on DM processes	Year 3	1
	Proposal of a workshop for a Software & Knowledge Engineering congress	Year 3	1
	Proposal for a book on DM Processes	Year 4	1
	Proposal of a European course within KDNet	Year 3	1

There are two ongoing PhD dissertations that we expect to be completed in 2007, because this project covers most of their results:

- "Definición de un Proceso de Data Mining basado en técnicas de Ingeniería del Software" (Definition of a Data Mining Process based on Software Engineering Techniques). Author: Gonzalo Mariscal Vivas.
- "Metodología para la definición de requisitos en proyectos de Data Mining (ER-DM)" (Methodology for defining data mining project requirements). Author: José Alberto Gallardo Arancibia.

And the following MSc dissertation was defended in 2006:

- "Uso de técnicas de educción para el entendimiento del negocio" (Using elicitation techniques to understand a business). Author: María Alejandra Ochoa

As regards publications, the result of the first stage of the project has been submitted to a JCR listed journal [5]. The second phase of the project ended in June 2006. As this coincided with the CRISP-DM consortium's announcement that it was going to review the process and intended to hold workshops to work on and review proposals, however, we decided to wait until we had participated in these workshops and got a first-hand evaluation of our results. We have submitted only one paper to a congress addressing the use of ontologies in the requirements phase [6]. In the light of the positive feedback that we received from the meeting we had in London at the last workshop in January 2007, we intend to pursue all the originally planned objectives, except for "Inclusion of results in the UPM's Expert in Business Intelligence course". Finally, it was suggested to us at the last CRISP-DM workshop that we should write a "Book on DM Processes" based on our results. This would signify another achieved objective.

4 References

[1] http://www.crisp-dm.org.
[2] I. Jacobson, G. Booch, and J. Rumbaugh. The Unified Software Development Process. AddisonWesley Longman Inc., 1999.

[3] S. W. Ambler, R. Jeffries, Agile Modeling: Effective Practices for Extreme Programming and the Unified Process, John Wiley and Sons Inc., 2002
[4] A. Ostenwalder, Y. Pigneur, and C.L. Tucci, Clarifying Business Models: Origins, Present, and Future of the Concept, Communications of AIS, Volume 15, Article 1.
[5] O. Marbán, E. Menasalvas, S. Eibe, J. Segovia, Towards Data Mining Engineering: a software engineering approach, The Knowledge Engineering Review, Cambridge University Press. Under review.
[6] S. Eibe, M. Valencia, E. Menasalvas, J. Segovia, and P. Sousa, Towards autonomous search engine mining: Conceptualizing context as metadata, AWIC 2007.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

ORIPI: Organización y Recuperación de Información Paralela e Inteligente TIN2004-05920

José Ranilla [i] Coordinador
Pedro Alonso [ii] Participante
Elías F. Combarro [iii] Participante
Raquel. Cortina [iv] Participante
Irene Díaz [v] Participante
Elena Montañés [vi] Participante

Abstract

This project tries to overcome some of the bottlenecks found in Information Retrieval (IR) systems. In order to achieve this goal, the different tasks of which IR is composed will be taken into account by studying their influence on the performance of the systems as well as the best way to combine them.

One of the focuses of the project will be on one important task in the building of IR systems: dimensionality reduction. To this extent, we will explore the use of relevance measures to reduce the number of attributes (terms) used in the document representation. We will also study, from a theoretical point of view, the feature selection measures seeking to establish a relationship between the characteristics of the corpora and the measures which achieve the best results.

Also, Self-Organizing Maps (and similar types of networks) will be used to cluster similar documents together, aiming to reduce the time complexity of document search. This clustering can also, in some cases, improve the performance of the retrieval. We will study the influence of the different parameters involved in the use of these networks on the results obtained with the system. Self-Organizing Maps will also be used to automatically generate word taxonomies which will be then used to improve the performance of the retrieval system.

Independently of the methodology or technology used to build an Information Retrieval system, there exists a common problem: the high computational cost, caused by the large amount of information to process. If we consider the necessities of a real-time answer and of the efficient adding of new information, it is obvious that sequential processing of information is insufficient. The use of a parallel system provides some advantages as are the improvement of the response times, the reduction of the search operations cost and the possibility of coping with large text collections.

Therefore, we will try to increase the efficiency and effectiveness of the Information Retrieval systems with the application of different techniques.

Keywords: Information Retrieval, Feature Selection, Relevance Measures, Self-Organizing Maps, Parallel Computing

[i] email: ranilla@uniovi.es
[ii] email: palonso@uniovi.es
[iii] email: elias@aic.uniovi.es
[iv] email: raquel@uniovi.es
[v] email: sirene@aic.uniovi.es
[vi] email: elena@aic.uniovi.es

1 Project aims

1.1 Introduction

In many situations we have a large collection of text files and we want to retrieve those relevant to a certain topic. This is known as Information Retrieval (IR).

Usually, IR systems manage a huge number of documents. Consequently, some of the methods used to store, classify and retrieve the information contained in the documents take a great amount of time. In the research project **ORIPI: Organización y Recuperación de Información Paralela e Inteligente,** we seek to study different ways of improving the efficiency and effectiveness of IR systems.

In the IR environments documents are usually represented by means of the words appearing in them (what is called the *lexicon*). Thus, the number of features that have to be considered is often very high. An important topic of research is the development of techniques for the reduction of the dimensionality of these *lexica*. In this project we will focus on the study of *Relevance Measures* for this reduction.

Furthermore, in order to reduce the number of comparisons that need to be carried out, one can cluster the documents and choose a representative of each cluster. In such cases, the query has only to be matched to the representatives, and not to the whole collection. Documents from the clusters with representatives closer to the query are selected. Sometimes, this clustering also has the effect of improving the efficiency of the retrieval. The clustering technique that we will be exploring in this project consists in the use of self-organizing networks, mainly Kohonen's Self-Organizing-Maps (SOM). This technique will also allow us to automate the task of creating word taxonomies which can be used in order to help the system in the task of retrieving relevant information.

Finally, the efficiency of the IR system can be benefit from the use of parallel algorithms and architectures. Thus, we will study the best way to parallelize the methods and techniques investigated.

1.2 Main objectives

The main objectives of this project are the following:

1. To develop methods for the selection of the relevant features of a document collection in order to index, manage and classify it. Study which feature selection measures achieve the best results on a certain corpus. Obtain feature selection measures involving lexical and syntactical information as well as genetic algorithms.
2. To organize the terms appearing in a document collection by means of self-organizing structures.
3. To classify and retrieve information in an efficient way by using self-organizing structures.
4. To study the best way to parallelize the processes involved in the construction of the information retrieval system. Implement parallel algorithms that increase the efficiency of the system.
5. To particularize the developed methods to the Spanish language.

1.3 Time schedule

Tasks	First year	Second year	Third year
T1.1	x|x| | | | | | | | | |	| | | | | | | | | | |	| | | | | | | | | | |
T1.2	| |x|x| | | | | | | |	| | | | | | | | | | |	| | | | | | | | | | |
T1.3	| | | |x|x|x| | | | |	| | | | | | | | | | |	| | | | | | | | | | |
T2.1	| | | | | | |x|x|x| |	| | | | | | | | | | |	| | | | | | | | | | |
T2.2	| | | | | | | |x|x|x|x	x|x| | | | | | | | |	| | | | | | | | | | |
T2.3	| | | | | | | | |x|x|x	x|x|x| | | | | | | | |	| | | | | | | | | | |
T2.4	| | | | | | | | |x|x|x	x|x|x|x|x|x| | | | |	| | | | | | | | | | |
T3.1	| | | | | | | | | | |x	x|x|x|x|x| | | | | |	| | | | | | | | | | |
T3.2	| | | | | | | | | | |	|x|x|x|x|x| | | | |	| | | | | | | | | | |
T3.3	| | | | | | | | | | |	| | |x|x|x|x|x| | |	| | | | | | | | | | |
T3.4	| | | | | | | | | | |	| | | | | |x|x|x| | |	| | | | | | | | | | |
T3.5	| | | | | | | | | | |	| | | | | | | |x|x|x	x|x|x| | | | | | | |
T4	| | | | | | | | | | |	| | | | |x|x|x|x|x|x|x	x|x|x|x|x|x|x| | | |
T5	| | | | | | | | | | |	| | | | | | |x|x|x|x|x	x|x|x|x|x|x|x| | | |
T6	| | | | | | | | | | |	x|x|x|x|x|x|x|x|x|x|x|x	x|x|x|x|x|x|x|x|x|x|x
C	x|x|x|x|x|x|x|x|x|x|x|x	x|x|x|x|x|x|x|x|x|x|x|x	x|x|x|x|x|x|x|x|x|x|x

Table 1: Time diagram and task distribution

The activities and tasks that will be conducted in order to achieve the objectives enumerated in the previous Section are:

- T1.1 General specifications
- T1.2 Identification, selection and analysis of document collections
- T1.3 Corpora processing
- T2.1 Study of the optimal measure for a given corpus
- T2.2 Use of genetic algorithms to obtain better relevance measures
- T2.3 Attributes characterization by means of implicit information
- T2.4 Computational study of feature selection methods
- T3.1 Construction of term hierarchies by self-organizing maps
- T3.2 Classification of document collections by self-organizing maps
- T3.3 Retrieval of documents from clustered collections
- T3.4 Computational study of self-organizing maps

- T3.5 Evaluation of parallelization alternatives
- T4 Implementation
- T5 Experimentation
- T6 Dissemination of results
- C Coordination

Table 1 below shows the time diagram for the coordination of the different tasks.

2 Project progress and achievements

2.1 Task 1

Document collections suitable for our study and widely used in the literature were successfully selected and processed. For clustering and retrieval purposes the collections *adi, cacm, cisi, cran, med* and *time* were selected. For text categorization and feature selection experiments we selected the *Reuters-21578*, *Ohsumed* and *20-newsgroups* collections.

2.1 Task 2

In the search for an optimal filtering measure for document collections we developed several new families of filtering measures. We first studied the use of concepts from the field of Machine Learning for the task of feature selection. This study concluded with the definition of several measures (D, D_{ir}, L L_{ir}, IL, IL_{ir}) which obtained, in most situations, better results than state-of-the-art measures. Later, we generalized this study and we were able to define two new families of measures (*linear* and *angular measures*) which offer even better performance. We then studied how to determine the measure of these families which is more suitable for each corpus by means of developing methods of automatically selecting the value of the parameters which offers the highest performance. This study leads to partial successes in the identification of the measures.

As another subtask included in the study of feature selection methods for document collections, we explored the use of lexical and syntactical information in order to determine which terms are more relevant for classification and indexing purposes. To this extent, we contacted experts in linguistics from the Spanish Philology Department of the University of Oviedo and developed tools to carry out a supervised selection of relevant features from tagged collections. We still have some work in progress in this topic, but prelimary results seems to show that the human effort needed for this task may not pay off in better classification results.

We also conducted experiments with genetic algorithms aiming to obtain better feature selection methods. However, the results obtained were not satisfactory. The genetic algorithms were usually slow and prone to overfitting while achieving results below those of the feature selection measures.

Different possibilities for the parallelization of all these methods were studied. Those offering the best performance in terms of efficiency and scalability were implemented.

2.3 Task 3

Based on the results of a prelimary study we decide to postpone the use of self-organizing maps for automatic generation of word taxonomies until some of the other tasks had been completed. Our intention is to use relevance measures in order to characterize the words for the clustering. Consequently, it is convenient to fully develop those measures before approaching this subtask.

However, the clustering and retrieval of document collections with SOMs was independent of the previous subtask and was successfully tackled. After a preliminary study conducted in order to determine the fittest kind of networks and the optimal parameters we implemented an IR system prototype which was subsequently used to organize the documents from the Gijón Town Council.

Our studies revealed a high dependency of the overall performance on the selection of the training parameters (especially the network size and radius) and also showed that this kind of maps can be successfully combined with dimension reduction techniques such as relevance measures, Latent Semantic Indexing (LSI) and Random Projection (RP).

The process of network training and the use of dimensionality reduction techniques (especially LSI and RP) are time consuming. In order to overcome this problem we opted for the use of high performance libraries for the linear algebra computations involved in the processes. The implementation of the prototypes was carried by means of highly efficient and scalable parallel algorithms.

3 Performance Indicators

3.1 Trainees

Mrs. Raquel Cortina Parajón, Professor at the University of Oviedo shall defend her doctoral thesis this year.

3.2 Summary of publications

Table 2: Summary of relevant publications

Type	Number	References
Papers in SCI journals	8	[1][2][3][7][9][10][12][14]
International Conference papers	5	[4][5][6][11][13]
Book chapters	2	[8][13]

[1] Alonso P., Cortina R., Díaz I., Ranilla J. Blocking Neville Elimination Algorithm for Exploiting Cache Memories. To be published in Journal of Computational and Applied Mathematics, 2007.

[2] Alonso P., Cortina R., Díaz I., Ranilla J. Scalability of Neville elimination using checkerboard partitioning. To be published in International Journal of Computer Mathematics, 2007.

[3] Alonso P., Cortina R., Díaz I., Ranilla J. Analyzing Scalability of Neville Elimination. Journal of Mathematical Chemistry, 40(1): 49-61, 2006.

[4] Combarro E.F., Montañés E., Díaz I., Cortina R., Alonso P., Ranilla J. NoW Architectures, Dimensionality Reduction and Self-Organizing Maps for Information Retrieval. In Proceedings of the 2006 International Conference of Computational Methods in Sciences and Engineering, 7: 110-113, Chania – Crete, Greece, 2006.

[5] Combarro E.F., Montañés E., Díaz I., Cortina R., Alonso P., Ranilla J. Dimension Reduction Techniques for Information Retrieval using Self-Organizing Maps. In Proceedings of the 11th Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2006), 1900-1906, Paris, France, 2006.

[6] Combarro E.F., Montañés E., Ranilla J., Díaz I. Angular Measures for Feature Selection in Text Categorization. In Proceedings of the 21st Annual ACM Symposium on Applied Computing, 1: 826-830 (SAC 2006), Dijon, France, 2006.

[7] Combarro E.F., Montañés E., Díaz I., Ranilla J., Mones R. Introducing a Family of Linear Measures for Feature Selection in Text Categorization. IEEE Transactions on Knowledge and Data Engineering, 17(9): 1223-1232, 2005.

[8] Combarro E.F., Ranilla J., Mones R., Vázquez N., Díaz I., Montañés E. Unsupervised Learning in Information Retrieval with NoW Architectures. Lecture Notes in Computer Science 3643: 179-188, Springer-Verlag, Germany, 2005.

[9] Miranda P., Combarro, E.F. On the structure of some families of fuzzy measures. To be published in IEEE Transactions on Fuzzy Systems, 2007.

[10] Montañés E., Quevedo J.R., Combarro E.F., Díaz I., Ranilla J. A Hybrid Feature Selection Method for Text Categorization. To be published in International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2007.

[11] Montañés E., Combarro E.F., Ranilla J., Díaz I. Finding Optimal Linear Measures for Feature Selection in Text Categorization. In Proceedings of the 21st Annual ACM Symposium on Applied Computing (SAC 2006), 1: 861-862, Dijon, France, 2006.

[12] Montañés E., Díaz I., Ranilla J., Combarro E.F., Fernández J. Scoring and Selecting Terms for Text Categorization. IEEE Intelligent Systems, 20(3): 40-47, 2005.

[13] Montañés E., Combarro E.F., Díaz I., Ranilla J. Towards Automatic and Optimal Filtering Levels for Feature Selection in Text Categorization. In Proceedings of the 6th International Symposium on Intelligent Data Analysis (IDA 2005), LNCS 3646: 239-248, Madrid, Spain, 2005.

[14] Ranilla J., Rodríguez-Muñiz L.J. A heuristic approach for learning rules out of fuzzy databases. To be published in IEEE Intelligent Systems, 2007.

3.3 Teaching

Part of the theoretical results has been used for teaching on the course *Análisis de Patrones Mediante Técnicas de Inteligencia Artificial*, on the interdepartmental doctoral program *Métodos Avanzados en Tecnologías de la Información y las Comunicaciones*.

3.4 Dissemination of results

According to the task schedule, web pages have been constructed with information about the results of the project and this can be accessed at the following addresses: http://www.aic.uniovi.es.

4 References

- Alonso P., Cortina R., Díaz I., Ranilla J. Blocking Neville Elimination Algorithm for Exploiting Cache Memories. To be published in Journal of Computational and Applied Mathematics, 2007.
- Alonso P., Cortina R., Díaz I., Ranilla J. Scalability of Neville elimination using checkerboard partitioning. To be published in International Journal of Computer Mathematics, 2007.
- Alonso P., Cortina R., Díaz I., Ranilla J. Analyzing Scalability of Neville Elimination. Journal of Mathematical Chemistry, 40(1): 49-61, 2006.
- Alonso P., Cortina R., Díaz I., Ranilla J. Scalability of Neville elimination using checkerboard partitioning. In Proceedings of the 6th International Conference on Computational and Mathematical Methods in Science and Engineering (CMMSE 2006), Madrid, Spain, 2006.
- Alonso P., Cortina R., Díaz I., Ranilla J. Analyzing Scalability of Neville Elimination. In Proceedings of the 5th International Conference on Computational and Mathematical Methods in Science and Engineering (CMMSE 2005), Alicante, Spain, 2005.
- Alonso P., Cortina R., Díaz I., Ranilla J. Blocking Neville Elimination Algorithm for Exploiting Cache Memories. In Proceedings of the 2005 International Conference of Computational Methods in Sciences and Engineering (ICCMSE 2005), Lecture Series on Computer and Computational Science, 4(B): 1770-1773, Korinthos, Greece, 2005.
- Combarro E.F., Montañés E., Díaz I., Cortina R., Alonso P., Ranilla J. NoW Architectures, Dimensionality Reduction and Self-Organizing Maps for Information Retrieval. In Proceedings of the 2006 International Conference of Computational Methods in Sciences and Engineering (ICCMSE 2006), 7: 110-113, Chania – Crete, Greece, 2006.
- Combarro E.F., Montañés E., Díaz I., Cortina R., Alonso P., Ranilla J. Dimension Reduction Techniques for Information Retrieval using Self-Organizing Maps. In Proceedings of the 11th Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2006), 1900-1906, Paris, France, 2006.
- Combarro E.F., Montañés E., Ranilla J., Díaz I. Angular Measures for Feature Selection in Text Categorization. In Proceedings of the 21st Annual ACM Symposium on Applied Computing (SAC 2006), 1: 826-830, Dijon, France, 2006.
- Combarro E.F., Montañés E., Díaz I., Ranilla J., Mones R. Introducing a Family of Linear Measures for Feature Selection in Text Categorization. IEEE Transactions on Knowledge and Data Engineering, 17(9): 1223-1232, 2005.

- Combarro E.F., Ranilla J., Mones R., Vázquez N., Díaz I., Montañés E. Unsupervised Learning in Information Retrieval with NoW Architectures. Lecture Notes in Computer Science 3643, pp. 179-188, Springer-Verlag, Germany, 2005.
- Díaz I., Montañés E., Ranilla J., Combarro E.F. Language Sensitive Term Selection. Submitted to a SCI journal.
- Díaz I., Lloréns J., Génova G., Fuentes J.M. Generating Domain Representations using a Relationships Model. Information Systems, 30(1): 1-19 (2005)
- Díaz I., Ranilla J., Montañés E., Fernández J., Combarro E.F. Improving Performance of Text Categorisation by Combining Filtering and Support Vector. Journal of the American Society for Information Science and Technology (JASIST), 55(7): 579-592 (2004)
- Miranda P., Combarro, E.F. On the structure of some families of fuzzy measures. To be published in IEEE Transactions on Fuzzy Systems, 2007.
- Mones R., Vázquez N., Díaz I., Ranilla J., Combarro E.F., Montañés E. An Open Low-Cost Parallel Architecture for Unsupervised Learning in Information Retrieval. In Proceedings of the 10th International Conference on Computer Aided Systems Theory (Eurocast 2005), Las Palmas de Gran Canaria, Spain, 2005.
- Montañés E., Alonso P., Combarro E.F., Díaz I., Cortina R., Ranilla J. Feature Selection for Text Categorization using Laplace and angular measures. Submitted to a SCI journal.
- Montañés E., Quevedo J.R., Combarro E.F., Díaz I., Ranilla J. A Hybrid Feature Selection Method for Text Categorization. To be published in International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2007.
- Montañés E., Combarro E.F., Ranilla J., Díaz I. Finding Optimal Linear Measures for Feature Selection in Text Categorization. In Proceedings of the 21st Annual ACM Symposium on Applied Computing (SAC 2006), 1: 861-862, Dijon, France, 2006.
- Montañés E., Díaz I., Ranilla J., Combarro E.F., Fernández J. Scoring and Selecting Terms for Text Categorization. IEEE Intelligent Systems, 20(3): 40-47, 2005.
- Montañés E., Combarro E.F., Díaz I., Ranilla J. Towards Automatic and Optimal Filtering Levels for Feature Selection in Text Categorization. In Proceedings of the 6th International Symposium on Intelligent Data Analysis (IDA 2005), LNCS 3646: 239-248, Madrid, Spain, 2005.
- Ranilla J., Cortina R., Alonso P., Montañés E., Díaz I., Combarro E.F. PirSOM: A Parallel Information Retrieval System. Submitted to a SCI journal.
- Ranilla J., Rodríguez-Muñiz L.J. A heuristic approach for learning rules out of fuzzy databases. To be published in IEEE Intelligent Systems, 2007.
- Ranilla J, Luaces O, Bahamonde A. A heuristic for learning decision trees and pruning them into classification rules. Artificial Intelligence Communication, 16(2): 71-87, 2003.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Adaptive Learning of Probabilistic Graphical Models in Data Mining and Automatic Personalisation TIN2004-06204-C03

Antonio Salmerón *
Universidad de Almería

Andrés Cano †
Universidad de Granada

José A. Gámez ‡
Universidad de Castilla-La Mancha

Abstract

In this project we intend to develop the methodology of adaptive learning of probabilistic graphical models, specially Bayesian networks, focused on its application to data mining and automatic personalisation. The background software is the Elvira platform [22], in which development the groups in this project took part, through projects TIC97-1137-C04 and TIC2001-2973-C05.

The most important part of the project is devoted to the development of applications, highly based on the Elvira platform. Each sub-project is responsible of two applications. Sub-project 1 (Almería) is developing a personalised academic advisor for students, based on the construction of a Bayesian network from the student database of the University of Almería. Also, it has implemented an application for bookmarks personalisation in web browsers. Sub-project 2 (Albacete) works in the implementation of an advisor for academic managers based on a Bayesian network obtained from the data provided by the University of Almería. Furthermore, it is developing a system for the classification of e-mail into folders. Sub-project 3 (Granada) is the responsible of the implementation of a system for personalising the result of a web search based on the user's preferences. Also, it works in the construction of a system for detecting urgent e-mail, specially useful when only a reduced number of messages can be read.

Keywords: Probabilistic graphical models, inference, supervised classification, unsupervised classification, applications to e-mail, information retrieval, web search, web browsers, automatic personalisation, data mining.

*Email: antonio.salmeron@ual.es
†Email: acu@decsai.ugr.es.es
‡Email: jgamez@info-ab.uclm.es

1 Objectives

We distinguish two types of objectives:

A. Objectives of methodological development.

B. Objectives of construction of applications.

Regarding methodological development, in the project application we established the objectives that we describe below, where Mn means that the planned deadline for a precise objective is the n-th month after the project start:

A1 *Data preprocessing.* Any knowledge discovery task requires a process of data preprocessing. This term refers to a process of transformation of the initial database in order to prepare it for our goals related to learning or inference. We proposed to study the problem of discretisation and establishment of hierarchies of concepts (M16), develop methodologies for variable selection (M15), as well as variable construction and discovery of new variables (M16).

A2 *Supervised classification.* In this area we planned to investigate the construction of classification models incorporating continuous variables, such as Gaussian models and MTEs (mixtures of truncated exponentials). Also, we consider the extension to models where the class variable is continuous (i.e. regression models). The deadline for this objective was M24.

A3 *Unsupervised classification.* The aim was to explore classification problems where the class variable is unknown (i.e. clustering problems), incorporating the feature of adaptation to the arrival of new data. The deadline was M15.

A4 *Learning of Bayesian networks.* Within this objective, we planned tasks consisting of learning in domains of high dimensionality, learning from data streams and incremental learning. The deadline was M28.

With respect to the construction of applications, we defined the following objectives:

B1 *Classification of e-mails into folders.* The goal of this application is to classify any incoming mail in the appropriate folder among those defined by the user. The classification model is constructed from a set of e-mails previously classified by the user. The folders may be structured as a tree. The application must be able to handle misclassification scenarios, by offering the user the possibility of correcting the initially proposed folder. The deadline for this objective is M36.

B2 *Classification of urgent e-mail.* The main goal of this application is to sort the incoming e-mails in terms of their degree of urgency. Obviously, the degree of urgency depends on the user, and thus the underlying model must be learnt from the user opinion, as well as from external information (as the date and time). This application may help a user who receives important amounts of messages, to attend them in a more productive way. The deadline for this objective is M36.

B3 *Bookmarks personalisation.* This application intends to increase the functionality of web browsers by automatically classifying the web sites stored by the user. The classification will take place within a set of folders, possibly hierarchically structured, an taking into account the preferences of the user as well as external information, such as the classification of the site provided by Google's tree of categories. The deadline for this objective is M36.

B4 *Personalised web search.* This application will improve web search engines by incorporating user preferences in connection with the way in which the search results are displayed. For instance, Google sort them using the page rank. Our aim is to incorporate the user opinion as its degree of satisfaction with the displayed results. The deadline for this objective is M36.

B5 *Personalised academic advisor for University students.* Spanish degree programs usually include an important amount of subjects among which the student must choose in order to create its own profile within the degree program. The aim of this application is to provide the user (the student) the list of subjects that he/she should choose in order to meet some criteria as maximising the probability of success, adaptation to the desired professional profile, etc. The model will be constructed from the administration databases provided by the Unit for Data Coordination of the University of Almería. The deadline for this task is M36.

B6 *Academic advisor for lecturers and administration.* This objective is related to B4 in the sense that the same data set will be used. The system will provide advise to lecturers related to the kind of students a lecturer can expect to find when approaching a new subject, classifications of the studentship in terms of their previous background, etc. With respect to administration, the system may help to establish pre-requisites for some given subjects, clustering of similar or related subjects, etc. The deadline for this objective is M36.

2 Status of project development

2.1 General considerations

In general we consider that the project is progressing in a satisfactory way, even though some of the objectives suffer a delay with respect to the initial plan. However, it must be taken into account that we made the schedule assuming that the the Ministry would provide funds to incorporate three research assistants to the project staff (one per group). Unfortunately, finally we only got funds for one research assistant (assigned to the group in Almería). Even though the human resources are lower than expected, we decided not to reduce the objectives of the project.

In connection with this, an added problem is the difficulty to find a research assistant to enroll in the project. The first candidate that was incorporated to the Almería group left the project after one year of work, due to the fact that he found a new position with better salary and more projection. Afterwards, the paperwork and bureaucracy required for opening the position again motivated that, during around 6 months our staff was also beneath expected.

2.2 Achievements regarding methodological tasks

A1 *Data preprocessing.*

With respect to discretisation, in this project we were more interested in the problem of grouping of cases for categorical variables with a high number of cases, an equivalent problem to that of discretisation of continuous variables. Thus, we started by implementing a well known method (KEX) in Elvira whose usefulness as a preprocessing tool was satisfactorily tried in [24]. Then, we have develop two semi-naive Bayes classifiers that search for dependent attributes using different filter approaches or the imprecise Dirichlet model [6], and join them into an only attribute. The algorithms also make use of a grouping procedure after two variables are merged in order to reduce the number of cases of the new variables.

Regarding hierarchies of concepts, our works are particularised to the classification of bookmarks into a folder structure, in connection with task B3.

With respect to variable selection, we have approached the problem from two different points of view: (1) through the construction of a general Bayesian network where the set of relevant features is taken as the Markov Blanket of the class variable. This approach has been tested in connection with the classification of satellite oceanic images [44]; and (2) by proposing a filter+wrapper approach in which the variables are first ranked by using a (fast) filter method, and then a wrapper approach is considered but taking the variables following the ranking instead of using all the variables as candidates at each stage. This second approach has been applied in combination with feature construction in breeding value classification [24, 25].

With regard to the construction of new variables, the previous algorithms [6] are methods to obtain new variables with higher predictive accuracy than the original variables. This task is still in a rather preliminary phase, and we are now investing effort on it in order to advance as much as possible. With the same goal, but in a concrete task as it is *e-mail foldering*, we have tried to construct new terms by integrating different terms by developing methods inspired in the X-of-N methodology. Concretely we have designed (wrapper) methods based on evolutionary computation to look for good constructed attributes [14] and based on greedy search [15]. These algorithms have been tested by using the well know ENROM e-mail messages data set.

A2 *Supervised classification.* The first group of achievements within this goal comprise supervised classification when the class variable is continuous (i.e. prediction models). We have developed three Gaussian models based on three different network structures but with multivariate Gaussian joint distribution in all the cases. The results have been applied to the prediction of the breeding value in Manchego sheep [34]. In order to approach scenarios in which continuous and discrete feature variables take part simultaneously, we have developed a prediction model based on the Naive Bayes structure where the joint distribution is approximated as a Mixture of Truncated Exponentials (MTE), testing the performance of the models in problems related to the prediction of performance indicators in higher education [41, 42].

Also, some effort has been dedicated to the problem of supervised classification when both the class and the attribute variables are discrete [3, 7]. We have developed new

classification methods using classification trees. It makes use of the imprecise Dirichlet model to estimate the probabilities of belonging to the respective classes.

Regarding the induction of the structure of the Bayesian network when the Naive Bayes restriction is not imposed, but still continuous and discrete variables are simultaneously considered, we have developed an algorithm for MTE models based on a hill-climbing and simulated annealing guided search methods [46].

For the estimation of conditional MTE distributions, we have refined previously existing methods by means of the approximation of the sample histogram using kernel functions [47]. The behaviour of the estimation algorithm for conditional MTE distributions is tested in a financial setting in [20].

In all the cases, such as it was contemplated in the project application, the software implementing all the above cited methods has been included in the Elvira system [22].

Once the model has been constructed, the classification is carried out through an inference process in which the distribution of the class variable is computed for the observed values of the feature variables. We have developed an approximate algorithm for carrying out this task when the model is of type MTE, which allows dealing with scenarios defined by a large amount of variables. The method is based on the so-called Penniless technique [17], which is based on adjusting the size of the data structures used to represent the density functions.

When all the variables in the network are discrete, the inference process relative to the classification task can be carried out more efficiently, but still there may be situations in which the complexity of this operations is too high. In order to approach these events, we have developed an approximate inference algorithm based on dynamic importance sampling that controls the level of approximation through the factorisation of the probability trees that represent the sampling distributions [37].

On the other hand, we have investigated a new family of Bayesian classifiers, in which the probabilistic graphical model is not a Bayesian network but a *dependency* network, that is, a network in which directed cycles are allowed. The main advantage of this model is that the parents of each variable can be learnt in parallel and so more efficient algorithms can be designed. In these two years we have explored the following scenarios: learning dependency networks classifiers from data by using score+search, independence tests and evolutionary algorithms [29, 30, 32].

Attending the explanations above, we can conclude that the works in this task progress according to the initial schedule.

A3 *Unsupervised classification.* In this task we have focused on two different problems.

Design of a new segmentation algorithm which receives as input a dataset with only discrete variables and outputs a full partitioning of the representation space by using a tree-shaped probabilistic structure [27]. Two additional advantages of this structure (still to be explored) are its use as an approximate factorisation of a probability distribution and its use for approximate inference.

On the other hand, we have developed an unsupervised clustering algorithm, which means that the class variable is hidden, which is able to deal with discrete and continuous

features simultaneously [35]. Again, this is possible due to the use of MTEs. Though still some points of the methods must be refined, the results are competitive considering well-known algorithms.

A4 *Learning of Bayesian networks.*

Within the framework of this objective, we have built a system based on Bayesian networks that is able to adapt itself to the user profile and that learns using a very large number of variables. The system is actually a computer chess program, which adapts its game strategy to the user's style, and dynamically re-learns its search heuristic according to several parameters that may change during the game. We think that chess may serve as a benchmark for testing the appropriateness of Bayesian networks for constructing adaptive systems. The system is called BayesChess and is described in [23].

Regarding learning in highly dimensional spaces, we are currently developing methods based on factorisation of probability trees, in which the learnt structure is a join tree rather than a Bayesian network. We have already developed the necessary theory behind approximate factorisation [36]. Also in this issue we have developed CHC (Constrained Hill Climbing) [33], an algorithm that improves the performance of classical Hill Climbing algorithm by constraining during the search the amount of comparisons to be done. The way in which the algorithm reduces the number of comparisons is by constraining the set of candidate parents for each variable, but with the advantage that this knowledge is acquired (and refined) during the search process.

A new methodology to treat several types of structural restrictions within algorithms for learning Bayesian networks have been proposed [16]. These restrictions may encode expert knowledge in a given domain, in such a way that a Bayesian network representing that domain should satisfy them. Three types of restrictions are formally defined: existence of arcs and/or edges, absence of arcs and/or edges, and ordering restrictions. It is analysed how the restrictions can be managed within Bayesian network learning algorithms based on both score + search and conditional independence paradigms, and then we particularise to two classical learning algorithms: a local search algorithm guided by a scoring function, with the operator of arc addition, arc removal and arc reversal, and the PC algorithm.

Some modifications on the PC basic learning algorithm have been proposed [4]. The variations are: to determine minimum size cut sets between two nodes to study the deletion of a link, to make statistical decisions taking into account a Bayesian score instead of a classical Chi-square test, to study the refinement of the learnt network by a greedy optimisation of a Bayesian score, and to solve link ambiguities taking into account a measure of their strength. It is shown that some of these modifications can improve PC performance, depending of the learning task: discovering the causal structure or approximating the joint probability distribution for the problem variables.

B1 *E-mail foldering.* This task refers to the problem of classifying incoming mail into the folders previously created by an user. The first stage in this task has been the study of text mining techniques, and the creation of a good state-of-the-art [45]. The way in which we have approached this problem has been by means of constructing highly discriminant new attributes from the available ones. The attributes constructed are inspired in the

X-out-of-N methodology, but we have modified the definition to a more suitable one in order to deal with numerical variables that represent word frequencies in the documents. With respect to the algorithms used to look for good constructed attributes we have designed algorithms based on evolutionary [14] and [15], and have been tested over a classical (though recent) benchmark: the ENROM dataset.

B2 *Classification of urgent e-mail.* There are not any available database for e-mails containing the degree of urgency for each mail and each user. Therefore we are developing a prototype program to build a database from the input e-mail of a given user. The program makes use of Elvira, Lucene and Weka. We plan to make a database with the list of e-mails of our team in order to prove different classification algorithms.

B3 *Bookmarks personalisation.* In this task we have addressed the following problems:

We have implemented a plug-in for Mozilla Firefox that automatically classifies the bookmarks saved by the user. The basis for the classification is the tree of categories used by Google. We are currently improving the system by incorporating the used preferences.

A typical problem in bookmarks management is that users usually limit themselves to add new bookmarks to its favourites in a single folder, but without creating a structure that makes easier its retrieval. With the aim to alleviate this problem we have designed a tool [43] that used hierarchical clustering to create a tree-shaped structure in which the bookmarks are placed depending on its content, which is analyised by using information retrieval classical techniques. Moreover, the tool allow the user to navigate over the constructed hierarchy and to fuse nodes if necessary.

Currently we are working on a different task related with bookmarks. Thus, although one user can store a large amount of bookmarks, only a few of them (less than 10) can be place in the *personal toolbar folder* of a browser, that is, those that are always shown as buttons in a bar of the browser. Of course, this bar is of great utility because we can achieve a new web page just by one mouse click. Now, we are working in the automatic modification of the content of this bar by inducing a classifier that takes into account not only the pages previously visited but also the day of the week, the month, ...

B4 *Personalised web search.* We have analysed factors involved in the relevance assessment of web pages [38, 39] for specific users. We study more than 150 web features (such as textual, visual/layout, structural) in order to investigate, using a machine learning approach, which ones are the most informative about the relevance of web pages. The results showed that while the textual and visual features were useful for the relevance prediction in a topic-independent condition, a range of features can be effective when the task knowledge was available.

We have also implemented a plug-in for Mozilla Firefox that recovers the user preferences when he makes a search using Google. This allows to build a local database for each user that will be used to reorder the list of links that retrieve Google in future searches.

B5 *Personalised academic advisor for University students.* We have completed the data preprocessing step required by this task, consisting of databases of performance indicators for all the subjects offered at the University of Almería since course 1993/1994. A thorough

analysis of the data is carried out in [40, 42]. We are currently implementing the web-interface to the final application.

B6 *Academic advisor for lecturers and administration.* This objective progresses in a parallel way to the former one although it accumulates some delay because the availability of the data from the UCLM. Thus, the initial dataset was received a few months ago, but an enrichment of it was required, that is, addition of extra variables as the professional category of the teacher, the type of subject, ... This new dataset has been just received in this month.

2.3 Additional work

Though not in the original plan, we have carried out some research that was considered necessary or useful for the project development. We want to mention the following contributions:

- *Evaluation of influence diagrams.* Influence diagrams are powerful tools for representing and solving complex decision-making problems. The evaluation may require an enormous computational effort and this is a primary issue when processing real-works models. We have proposed [18] an approximate inference algorithm to deal with very large models, when an exact solution is infeasible. It is an anytime algorithm that returns approximate solutions which are increasingly refined as computation progresses.

- *Inference in credal networks.* A credal network is a graphical structure which is similar to a Bayesian network, where imprecise probabilities are used to represent quantitative knowledge. Credal networks can be applied for example to supervised learning problems. We have proposed two new algorithms [19] for inference in credal networks. These algorithms enable probability intervals to be obtained for the states of a given query variable.

- *Properties of uncertainty measures.* We have studied properties of uncertainty measures for imprecise probabilities [5, 8, 9, 10, 11]. These works have been applied in posterior papers to problems of supervised learning and selection of variables.

- *Applications to genome analysis of linkage disequilibrium (LD).* BMapBuilder builds maps of pairwise linkage disequilibrium (LD) in either two or three dimensions. The optimised resolution allows for graphical display of LD for single nucleotide polymorphics (SNPs) in a whole chromosome. We have used a Bayesian estimator of LD [1, 2] that greatly reduces the effect of ascertainment bias in the inference of LD decay.

- *Applications of Decision support system in soaring site recommendation.* We have developed a recommendation system for the problem of adventure tourism by means of the integration of a data warehouse and a Decision Support System in order to help retrieve data from different databases and information sources and analyse them in order to provide useful and explicit information [12, 13].

- *Applications to evolutionary computation.* In this point new estimation of distribution algorithms have been designed by using dependency networks as the probabilistic graphical model that guide the search [31].

3 Results indicators

The following indicators support, in our opinion, the results of the project:

- The number of publications that have generated by the project works. The list of references include the publications that have received financial support from the project funds. The list includes the most reputed journal and conferences in the field. Also, several papers are currently under review, and for sure, new striking publications will arise in the next future.
- Two PhD. theses (Julia Flores -UCLM- and María Morales -UAL-) have been approved during the first part of the project, and other theses are about to be submitted.

References

[1] M. M. Abad-Grau, R. Montes, P. Sebastiani. Building chromosome-wide LD maps. *Bioinformatics*, Vol. 22, no 16, pp. 1933-1934, 2006.

[2] M. M. Abad-Grau, P. Sebastiani. Bayesian correction for SNP ascertainment bias. *Lectures Notes in Computer Science* Vol 3885, Modeling Decisions for Artificial Intelligence (Conference on Modeling decisions for Artificial Intelligence, Tarragona, Spain, 3-5 Abril), pp. 262-273, 2006.

[3] J. Abellán. Completing an Uncertainty Criterion of Classification. *Mathware and Soft Computing*, Vol XII, pp. 83-95, 2005.

[4] J. Abellán, M. Gómez-Olmedo and S. Moral. Some Variations on the PC Algorithm. *Proceedings of the Third European Workshop on Probabilistic Graphical Models*, Prague, Czech Republic, pp 1-8, September, 2006.

[5] J. Abellán. Uncertainty measures on probability intervals from the imprecise Dirichlet model. *International Journal of General Systems*, Vol. 35, No 5, pp. 509-528, October 2006.

[6] J. Abellán. Application of uncertainty measures on credal sets on the naive Bayesian classifier. *International Journal of General Systems*, Vol. 35, No. 6, pp 675-686, 2006.

[7] J. Abellán, S. Moral, M. Gómez and A. R. Masegosa. Varying Parameter in Classification Based on Imprecise Probabilities. *Soft Methods for Integrated Uncertainty Modelling. Advances in Soft Computing.* Vol 6, pp 231-239. Springer-Verlag, ISBN: 3-540-34776-3, 2006.

[8] J. Abellán, G. J. Klir. Additivity of uncertainty measures on credal sets. *International Journal of General Systems*, Vol 34, No 6, pp.. 691-713, December 2005.

[9] J. Abellán, G. J. Klir, S. Moral. Disaggregated total uncertainty measure for credal sets. *International Journal of General Systems*, Vol 35, No 1, February 2006, pp. 29-44, 2006.

[10] J. Abellán, M. Gómez. Measures of divergence on credal sets. *Fuzzy sets and systems*, Vo. 157, pp 1514-1531, 2006.

[11] J. Abellán, S. Moral. An algorithm to compute the upper entropy for order-2 capacities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. Vol. 14, no 2 pp. 141-154, 2006.

[12] F. Araque, A. Salguero and M.M. Abad-Grau. Application of data warehouse and Decision Support System in soaring site recommendation. *Proceedings of the the Thirteenth Conference on Information and Communication Technologies in Tourism*, ENTER 2006, M. Hitz, M. Sigala and J. Murphy (eds.), Wien-NewYork: Springer-Verlag Computer Science, pp. 308-319, 2006.

[13] F. Araque, A. Salguero, M. M. Abad-Grau. Supporting decision making as a plus in adventure tourism. Submitted to *Information Technology and Tourism*, 2007.

[14] P. Bermejo, J.A. Gámez, and J.M. Puerta. Construcción de atributos: un caso de estudio en clasificación de correo-e. In *Actas del V Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados, MAEB-07.*, 2007 (accepted).

[15] P. Bermejo, J.A. Gámez, and J.M. Puerta. Attribute Construction for E-mail foldering by using wrappered forward greedy search. 2007 (Submitted).

[16] L. M. de Campos, J. G. Castellano. Bayesian network learning algorithms using structural restrictions. To appear in *International Journal of Approximate Reasoning*, 2007.

[17] A. Cano, S. Moral, A. Salmerón. Penniless propagation in join trees. *International Journal of Intelligent Systems*, 15:1027–1059, 2000.

[18] A. Cano, M. Gómez, S. Moral. A forward-backward Monte Carlo method for solving influence diagrams. *International Journal of Approximate Reasoning*, Vol. 42, pp 119-135, 2006.

[19] A. Cano, M. Gómez, S. Moral and J. Abellán. Hill-climbing and Branch-and-bound algorithms for exact and approximate inference in credal networks. To appear in *International Journal of Approximate Reasoning*, 2007

[20] B. Cobb, R. Rumí, and A. Salmerón. Especificación de distribuciones condicionadas de variables continuas en redes Bayesianas. In: *Actas de las VI Jornadas de Transferencia de Tecnología en Inteligencia Artificial*. Pag. 21–28. 2005.

[21] B. Cobb, P.P. Shenoy, and R. Rumí. Approximating probability density functions in hybrid Bayesian networks with mixtures of truncated exponentials. *Statistics and Computing*, 16:293–308, 2006.

[22] Elvira Consortium, Elvira: An environment for creating and using probabilistic graphical models, in: J. Gámez, A. Salmerón (Eds.), Proceedings of the First European Workshop on Probabilistic Graphical Models, 2002, pp. 222–230.

[23] A. Fernández and A. Salmerón. BayesChess: programa de ajedrez adaptativo basado en redes bayesianas. In: *Actas del Campus Multidisciplinar en Percepción e Inteligencia (CMPI-2006)*. Pages 613-624. 2006.

[24] Julia Flores and José A. Gámez. Breeding value classification in manchego sheep: a study of attribute selection and construction. In *Proceedings of the 9th International Conference on Knowledge-Based & Intelligent Information & Engineering Systems (KES-2005)*, volume 3682 of *Lecture Notes in Computer Science*, pages 1338–1346, Melbourne, 2005. Springer Verlag.

[25] Julia Flores, José A. Gámez, Juan L. Mateo, and José M. Puerta. Selección genética para la mejora de la raza ovina manchega mediante técnicas de minería de datos. En *Inteligencia Artificial (Revista Iberoamericana de Inteligencia Artificial)*, 29:69–77, 2006.

[26] M. J. Flores, J. A. Gámez and S. Moral. Finding a partition of the explanation space in bayesian abductive inference. *Proceedings of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty* (ECSQARU-2005). Volumen 3571 of Lecture Notes in Computer Science, Barcelona, Springer Verlag, pp 63-75, 2005.

[27] M. J. Flores, J. A. Gámez and S. Moral. The *Independency tree* model: a new approach for clustering and factorisation. *Proceedings of the Third European Workshop on Probabilistic Graphical Models*, Prague, Czech Republic, pp 83-90, September, 2006.

[28] J.A. Gámez, I. García-Varea and J.Hernández-Orallo. *Minería de Datos: Técnicas y Aplicaciones*. Ediciones del Departamento de Informática de la UCLM, 2005.

[29] J.A. Gámez, J.L. Mateo, and J.M. Puerta. Dependency networks based classifiers: learning models by using independence. In *Proceedings of the 3rd European Workshop on Probabilistic Graphical Models*, pages 115–122, 2006.

[30] J.A. Gámez, J.L. Mateo and J.M. Puerta. Aprendizaje automático de clasificadores basados en redes de dependencia mediante búsqueda heurística. In *Actas del V Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados, MAEB-07.*, 2007. (accepted)

[31] J.A. Gámez, J.L. Mateo and J.M. Puerta. EDNA: Un Algoritmo de Estimación de Distribuciones basado en Redes de Dependencia. In *Actas del V Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados, MAEB-07.*, 2007. (accepted)

[32] J.A. Gámez, J.L. Mateo and J.M. Puerta. Learning Bayesian classifiers from dependency network classifiers. In *Proceedings of the 8th International Conference on Adaptive and Natural Computing Algorithms (ICANNGA-07)*, Lecture Notes in Computer Science, Warsaw, 2007. Springer Verlag. (in press)

[33] José A. Gámez and José M. Puerta. Constrained score+(local)search methods for learning Bayesian networks. In *Proceedings of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU-2005)*, volume 3571 of *Lecture Notes in Computer Science*, pages 161–173, Barcelona, 2005. Springer Verlag.

[34] J.A. Gámez and A. Salmerón. Predicción del valor genético en ovejas de raza manchega usando ténicas de aprendizaje automático. In *Actas de las VI Jornadas de Transferencia de Tecnología en Inteligencia Artificial*, pages 71–80. Paraninfo, 2005.

[35] J.A. Gámez, R. Rumí and A. Salmerón. Unsupervised naive Bayes for data clustering with mixtures of truncated exponentials. In *Proceedings of the 3rd European Workshop on Probabilistic Graphical Models (PGM'06)*. Pages 123-132. 2006.

[36] I. Martínez, S. Moral, C. Rodríguez, and A. Salmerón. Approximate factorisation of probability trees. *ECSQARU'05. Lecture Notes in Artificial Intelligence*, 3571:51–62. 2005.

[37] I. Martínez, C. Rodríguez, and A. Salmerón. Dynamic importance sampling in Bayesian networks using factorisation of probability trees. In *Proceedings of the 3rd European Workshop on Probabilistic Graphical Models (PGM'06)*. Pages 187-194. 2006.

[38] A. Masegosa, H. Joho and J. Jose. Identifying Features for Relevance Web Pages Prediction. *1st International Worskshop on Adaptive Information Retrieval*, Glasgow, UK, pp 36-37, 2006.

[39] A. Masegosa, H. Joho and J. Jose. Evaluating the query-independent object features for the relevancy prediction. To appear in *Lecture Notes in Computer Science* (29th European Conference on Information Retrieval, Roma), 2007.

[40] M. Morales, C. Rodríguez, and A. Salmerón. Estudio de dependencias de indicadores de rendimiento del alumnado universitario mediante redes bayesianas. In *Actas de las VI Jornadas de Transferencia de Tecnología en Inteligencia Artificial*, pages 29–36, 2005.

[41] M. Morales, C. Rodríguez, and A. Salmerón. Selective naive Bayes predictor with mixtures of truncated exponentials. In *Proceedings of the International Conference on Mathematical and Statistical Modeling*, 2006.

[42] M. Morales. Modelización y predicción en estadística universitaria. PhD Thesis. Department of Statistics and Applied Mathematics. University of Almería. 2006.

[43] J.A. Moreno. Gestión semiautomática de los marcadores de un navegador Internet usando técnicas de minería de datos. Proyecto Fin de Carrera. Ingeniería Informática. Universidad de Castilla-La Mancha. Diciembre, 2006.

[44] J.A. Piedra, A. Salmerón, F. Guindos, M. Cantón (2005). Reduction of Irrelevant Features in Oceanic Satellite Images by means of Bayesian networks. In *Actas de las VI Jornadas de Transferencia de Tecnología en Inteligencia Artificial*, pages 133–140.

[45] J.M. Puerta. Introducción a la Minería de Textos. En [28], 67–85. 2005.

[46] V. Romero, R. Rumí, and A. Salmerón. Learning hybrid Bayesian networks using mixtures of truncated exponentials. *Intnl. Journal of Approximate Reasoning*, 42:54–68. 2006.

[47] R. Rumí, Kernel methods in Bayesian networks, in: Proceedings of the International Mediterranean congress of mathematics, 2005, pp. 135 – 149.

[48] R. Rumí and A. Salmerón. Penniless propagation with mixtures of truncated exponentials. *Lecture Notes in Computer Science*, 3571:39–50, 2005.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

HAMOPA-CAD: Herramientas avanzadas para modelado paramétrico en CAD TIN2004-06326-C03

Robert Joan Arinyo *
Escola Tècnica Superior d'Enginyeria Industrial
Universitat Politècnica de Catalunya

Juan Carlos Torres Cantero †
Departamento de Lenguajes y Sistema Informàticos
Universidad de Granada

Francisco Feito Higueruela ‡
Escuela Politécnica Superior
Universidad de Jaén

Abstract

Cooperative design is the term used to refer to a design process where experts from different areas such as conceptual design, shape design, functional design, manufacturing, process planning, etc. work simultaneously on different aspects concerning an ongoing design project. The goal of cooperative design is twofold: shortening the design-to-market cycle and to improve quality.

Generally, the teams of engineers involved in cooperative design are located in different geographic sites and the cooperation is based on distributed information systems. In this scenario, successful cooperative work is only possible if tools to effectively transfer information between sites, and methods to coordinate multi-view systems are available. There are many functionalities in current CAD systems that should be improved to take full advantage of their possibilities. This project aims at contributing to solve some of these deficiencies by exploring and developing advanced tools for parametric modeling and progressive transmission of models.

Keywords: Computer-Aided Design, Collaborative design, parametric modelling, progressive transmission.

*Email: robert@lsi.upc.edu
†Email: jctorres@ugr.es
‡Email: ffeito@ujaen.es

1 Project Goals

This project aims at contributing to solve some deficiencies in the current CAD systems by exploring and developing advanced tools for parametric modeling and progressive transmission between sites.

1.1 Goals

Specifically, the aim is to study the following problems:

- To develop multi-view, parametric geometric constraint-based design techniques to allow each user to specify the object under design in a natural way according to the designer's specific view.
- To develop new modeling schemes and paradigms to represent free form and sculptured solid models, and volumetric properties.
- To develop multiresolution schemes for free form and sculptured solid models, and volumetric properties to facilitate progressive transmission.
- To seek new geometric algorithms to improve the kernel throughput in basic computations.

1.2 Resources

Resources available to develop the project belong to two different categories: human and material. Human resources are

- 6 PhD Senior researchers.
- 11 PhD Junior researchers.
- 11 Researchers pursuing a PhD degree.
- 4 Development and staff engineers.

Among the material reources we have

- Personal and general computing and printing resources.
- Scanner MINOLTA VI-900, manual scaner Polhemus.
- 3D SYSTEMS SLA-250/50 wax prototyping unit, ZCORP Z402-C prototyping unit.
- SenseGraphics 3D-IW immersive workbench.

1.3 Schedule

Developments were organized into two categories of tasks. One category includes those tasks devoted to the project management. The other category includes 26 different tasks which

define the specific research work to be developed. In what follows we shall focus on these tasks that will be detailed in Section 2.1.

The project is a three years project, starting on December 2004 and ending on November 2007. The time unit considered in the project proposal was one month. Therefore the project spans 36 months. Current month is month number 26. The timing planned for the development tasks was as follows

Task	1	2	3	4	5	6	7	8	9	10	11	12	13
Starting month	1	13	19	1	19	25	1	13	25	1	19	1	12
Ending month	24	36	36	24	36	36	12	24	36	18	36	12	30

Task	14	15	16	17	18	19	20	21	22	23	24	25	26
Starting month	1	1	1	1	1	1	1	1	1	1	1	13	1
Ending month	24	34	24	34	36	30	36	36	36	36	36	36	36

2 Goals Accomplishment

For each task we give the current state of the developments along with the problems found, if any, and relevant results.

2.1 Accomplishments

1. Decomposition and completion of geometric constraint graphs.

   We have designed an initial version of a recursive algorithm to decompose contraint graphs as decomposition trees. The algorithm has been implemented and included in the `solBCN` solver. Using this implementation we have conducted a set of experiments to assess both efficiency and reliability. The success lead to replace the old algorithm in the `solBCN` kernel with this new algorithm.

   Currently we are working on a deeper understanding of the algorithm to achieve two results. First we want to develop a formal proof of the algorithm correctness based on graph theory. Then we want to improve the efficiency by reusing in subsequent steps data generated in previous recursive steps.

2. Incremental solving of constraint-based geometric problems.

   Almost no progress has been made in this task. The problem arises from the fact that in paradigms currently applied in the user-system interaction. In general, users can, at their will, add to the current problem either geometry or constraints. In this context, every time the user interacts with the system it is mandatory to check whether the whole problem must solved again.

   Currently we are exploring a new user-system interaction paradigm. Basically we will require the user to define the intended object as a sequence of constructions where at each construction level the current object is defined as a set of smaller welldefined objects which are properly related by geometric constraints.

3. Construction plans based on oriented geometry.

   `solBCN` is a solver developed by our research team as part of the work in the project TIC2001-2099-C03. Now the goal is to rebuild its geometric engine to allow further geometric reasoning. We have redefined the geometric elements the solver can handle associating with each category of geometric element an orientation. Currently we are developing a mathematical library to deal with oriented geometry.

4. The Root Identification Problem. Genetic algorithms: Parameters optimization.

   We have conducted an empirical statistical study to establish the influence of the driving parameters in Genetic Algorithms when applied to solve the Root Identification Problem. We have also identified ranges for the parameters values that optimize the genetic algorithms performance.

   This task has has been succesfully completed and lead to the PhD dissertation developed by the researcher E. Barreiro.

5. The Root Identification Problem. Genetic algorithms: Connecting them with a knowledge database.

   The goal is to build a system to automatically assign values to Genetic Algorithm parameters which optimize the algorithm performance in the Root Identification Problem. To this end, we are developing a framework that integrates Case-Based Reasoning (CBR) and Bayesian Networks (BN). CBR provides the mechanism to acquire knowledge concerning the Root Identification Problem domain. BN provide the tool to model quantitative and qualitative relationships between those parameters that govern the Genetic Algorithm behavior. Results yielded by the prototype so far developed are promising.

6. Continuous searching in the constructive-based geometric constraint solving solution space.

   According to the scheduling this task started in month number 25, just one month ago.

7. Improvement of the edge-based marching cubes algorithm.

   We have developed a method to render the boundary of a volume solid model. The method first tessellates cubes into tetrahedra, then extracts from the tetrahedra an isosurface without ambiguities. The tessellation does not depend on the visualization threshold and is carried out just once therefore the computational effort is smaller than that rerquired by methods which extract the isosurface directly from the cubes. The rendering is performed by marching over the tetrahedra, a method faster and simpler than the extended marching cubes. Moreover, our method allows to smoothly move the isosurface accordingly to small chnanges in the isovalue.

8. Adaptive visualization of volume models.

   Two different approaches has been used to study this task: the tetrahedra tessellation derived on task 7, and the dual cell representation derived in task 9. The first approach allowed us to simplify the original tetrahedra mesh by collapsing those edges that fulfil some criterion. The output of this process is a mesh that represents a low resolution volume. Then a local refining of thsi mesh allows to render the area of interest at full

resolution while the rest of the object is rendered at a law level resolution. This subtask is still in progress. In the second approach we take advantage of the hierarchical structure of the dual cell octree to traverse the octree at different depth according to the user view point or interests. This subtask is completed.

9. Improvement of the cell octree solid modeling scheme.

   We have devised a dual cell octree representation to store the dataset model on a compact octree. We have also designed and implemented a rendering algorithm that builds dual cells and extract isosurfaces on the fly while traversing the octree.

   This task was scheduled to start at month 25 however, we started it earlier to be able to use the resulting representation in the adaptive volume visualization. This task is now completed.

10. Development of efficient algorithms to manage heterogeneous solids bounded by hyper-patches.

    A representation for heterogeous solid has been defined. Also, a set of basic algorithms including rendering, inclusion and conversion has been designed and implemented.

11. Implementation of a modeling system based on heterogeneous solids bounded by hyper-patches.

    We are actually working on the implementation of a test application using the reprentation developed in the precedent task.

12. Optimization of the SP-Octree representation scheme.

    The SP-Octree representation has been improved by using spatial indexes and a slight variation of the data structure. The representation now allows to efficiently manage models including up to several million polygons. Also, model construction and rendering times have been reduced. This new data structure is named BP-Octree (Bounding-Planes Octree), and is the basis for the adaptive visualization of large models. This new data structure allow fast collision detections, progressive transmission and a multiresolution visualization of huge three-dimensional models.

13. Progressive editing of solid models.

    This task has been delayed until completion of task 12. In this way we will be able to develop it using the new BP-octree structure instead of the old SP-octree representation.

14. Adaptive visualization of huge solid models.

    A new visualization algorithm has been developed for the BP-octree representation that allows to change the resolution of selected parts of a 3D model as a function of the viewing position. This is done by using the spatial hierarchy of bounding planes. We achieve a real time frame rate. To avoid distortions at lower levels of detail, we have proposed the use of view dependent impostors.

15. Implementation of a solid modeling system prototype for huge solid models.

    So far, we have set up an immersive desktop system with haptic interaction. To test the system performance, we have implemented on it a teeth cleaning simulation system

using large scanned models (over 200.000 triangles). The next step is to integrate the BP-Octree adaptative visualization algorithm into the immersive workbench and integrate those edition methods derived in task 13.

16. Non photorealistic visualization of solid and volume models.

    We have developed a new light to illuminate volumetric and solid models to produce illustrations based on the HSV colour space. This method can be used both to detect and to explain elements in volumes. Hence, the system can help doctors to diagnose diseases and to explain them to the patients. In addition, we have developed a new method to detect silhouettes in volumes using geometric buffers. The system can be used with lights based on the HSV colour space. We have also expanded the original model of virtual lights to be used with solids or volumes in the same way. These new virtual lights, called *volumetric virtual lights*, are lambda functions which have a position or a direction, and can be composed in an executable pipeline. This model allows the user to compose any volumetric virtual light to generate different images, which can be examined by a doctor or a patient.

17. Implementation of a solid modeling system prototype for volume models.

    We have developed basic algorithms to interactively edit our dual volume-octree representation by either adding or subtracting material from the volume. These algorithms have been implemented and tested on our haptic-inmersive workbench. Currently we are implementing a full virtual sculpting application based on these algorithms.

18. Building a repository of solid models stored in a number of different schemes.

    The design and test of the repository is almost finished. Minor changes will be probably needed. A user interface to interact with the repository through the web has been designed and implemented. Temporary elements of the interface have been designed according to the new standard Web 2.0. Scripts to manage information through the user interface and to run queries on the database have been implemented.

    We expect to develop a set of server-based applications in the last stage of the project. They will allow the user to perform standard queries as well as uploading and downloading files. New functional features to be added in the next year include tools for the automatic processing of 3D meshes to extract the associated topological information.

19. Automatic acquisition and treatment of solid models.

    We have acquired several solid models using our 3D scanners. Some of these models are Iberian art pieces, which are available thanks to a joint project with the University of Jaén aiming at setting up and managing museums devoted to Iberian art.

20. Detecting collisions in 2D, 2D1/2 and 3D solid models.

    We have developed a technique to detect collisions between objects structured in two standard phases: A wide phase and a narrow phase. The wide phase detects situations where objects are clearly apart from each other. When the wide phase cannot elucidate whether there is collision or not, the narrow phase applies specific local tests. Our technique for the wide phase based on polar diagrams has shown to be a convenient structure

to solve the problem. Besades, the technique developed for spatial decomposition methods based on simplicial chains leads to efficient algorithms to detect collisions among complex polihedral objects defined with both complex faces and triangle meshes.

21. Representation of solid models by means of the Extended Layers scheme. Parallel algorithms. Applications.

    We have achieved valid results from the decomposition of objects in extended layers. We have also developed a new object decomposition method, named *multi-Lrep*, which have proven to be efficient with 2D solids. We are now working to extend this method to represent 3D solids. We expect to develop and implement a parallel algorithm to manage multi-Lrep representations embeded in commercial GPU boards.

22. Multiresolution-based terrain visualization.

    We have developed a terrain visualization tool that can render terrains with different levels of detail. It is based on the Digital Elevation Model and textured images. The application computes a triangle mesh which approximates the terrain. The quality of the approximation depends on the terrain characteristics and the camera position. The rendering is performed in real time. The tool also allows the user to define virtual tours by defining checkpoints in the model.

23. Boolean operations on simplicial chains based solid models.

    We have developed algorithms to compute boolean operations in 2D and 3D solid models. We are now working in the optimization of the algorithms and the dissemination of the results, as well as in the theoretical study of new applications. Works in this task have been completed.

24. Modeling and handling of solid models bounded by parametric, triangular patches.

    The study of the algebraic representation of CSG solids built from free-form primitives is finished. The developed model stores information concerning the solid boundary using triangular Bézier patches, and concerning the volume occupied by the solid. Algorithms to render the models have been developed. Our efforts now focus on devising algorithms to compute a Brep for the solids defined by the CSG representation. Works in this task have been completed.

25. Application of simplicial chains representations to virtual reality.

    Preliminary studies show that coverings based on simplicial chains are suitable for virtual reality applications.

26. Implementations on Graphic Processing Unit boards

    The point-in-solid test is a key component in Computer Graphics algorithms, therefore having an algorithm that efficiently solves the problem is paramount in a CAD system. So far we have developed a hardware implementation of this algorithm embeded in a commercial GPU board. The throughput achieved on GPU clearly outperforms the algorithm run in the CPU. We are currently working in the embedding in GPUs of algorithms to compute boolean operations, solid voxelization and collision detection. Preliminary results are very promising.

2.2 Summary

To summarize, we can say that the status of the tasks scheduled in Section 1.3 is the following

- Tasks 4, 7, 9, 10, 12, 13, 14 and 16 have been completed.
- Tasks 3 and 8 are almost completed.
- The rest of the tasks are still under study.
- Management tasks will be active until the end of the project.

3 Results Valuators

To asses the merits of the work so far developed, in what follows we include lists of: PhD dissertations completed by researchers in the project team within the project time span, PhD dissertations under development, papers published in international journals, papers published in international conferences, joint projects with private companies and public institutions aiming at transferring results, and international collaborations that either have already been established or are currently in the process of being established. For the sake of brevity, national papers will only be included in the summary given in Section 3.5.

3.1 Completed PhD Dissertations

1. G. Arroyo. *Visualización expresiva de volúmenes.* Universidad de Granada. May 2006.
2. E. Barreiro. *Modelización y optimización de algoritmos genéticos para la selección de la solución deseada en resolución constructiva de restricciones geométricas.* Universidad de Vigo, March 2006.
3. J.J. Jiménez. *Detección de colisiones mediante recubrimientos simpliciales.* Universidad de Granada, December 2006.
4. C.J. Ogayar. *Optimización de algoritmos geomètricos básicos mediante el uso de recubrimientos simpliciales.* Universidad de Granada, June 2006.

3.2 PhD Dissertations in Progress

1. A. Aguilera. *Modelos multirresolución de terrenos.* Universidad de Jaén.
2. F. Conde. *Representation of non-homogeneous solids using hyperpatches.* Universidad de Granada.
3. M. Freixas. *An Enriched Model for Dynamic Geometry.* Universitat Politècnica de Catalunya.
4. A.L. García. *Modelado de sólidos de forma libre basado en superficies paramétricas triangulares de grado bajo.* Universidad de Jaén.

5. J.R. Jiménez. *Algoritmos eficientes de iluminación global en presencia de medios participativos complejos*. Universidad de Jaén.

6. A. León. *Multiresolution representation of volumes*. Universidad de Granada.

7. F. Martínez. *Representaciones Multi-Lrep. Aplicaciones*. Universidad de Jaén.

8. F.J. Melero. *Hierarchical representation of triangle meshes*. Universidad de Granada.

3.3 Papers in International Journals

1. S.M. Heras, A. Valenzuela, C.J. Ogayar, A.J. Valverde, J.C. Torres. *Computer-based production of comparison overlays from 3D-scanned dental casts for byte mark analysis*. Journal of Forensic Sciences, Vol 50, Vol. 1, 2005.

2. C.M. Hoffmann, R. Joan-Arinyo. *A Brief on Constraint Solving*, Computer-Aided Design & Applications, Vol. 2, No. 5, 2005 pp. 655-663.

3. J.R. Jiménez, X. Pueyo. *Interactive rendering of globally illuminated scenes including anisotropic and inhomogeneous participating media*. The Visual Computer, Vol 21, No. 7, 2005.

4. A.J. Rueda, F.R. Feito, L.M. Ortega. *Layer-based decomposition of solids and its applications*. The Visual Computer. Vol 21, No. 6, 2005.

5. C.J. Ogayar, R.J. Segura, F.R. Feito. *Point in solid strategies*. Computers & Graphics. Vol 29, No. 4, 2005.

6. L. Ortega, F.R. Feito. *Collision detection using polar diagram*. Computers & Graphics, Vol 29, No. 5, 2005.

7. R.J. Segura, F.R. Feito, J. Ruiz de Miras, J.C. Torres, C.J. Ogayar. *An efficient point classification algorithm for triangle meshes*. Journal of Graphics Tools, Vol. 10, No. 3, 2005.

8. G. Arroyo, F. Velasco, D. Martín. *Silhouette detection in volumetric models based on a non-photorealistic illumination system*. Computers & Graphics, 29(2). pp. 209-216, 2005.

9. F.J. Melero, P. Cano, J.C. Torres. *Combining SP-octrees and impostors for the visualisation of multiresolution models*. Computers & Graphics. 29(2), pp. 225-233, 2005.

10. R. Joan-Arinyo, A. Soto-Riera, S. Vila-Mata. *Constrained-Based Techniques to Support Collaborative Design*, Journal of Computing and Information Science in Engineering, June 2006, Vol. 6, pp. 136-148.

11. L. Ortega, A. Márquez, C. Grima. *A new 2D tessellation for angle problems: The polar diagram*, Computational Geometry: Theory and Applications, Vol. 34, No. 2, 2006.

12. J.J.Jiménez, F.R. Feito, R.J. Segura, C.J. Ogayar. *Particle oriented collision detection using simplicial coverings and tetra-trees*, Computer Graphics Forum, Vol 25, No. 1, 2006.

13. F. Martínez, A.J. Rueda, F.R. Feito. *The multi-L-REP decomposition and its application to a point-in-polygon inclusion test.* Computers & Graphics, Vol. 30, No. 6, 2006.

3.4 Papers in International Conferences

1. R. Pavón, F. Díaz, M.V. Luzón. *Adjusting control parameters of genetic algorithms using Bayesian Networks*, 3rd IWPAAMS, Burgos, Spain, 2004.

2. R. Pavón, F. Díaz, M.V. Luzón. *Parameter control in genetic algorithms using Bayesian Networks*, LEARNING'04, Elche, Spain, 2004.

3. F. Martínez, A.J. Rueda, F.R. Feito. *Constructing the Layer Representation of Polygons in Paralle.* International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2005, 31 January-4 February, Plzen (Chekia), 2005.

4. O.L. García, J. Ruiz de Miras, F.R. Feito. *Algebraic Representation of CSG Solids Built from Free-Form Primitives.* International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2005, 31 January-4 February, Plzen (Chekia), 2005.

5. A.J. Rueda, F. Martínez, F.R. Feito. *Layer-based Decompositions of Polyhedra.* International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2005, 31 January-4 February, Plzen (Chekia), 2005.

6. J.R. Jiménez, K. Myszkowski, X. Pueyo. *Interactive Rendering in Dynamic Participating Media Using Selective Photon Tracing.* Spring Conference on Computer Graphics '05. May 12-14, Budmerice (Eslovaquia), 2005.

7. A. León, J.C. Torres, F. Velasco. *Volume Octree contouring by dual cell without explicit cell representation.* Ibero-American Symposium on Computer Graphics, Nu. 2, pp.84-87. Santiago de Compostela (Spain). 2006.

8. R. Pavón, F. Díaz, M.V. Luzón. *An adjustment model in a geometric constraint solving problem.* ACM Symposium on Applied Computing 2006, April 23-27, Dijon, France, 2006.

9. D. Colnet, G. Marpons, F. Merizen. *Reconciling subtyping and code reuse in object-oriented languages. Using inherit and insert in SmartEiffel, the GNU Eiffel compailer.* ICSR9, June 11-15, Turin, Italy, 2006.

10. R. Joan-Arinyo, J. Vilaplana. *Interactive cutting in voxel-based objects. Application to simulate transurethral resection of the prostate.* SIACG'06, Santiago, Spain, 2006.

11. R. Pavón, R. Laza, F. Díaz, M.V. Luzón. *A Bayesian case based reasoning system for automated parameter tuning.* 8th European Conference on Case-Based Reasoning. September 4-7, Ölüdeniz, Turkey, 2006.

12. H. Borges de Barrios, L. Pérez Vidal, *Towards K-shortest paths with behavioral impedance domain from the intermodal public transportation system perspectice*, XIV Congreso Panamericano de Ingenieriía, Tránsito y Transporte, September 20-23, Las Palmas, Spain, 2006.

13. C.J. Ogayar, F.R. Feito, R.J. Segura, M. Rivero. *GPU-based evaluation of operations on triangulated solids*. 3rd Ibero-American Symposium in Computer Graphics. July 5-7, Santiago de Compostela, 2006.

14. O.L. García, J. Ruiz de Miras, F.R. Feito. *Adaptive trimming of cubic triangular Bézier patches*. 3rd Ibero-American Symposium in Computer Graphics. July 5-7, Santiago de Compostela, 2006.

15. L. Ortega, A.J. Rueda, F.R. Feito. *Visibility Map determination using angle preprocessing*. 22nd European Workshop on Computational Geometry. March 27-29, Delfos (Greece), 2006.

16. J.J. Jiménez, C.J. Ogayar, R.J Segura, F.R. Feito. *Collision detection between a complex solid and a particle cloud assisted by programmable GPU*. Third Workshop in Virtual Reality Interactions and Physical Simulations. Móstoles (Madrid), October 2006.

17. F. Velasco, J.C. Torres, A. León, F. Soler. *Adaptive cube tessellation for topologically correct isosurfaces*. 2nd International Conference on Computer Graphics Theory and Applications, Barcelona (Spain), 8-11 March, 2007.

3.5 Summary

The total number of international and national papers is summarized in the following table.

Category	# of papers
International journals	13
International conferences	17
National conferences	25
Chapters in national books	1

3.6 Technology Transfer

1. Institut Català d'Urologia i Nefrologia S.L. *Computerized simulator for training surgeons in transurethral resection of the prostate.*

3.7 International Collaboration

3.7.1 Universitat Politècnica de Catalunya Team

Guillerm Marpons, PhD candidate and a researcher in the team, visited Prof Colnet of LOIRA (CNRS-INPL-INRIA) at Nancy (France) from January through April 2006. The goal was to

conduct research in objected-oriented languages (specifically in the Eiffel Language) and its applications to geometric repesentations.

In the well established framework of collaboration with Prof C.M. Hoffmann of Purdue University (USA), Prof Hoffmann and R. Joan-Arinyo drafted an overview of the state of the art in Geometric Constraint Solving. The result was an invited lecture delivered in the CAD'05 Conference held in Bangkok in 2005.

3.7.2 Universidad de Granada Team

F.J. Melero completed a research stay of three months in the Department of Computer Graphics, Max Planck Institut für Informatik Saarbrücken (Germany). Currently, we are planning new visits.

A. León completed a research stay of three months with Prof Dirk Bartz, head of the Visual Computing for Medicine Group. Graphisch-Interaktive Systeme, Wilhelm-Schickard-Institut, Tübingen University.

3.7.3 Universidad de Jaén Team

Roberto Jiménez, a researcher in our team, completed a research stay in the Department of Computer Graphics, Max Planck Institut für Informatik (Saarbrücken, Germany). Currently, we are planning new visits.

Prof R. Scopigno, Research Director of Visual Computing Lab, Pisa (Italy) reviewed several PhD Thesis authored by members of our research group.

Our reseracher Juan José Jiménez has been formally invited by the GAMMA Group of the University of North Carolina (USA) for a longstay research visit within year 2007.

Recently we have established preliminary contacts with Prof Bryson R. Payne of North Georgia College and State University, whose research interests include GPU programming. The aim is to colaborate in the applications of GPUs to current open problems in solid modeling. We have also approached Prof A. Augusto de Sousa of the Instituto de Engenharia de Sistemas e Computadores do Porto.

We have signed several agreements with european universities to host Erasmus students. Other collaboration agreements to teach and conduct research have been established with universities in the USA.

Since 1996 researchers in our group have been serving as members in the Executive Committee of the Spanish Chapter of the Eurographics Association.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Developing Distributed Computing Techniques for Scheduling. Application to Transportation Problems TIN2004-06354-C02

Federico Barber*
Pilar Tormos, Antonio Lova,
Miguel A. Salido
Polytechnic University of Valencia

Beatriz López**
Santiago Esteve, Miquel Montaner
Dídac Busquets, Silvia Suárez
University of Girona

Abstract

This project aims at the development of distributed computing techniques and multi-agent systems for the resolution of constraint satisfaction problems (CSP), these techniques particularly guided towards planning and scheduling processes. So that, it is a coordinated project in which two research groups work together with complementary experience on: (i) CSP techniques, distributed CSP, planning and scheduling and (ii) distributed and open frameworks, multi-agent systems and multi-agent CSPs. The scientific-technique objectives are very relevant and important allowing us to manage problem typologies in which traditional techniques are not appropriate. Additionally, in the research project it is sought to apply and evaluate the developed techniques in evaluation environments broadly contrasted in the scientific community and, mainly, with two concrete and real frameworks based on rail transport and intercity transport in which it is maintained respective agreements of collaboration.

The scientific capacity of the teams is contrastable by means of the results obtained in previous collaboration research projects. The last research project was carried out in conjunction by both teams. Furthermore, both groups have participated in projects and European networks of similar thematic and the relationships with other national and international groups.

Keywords: Constraint satisfaction, scheduling, distributed techniques.

1 Project Objectives

The coordinated project is focused towards the achievement of the following goals:

1. Development of scheduling techniques, by integrating mathematical algorithms, operational research techniques, intelligent techniques and topological methods, toward the efficient management of complex problems.
2. Enlarge the application of Constraint Satisfaction Problems (CPSs) by means of multi-agent techniques and distributed CSPs.

* Email: fbarber@dsic.upv.es (leader of Subproject TIN2004-06354-C02-01)
** Email: blopez@silver.udg.es (leader of Subproject TIN2004-06354-C02-02)

3. Development of an integrated architecture for planning and scheduling.
4. Application of the architectures and techniques developed on real-world environments proposed by companies as well as its application on relevant benchmarks generated by the scientific community of planning and scheduling.

The common goal of the research project is to develop distributed techniques and multiagent approaches to manage CSPs, focused on solving planning and scheduling processes. These goals include the integration of both processes, and the evaluation of these techniques in real life problems.

The scientific capacity of the teams is contrastable by means of the results obtained in previous collaboration research projects (see, for example, publication [LBG05]). The scientific expertise of subgroup 1 in CSP and scheduling techniques is very appropriate to reach the proposed objectives. Furthermore, the expertise of subgroup 2 in multi-agent system complements the necessary knowledge for the group to reach the proposed objectives focused on distributed CSPs.

Each subproject is committed to apply the developed models and techniques to common optimization contexts in transportation scheduling problems.

1.1 Subproject's Objectives

The objective of the Project is distributed between both subprojects in the following way:

Objectives of Subproject-1 (UPV subgroup): "Development of techniques for Distributed Constraint Satisfaction Problems and their application to planning and scheduling"

- Development of methods for constraint satisfaction problems, by integrating mathematical algorithms, operational research techniques, heuristic techniques and topological methods, toward the efficient management of complex scheduling problems
- Development of methods for distributed CSP, by identifying types and hierarchies of constraints. Development of methods and heuristics for solving distributed CSP.
- Development of an integrated architecture for planning and scheduling, satisfying temporal constraints and those related to the limited availability of shared resources.

Objectives of Subproject-2 (UdG subgroup): "Development and application of scheduling techniques in multi-agent environments".

- Development of multi-agent techniques for solving constraint satisfaction problems. For such a purpose, the study of the current distributed techniques and multi-agent paradigms are required.
- Development of collaboration and negotiation techniques for competitive agents that share common resources
- Development of partition techniques in order to solve a scheduling problem with the collaboration of heterogeneous agents with different capabilities.

These objectives address the applicability of the state of the art techniques on constraint satisfaction problem solving. Their achievement would allow coping with quite complex problems, particularly significant in the industrial environment due to their practical interest. For this reason, the project has been focused on practical issues. In fact, the developed architectures and techniques are applied in real environments in which the research group has a large experience throughout its collaboration with companies. Such applications will be those related to the optimization of the use of railways infrastructures and the inter-urban passenger road transportation.

In the following temporal graph, we present the different tasks of the project. The first block of tasks are common tasks, which must be carried out between both groups. The second block of tasks corresponds to the UPV group (subproject-1) and the third group of tasks is assigned to UDG group (subproject-2).

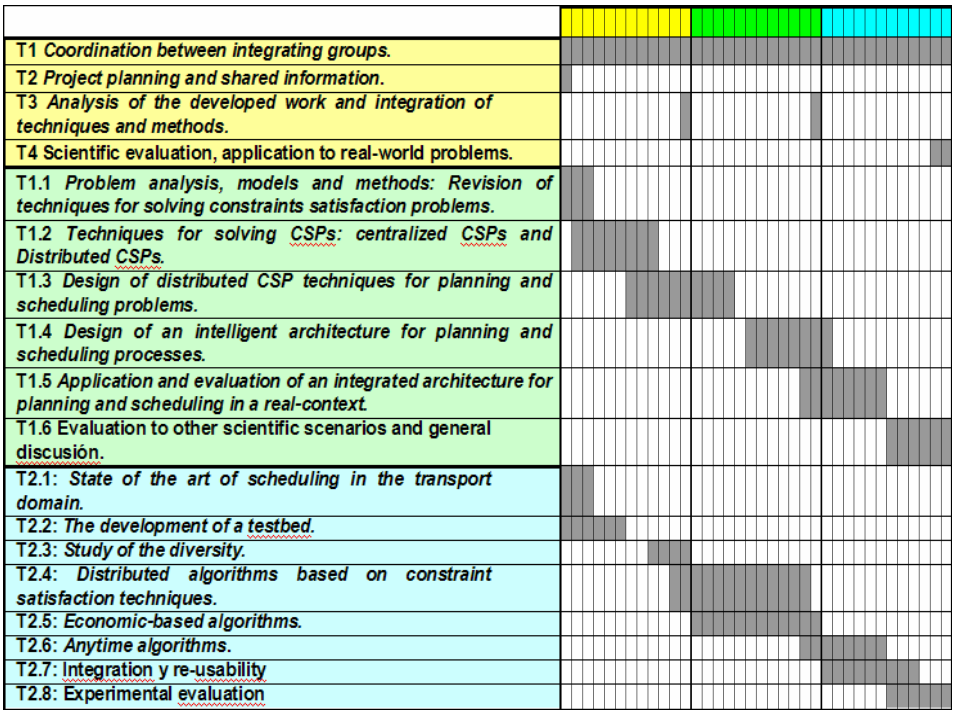


Figure 1.- Research Task of the Project

2 Degree of performance

In this section the activities and achievements of the project are described, together with the difficulties found during their deployment and the proposed solutions.

2.1 Activities and achievements

This section describes the achievements during the firsts two years of the project organized according to the planned tasks. First, the activities and achievements of the coordination tasks are described, and following the activities of each subgroup.

2.2.1 Coordination activities

T1: Coordination between integrating groups. The coordination between both groups has been held regularly mainly by electronic tools (emails, webpage). A kick-off meeting and other on-site meetings have been also carried out during the project. International and national events have given also the opportunity to meet and interchange information about the research progresses.

T2: Project planning and shared information. Several reports and other papers and documentation have been shared between the UPV and UdG groups. Planning of tasks has been detailed during the meetings, where the coordinators have defined the different tasks to develop. A webpage (http://www.dsic.upv.es/users/ia/gps/DiSched/index.html) has been set up to coordinate the project.

T3: Analysis of the developed work and integration of techniques and methods. At several stages, a revision of the developed work and exchange of experiences have been carried out by emails and actual meetings. The work performed on integration of techniques and their evaluation on a common problem has given rise to submit a paper to an international journal [LMM+06].

2.1.2 Techniques for distributed constraint satisfaction problems

Up to date, the subproject-1 has completed the first four tasks of the project:

UPV. T1.1: Problem analysis, models and methods: Revision of techniques for solving constraints satisfaction problems. This task was completed during the first three months of the projects. The research group reviewed the latest and innovative techniques and current trends for CSP's and intelligent scheduling developed in the scientific community.

UPV. T1.2: Development of techniques for solving CSPs: centralized CSPs and Distributed CSPs. This task was focused on the development of new CSP techniques, both for classic and for distributed CSP's. Centralized techniques were focused on solving the widely known Resource Constrained Project Scheduling Problem (RCPSP). Most existing algorithms for solving this problem are related to priority rule-based sequencing heuristics, metaheuristics and exact methods. We have developed a new method based on single and multi pass very-efficient heuristics, based on Parameterised Regret-Based Random Sampling methods, which is especially useful because it intensifies the search in promising areas of the search space ([LTB06]). On the other hand, developed distributed techniques were based on graph partitioning, where the completed graph is divided into several sub-graphs as independent as possible. To carry out this partition, we have used METIS, a well known graph partitioning tool [5]. We have developed a distributed technique that can solve a CSP by dividing the CSP into several sub-CSPs. This technique solves concurrently all sub-CSPs and the efficiency of this technique has been contrasted with some of the well-known CSP techniques. Developed works have been published in different states of development in several conferences ([SAB+05], [ASB+05]) and finally in the journal referenced in [SB06].

UPV. T1.3: Design of distributed CSP techniques for planning and scheduling problems. We have developed new models for distributing CSP's, both, as general-oriented and as problem-oriented approaches. General-oriented approach is based on distributing the CSP by means of trees, so that the resultant sub-CSP are solved efficiently without backtracking. The main difficulty of this proposal is the way for obtaining the selected trees. The problem of obtaining tress from a CSP is a combinatorial problem, so that, we have selected a technique that minimize the number of backtracking between different trees. Problem-oriented approaches were focused on planning and scheduling problems (see T1.4), which also give us the opportunity of refining the general distributed technique into a domain dependent distributed model. Results of this task can be seen in [LMM+06], and [ASB07].

UPV T1.4: Design of an intelligent architecture for planning and scheduling processes. In this task, we have considered the planning process as a collection of interacting, autonomous and

flexible components (distributed CSP's), aiming at representing the problem among several distributed processes, solving subproblems incrementally and aggregating subproblem solutions to achieve a global solution. Therefore, constraints should be distributed among several CSP's. Here, we have analysed the use of both general methods (graph partitioning techniques and problem-oriented methods) to distribute the problem. One of the first conclusions is that domain-dependent distributed constraint models maintain better behaviours than general distributed models. Our study has been focused on the railway scheduling problems, where specific railway characteristics make the problem domain-dependent. Each instance depends on the number of trains in each direction, the number of stations, the distance between stations, etc. Our evaluation shows that general distributed models have a better behaviour than the centralized model, but it also shows that domain-dependent distributed models are more efficient than general ones. Domain-dependent models make a distribution that is based on trains (which can be assimilated to the tasks to be done) and on stations (which can be assimilated to resources) [SAB+07] [TAS+06]. Research on models for domain-based distribution is a promising area of research in distributed CSPs, which can also be applied to a distributed architecture for planning and scheduling processes where each subCSP collaborated to find a global solution. On the other hand, developed work in this task has provided a better understanding of planning and scheduling problems, new results, and new research lines in the area [ABS+06], [ILB+06], [SAB+06] and [IBT+07].

UPV T1.5 Application and evaluation of an integrated architecture for planning and scheduling in a real-context. This task is closely related to T1.2, T1.3 and T1.4 tasks, and its goal is the application and evaluation of centralized and distributed CSP-based paradigms to real contexts of planning and scheduling. These evaluations are allowing us to obtain interesting conclusions about heuristics for distributing problems and for solving the resulted distributed CSP's. We are applying developed methods on solving real-world railway scheduling problem, where the main objective is to plan several trains, that share common resources, such as platforms, section tracks, and even drivers and conductors. Preliminary results of performed evaluations can be seen in [LTB+06], [BTL+06] and [ASB+06]

2.1.2 Techniques in multi-agent environments

The tasks addressed by the second subgroup have been the following:

UdG T2.1: State of the art of scheduling in the transport domain. This task has been completed. We have followed an experimental approach to the state of the art. Then, a problem type has been defined, and different models have been developed, one for each technique, and then their feasibility has been tested in this problem type. Next, by the use of a testbed (see task T2.2), all the techniques have been evaluated regarding different measures: time, expressiveness, and optimality among others. Bio-inspired and metaheuristics approaches have been shown to be performed the best, regarding time and optimality. Classical integer programming techniques have been proved to work well, even than it is not clear their scalability. Distributed and search approaches have been shown their power regarding the problem modelling (language expressiveness), even that much more work should be done in order to achieve response time closer to other classical approaches. The state of the art has been submitted to the Artificial Intelligence Review journal (see the [LMM+06] publication in section 3.5)

UdG T2.2: The development of a testbed. This task has been also completed. Two different test beds have been developed: a synthetic and a real data sets. Both data sets have been defined in the

road passenger transportation domain. On one hand, the synthetic datasets consists of up to 70 cases, of different complexity. The first case is conformed by a single driver and a single service. The second case is a problem in which two drivers and two services are defined, and so on until the 70 case. The data corresponding to the services and drivers have been randomly generated. On the other hand, the real data set consist of a real case of a company (one of the companies interested on the project or EPO) corresponding to a real situation of a work day of the company. This case has up to 60 services. So in this sense, the synthetic data set is even more complex than the real one. As a result of this task, a report formalizing the problem has been produced [L05], and in publication [ML06] the full explanation of the problem type is given together with some preliminary experiments.

UdG T2.3: Study of the diversity. This task has also been completed. The national and European legislation regarding driver journeys in the road passenger transportation services has been studied. Different kind of resources has been identified, that is, different kind of buses (microbuses, buses, coaches), transports (road and passenger), and services (regular, especial, shuttle, and discretionary). The main difficulty has been to separate from the information available the issues regarding regular services and just-in-time services (discretionary), the focus of the current research problem. A key point in this study has been to identify the different time measures that are currently applied to drivers, and that we have grouped in four main time measures: journey length, driving time, waiting time (or presence time) and weekly resting time. The legislation regarding all these different times is complex to formalize due to the amount of exceptions that it contemplates. For its understanding, several examples have been developed in [UL05].

UdG T2.4: Distributed algorithms based on constraint satisfaction techniques. This task has been completed. The starting point has been the asynchronous backtracking algorithm of Makoto Yokoo [1] and its extension to deal with optimization problems (ADOPT [2]). Theses algorithms presents several drawbacks mainly due to the fact they require binary constraints, and for our test bed, such a requirement means that wee need to deal with thousands of constraint, that make the algorithm exhibit a very low performance. The results of this task are described in [LMM+06].

UdG T2.5: Economic-based algorithms. This task, related to economic algorithms, is still running. Our starting point has been the study of marked-based mechanisms, and particularly auctions. First, we have studied the applicability of combinatorial auctions. For this purpose, bid policies for using the CASS [3] algorithm in the transportation domain have been developed. Additionally, a re-scheduling approach to deal with incidences has been integrated in the MAGNET system [4], a framework designed at Minnesota University for supporting multiple agents in negotiating contracts for tasks with temporal and precedence constraints[1]. Developed works have been the source of the following publications: [LU05], [LUA05], [SCB05], [SL06]. Finally, different inverse auction mechanisms have been developed for a particular kind of road transportation: the urgent patient transportation. In these latter algorithms the concept of "coverage" has been incorporated in the decision procedure, so a multi-attribute clearing market algorithm has been developed. That is, instead of taking into account a single attribute (response time), the time the driver would be available again is also taken into account when deciding the scheduling. Different trust measures regarding the available resources have also been incorporated in the algorithms. This work is published in [LIB06], [LAI+05], [IBL06], and [LIA+05].

[1] This integration approach is part of task 2.7 (Integration and re-scheduling) that will be continued in the third Project year.

UdG T2.6: Anytime algorithms. This task has been started and it is still running. Some of the classical constraint optimization algorithms have the possibility of providing the best solution found so far at any time, but in a distributed environment, this feature is not being tackled in most of the algorithms. It is the possibility to stop the algorithm, but not to resume it again. So, further research is required for the third project year.

2.2 Difficulties and proposed solutions

The fact of being a coordinated project has been useful to beat the difficulties found during the project, since the exchange of advices and experiences between the research groups involved has enriched the common research of the project. For example, the experience of the UPV group on CSP has benefit the UdG group on coping with the complex modelling of the road passenger transportation problem, while the UdG group has provided the background on the problem. Additionally, at the completion of the design and development tasks, scientific and experimental evaluations will be carried out during the third year, with a final phase of useful discussions and conclusions. However, particular difficulties have been found in each subproject, and they are described in the remaining of this section.

2.2.1 Techniques for distributed constraint satisfaction problems

The main problems found during the development of the project have been related to the accessibility of existing distributed techniques to evaluate the performance of our distributed techniques, so we have evaluated our distributed techniques against our centralized techniques. Moreover, we should remark the difficulty of working on real-life problems, with a very big amount of data and very complex constraints to be satisfied. However, at the same time, these real cases allow us to evaluate developed methods on interesting, useful and realistic problems.

2.2.1 Techniques in multi-agent environments

Some of the problems encountered during the work have been the response time of the distributed algorithms. It seems clear that during the development of the project, the research community is aware that for a reasonable problem size the use of distributed approaches cannot be justified because of response time. However, the use of a distributed approach relays on the fact that the problem information is distributed among different agents, or there exist privacy requirements, and there is no possibility that a single entity keeps all the information. In this latter case, economic approach as auctions allows a good combination of both situations. On one hand, information is natural distributed to each agent (for example, each driver is aware about its current agenda, hours driving during the week, exceptions, incidences, etc), and a single agent who request a summarized information to all the agents (bids) can use of a very well known central scheduling algorithm to solve the allocation in a reasonable computational time.

3 Result indicators

The results of the Project are here described in terms of trainees, publications, technology transfer, collaborations with international groups, and publications.

3.1 Trainees

- **Montserrat Abril (UPV)**, computer engineer, has been collaborating in the project since its beginning. Working area: Distributed CSP's, Capacity in scheduling systems. During the project she has got the DEA degree and, currently, she is concluding the evaluation phase of her Ph. Doctoral Thesis "General versus problem-oriented models for distributing and solving distributed constraint satisfaction problems", which is planed to be presented on May'2007.
- **Laura Ingolotti (UPV)**, computer engineer, has been collaborating in the project since its beginning. Working area: Heuristics for CSP and Scheduling problems. Currently, she has finished the works about of her Ph. Doctoral Thesis "An architecture and heuristic proposals for solving very-complex constraint satisfaction problems. Application to Railway scheduling and timetabling problems", which is planned to be presented on January'2007.
- **Mariamar Cervantes (UPV)**, computer engineer, has been collaborating in the project since Oct'06, in the second year of the project, due to a research grant of the 'Fundación Carolina'. Working area: Robustness in CSP solutions, Scheduling.
- **Silvia Suárez (UdG)**, computer engineer, has been collaborating since the beginning of the project. Working area: scheduling in dynamic environments using combinatorial auctions.
- **Pere Urra (UdG)**, computer engineer, collaborates occasionally in the project. Since the subproject has not been awarded by any FPI gran, he shares the research with a conventional work. Working area: analysing the diversity of the problem transportation domain, and its suitability to current distributed algorithms. He has started his PhD courses on the first year of the project, and it is expected that he gets the DEA on October 2007.
- **Javier Murillo (UdG)**, computer engineer, is collaborating in the project since the second year. Working area: development of classical algorithms for the experimental analysis. He has started his Master courses during the second Project year.
- **Victor Muñoz (UdG)**, computer engineer, is collaborating in the project since the second year. He shares the research with a conventional work.. Working area: development of distributed algorithms for the experimental analysis. He has started his PhD courses, and we expect that he gets the DEA on June 2007.

3.2 Employees

3.2.1 Subproject TIC2004-06354-C02, UPV

- **Laura Ingolotti** has been hired as a computer engineer during July'06 / April'07.

3.2.2 Subproject TIC2004-06354-C02, UdG

- **Javier Murillo** has been hired as a computer engineer during October-2006-December 2006.

3.3 Technology transfer

The project has been conceived since the beginning from a practical point of view, so most of the solutions found have an applicable to real environments, as shown the following technology transfer achieved. Furthermore, the research groups plan to make available for the scientific community the developed tools and systems at the web page of the project.

3.3.1 Subproject TIC2004-06354-C02, UPV

Many of the basic techniques developed during this project, especially those related to centralized and distributed CSPs techniques as well as heuristics for solving very complex planning and scheduling problems, have brought on transferences of technology in different environments:

- Optimization of train timetables is a hard, difficult and time-consuming task that implies the consideration of complex and hard traffic constraints, a huge amount of distributed data, several and distributed priorities to be balanced & optimized, etc. In this area, an agreement for transference of technology (Dec'05/Dic'08) has been formalized with the "Administrador de Infraestructuras Ferroviarias (ADIF)". Prototypes developed are currently being used in the company and considered as a very useful supporting tool. More details can be seen at: http://www.dsic.upv.es/users/ia/gps/MOM/.
- Staff allocation and timetable generation imply to handle a big number of constraints. Furthermore, several criteria have to be optimized and highly efficient procedures are required in the complex and distributed environments existing in real life. In this area, an agreement for transference of technology has been signed up in Jun'06 with an important company of commercial distribution. The problem we are facing implies the assignment to near to 55,000 employees in more than one thousand of centers. Prototypes developed are being evaluated and considered by the company as a very successful approach to the problem.

3.3.2 Subproject TIC2004-06354-C02, UdG

Thanks to the experience acquired on scheduling problems during the first year of the project, the MANTER Company has signed a contract with the university in order to develop a system related to cutting problems.

During the second year, a demo has been given to the responsible of an ambulance company regarding the applicability of auction techniques to urgent patient transportation. We are still in contact with this company in order to study some future collaboration.

3.4 Participation in international projects, relations with other groups

There are several research groups abroad sharing related research topics to the Project, from which several collaborations and projects have been set up as explained in this section.

3.4.1 Subproject TIC2004-06354-C02, UPV

The knowledge gathered along this project has facilitate the participation of the subgroup of the UPV in "Algorithms for Robust and online Railway optimization: Improving the Validity and reliAbility of Large scale systems (ARRIVAL), a Specific Targeted Research Project funded by the FET (Future and Emerging Technologies) Unit of the European Commission, priority IST (Information Society Technologies), within the 6th Framework Programme, under contract no. FP6- 021235–2. This project has three years of duration and it started in Feb'06.

The participation of the UPV in the project is mainly related to algorithmic research by tackling optimization questions in perhaps the most complex and largest in scale setting: that of railway systems. Railway optimization deals with planning and scheduling problems over several time horizons. We focus on two important and actually unexplored facets of planning that pose even

harder optimization questions: robust planning and online (real-time) planning. Our main goal is to develop new knowledge through an integrated, interdisciplinary approach drawn from algorithmic, operations research, and constraint programming techniques. The project has 13 partners of several universities and research centers. More details can be seen at http://arrival.cti.gr/.

3.4.2 Subproject TIC2004-06354-C02, UDG

One of the research members of the UdG subgroup has been moved to the Minnesota University during three months the last two years to work with professor John Collins. The goal of the stay has been to know the advances of Dr. Collins related to combinatorial auction techniques that take into account precedence relationships. One of the publications shows the results achieved (see section 3). PhD students have been enrolled in International advanced courses, as the ones organized by ICAPS on planning and scheduling and Agentlink on agents(EASSS).

3.5 Publications

3.5.1 Journal papers

- [ASB07] "Distributed Models for solving CSPs" M. Abril, M.A. Salido, F.Barber. INFOCOMP, Journal of Computer Science, ISSN: 1807-4545 Fed. Univ. of Lavras (Brasil). (accepted, to be published) (2007)
- [LTB06] "Multi-Mode Resource Constrained Project Scheduling: Scheduling Schemes, Priority Rules and Mode Selection Rules" A. Lova, P. Tormos, F. Barber. Inteligencia Artificial, Revista Iberoamericana de I.A. Vol.10, No.30 (2006), ISSN: 1137-3601. pp. 69-86.
- [SAB+07] "Domain Dependent Distributed Models for Railway Scheduling" Salido, Miguel A., Abril, M., Barber, F., Ingolotti, L., Tormos, P. and Lova, A. International Journal Knowledge Based Systems. (Elsevier). (Accepted – to be published in 2007).
- [SB06] "Distributed CSPs by Graph Partitioning" Miguel A. Salido, F. Barber. Applied Mathematics and Computation (Elsevier). Vol 183, pp:491-498, 2006.
- [ABS+06] Abril, M., Barber, F., Salido, Miguel A., Ingolotti, L., Tormos, P. and Lova, A. "Applying Computer-Based Analytical and Empirical Methods to the Assessment of Railway Capacity", Transportation Research - Part E (under the second revision)
- [LIB06] A Multiagent System to Support Ambulance Coordination of Urgent Medical Transportation. Beatriz Lopez, Bianca Innocenti, Didac Busquets. IEEE Intelligent Systems (submitted on September, 21th, 2006).
- [LMM+06] Experimental analysis of optimisation techniques on the road passenger transportation problem. Beatriz Lopez, PhD; Victor Muñoz; Javier Murillo; Federico Barber, PhD; Miguel Angel Salido, PhD; Montserrat Abril; Mariamar Cervantes; Luis Fernando Caro; Mateu Villaret, PhD. Artificial Intelligence Review (submitted on September 15th, 2006).

3.5.2 Papers at conference proceedings

- [ASB+05] "A Heuristic Technique for the Capacity Assessment of Periodic Train", Abril, M., Salido, M.A., Barber, F., Ingolotti, L., Lova, A., Tormos, P., Frontiers in Artificial Intelligence and Applications (IOS Press) 131, 339-346, (2005).

- [ASB+06] "Distributed Models in Railway Industry" M. Abril, M. A. Salido, F. Barber, L. Ingolotti, A. Lova, P. Tormos Proc. of the 1st Workshop on Industrial Applications of Distributed Intelligent Systems (INADIS'2006), 2006. CD-ROM. ISBN 85-87837-11-7
- [BTL+06] "A Decision Support System for Railway Timetabling (MOM): the Spanish Case", F. Barber, P. Tormos, A. Lova, L. Ingolotti, M. A. Salido, M. Abril In: WIT Trans. in The Built Environment: Computer in Railways (ISSN: 1743-3509). WIT Press.Vol.10, 235-244. Selected from: 10th Int. Conf. on Computer System Design and Operation in the Railway and Other Transit Systems (Comprail'2006). Prague, Czech Republic (2006)
- [IBL06] "Coordinación del transporte sanitario basado en subastas". Bianca Innocenti, Dídac Busquets, Beatriz López. INFORMED, Murcia (2006), pp. 61-66.
- [IBT+07] "A scheduling order-based method to solve timetabling problems" L. Ingolotti, F.Barber; P. Tormos, A. Lova; ; M. A. Salido M. Abril, Lecture Notes in Conmputer Science, LNAI. Springer (Accepted- to be Published).
- [ILB+06] "New Heuristics to Solve the CSOP Railway Timetabling Problem". L. Ingolotti, A. Lova, F. Barber, P. Tormos, M. A. Salido, M. Abril. In: Advances in Applied Artificial Intelligence. Lecture Notes in Computer Science, (Springer), Vol. 4031, pp.400-409. Selected from "The 19th International Conference on Industrial, Engineering and other Applications of Applied Intelligent Systems", IEA/AIE 2006, Annecy, France (2006).
- [LIA+05] "A Multi-Agent System to Support Ambulance Coordination in Time-Critical Patient Treatment." B. López, Innocenti, B., Aciar, S. and Cuevas, I. 7º SImposio Argentino de Inteligencia Artificial, ASAI 2005, Rosario, Argentina, 29 Agosto – 2 Setiembre, 2005. En CD ROM de las 34 Jornadas Argentinas de Informática e Investigación Operativa (ISSN 1666 1141), ASAI 2005 (ISSN 1666 1079) pp.43-54
- [LSI+05] "How multi-agent systems suport acute stroke emergency treatment." López, B. Aciar, S., Innocenti, B. and Cuevas, I. IJCAI Workshop on Agents Applied in Health Care, pp. 51-59, Edingourgh, Scotland, July 30, 2005.
- [LTB+06] "Técnicas para la programación del Tráfico Ferroviario Heterogéneo" A. Lova, P. Tormos, F. Barber, L. Ingolotti, M. Abril, M.A. Salido Procc. of the XXIX Congreso Nacional de Estadística e Investigación Operativa. Tenerife (ESPAÑA) 2006
- [LU05] "Decentralized resource allocation for road passenger transportation" B. López, P. Urra. Workshop sobre planificación, Scheduling y Razonamiento Temporal (asociado a CAEPIA), Santiago de Compostela, 13 Noviembre 2005, pp. 9-19.
- [LUA05] "Multi-agent resource allocation for road passenger transportation". López, B. , Urra, P., and Abelló, X. In: I.García and F.J. Rodríguez (eds): IV Internacional Workshop on Practical Applications of Agents and Multiagent Systems (IWPAAMS 2005), Universidad de León, ISBN 84-9773-222-7, pp: 207-216.
- [ML06] "An empirical study of planning and scheduling interactions in the road passenger transportation domain". J. Murillo and B Lopez. The 25th Annual Workshop of the UK Planning and Scheduling Special Interest Group (PlanSIG), Nottingham, UK, December 2006, pp. 129-136.
- [QRV05] "Studies about the Atomic Capabilities Concept for Linear Control Systems in Physical Multi-Agent Environments". Christian G. Quintero, Josep Ll. De la Rosa, J. Vehí. 6th

IEEE International Conference on Computational Intelligence in Robotics and Automation, Espoo, Finland, pp. 727-732 (2005).

- [SAB+05] "Topological Constraints in Periodic Train Scheduling", Salido, M.A., Abril, M., Barber, F., Ingolotti, L., Tormos, P., Lova, A. Frontiers in Artificial Intelligence and Applications (IOS Press) 117, 11-20, (2005).
- [SAB+06] "Domain Dependent Distributed Models in Railway Industry" Salido, Miguel A., Abril, M., Barber, F., Ingolotti, L., Tormos, P. and Lova, A. The 26th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence. AI-2006. Vol.14, pp.163-176 (2006)
- [SCL05] "Improving Rescue Operations in Disasters: Approaches about Task Allocation and re-scheduling". Suárez, S., Collins, J., López, B. The 24th Annual Workshop of the UK Planning and Scheduling Special Interest Group (PlanSIG), London, UK, Dec'2005, pp: 66-75.
- [SL06] "Reverse Combinatorial Auctions for allocating Resources in Rescue Scenario". Suarez, S, López, B. ICAPS Workshop on Constraint Satisfaction Techniques for Planning and Scheduling Problems, Cumbria, UK. 62-65, (2006).
- [TAS+06] "Distributed Constraint Satisfaction Problems to Model Railway Scheduling Problems" P. Tormos, M. Abril, M. A. Salido, F. Barber, L. Ingolotti, A. Lova. In: WIT Trans. in The Built Environment: Computer in Railways (ISSN: 1743-3509). WIT Press.Vol.10, 289-297. Selected from: 10th Int. Conf. on Computer System Design and Operation in the Railway and Other Transit Systems (Comprail'2006). Prague, Czech Republic (2006)

3.5.3 Reports and others

- [L05] "Time control in road passenger transportation. Problem description and formalization.". B. López. IIiA Research Report 05-04, University of Girona.
- [LBG05] "Disseny i planificació d'agents físics dinámics. Aplicació a Robocup Rescue". López, B., Barber, F., García, M.A. Butlletí ACIA núm. 34, 20-21, Barcelona, Tardor, 2005.
- [UL05] "Estudi de la diversitat en el transport de passatgers per carretera". P. Urra, B. López. IIiA Research Report 05-05, University of Girona.

4 References

[1] Yokoo, M. and Hirayama, K. (2000). Algorithms for distributed constraint satisfaction: A review. Autonomous Agents and Multi-Agent Systems, 3:185-207.

[2] Modi, P., Shen, W., Tambe, M., and Yokoo, M. (2005). Adopt: Asynchronous distributed constraint optimization with quality guarantees. Artificial Intelligence Journal, 161:149-180.

[3] Leyton-Brown, K., "Resource Allocation in Competitive Multiagent Systems". Ph.D. thesis, Standford University, August 2003.

[4] MAGNET System web. Intelligent Agents for Electronic Commerce. http://www.cs.umn.edu/magnet/

[5] G. Karypis and V. Kumar. MeTiS: A software package for partitioning unstructured graphs, partitioning meshes, and computing fill-reducing orderings of sparse matrices. Version 4.0, Univ. of Minnesota, Dept. of Computer Science, 1998. Available at http://wwwusers.cs.umn.edu/Nkarypis/metis/metis.html

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

IN2GESOFT: Innovation and Integration of Methods for the Development and Quantitative Management of Software Projects TIN2004-06689-C03

José Javier Dolado *
Universidad del País Vasco - Euskal Herriko Unibertsitatea

Isabel Ramos †
Universidad de Sevilla

Javier Tuya ‡
Universidad de Oviedo

Abstract

This coordinated project intends to introduce new methods in *software engineering project management*, integrating different quantitative and qualitative technologies in the management processes. The underlying goal to all three subprojects participants is the generation of information adapted for the efficient performance in the directing of the project. The topics that are investigated are related to the capture of decisions in dynamical environments and complex systems, software testing and the analysis of the management strategies for the process assessment of the software in its different phases of the production.

The project sets up a methodological, conceptual framework and supporting tools that facilitate the decision making in the software project management. This allows us to evaluate the risk and uncertainty associated to different alternatives of management before leading them to action. Thus, it is necessary to define a taxonomy of software models so that they reflect the current reality of the projects. Since the *software testing* is one of the most critical and costly processes directed to guarantee the quality and reliability of the software, we undertake the research on the automation of the process of software testing by means of the development of new technologies test case generation, mainly based in metaheuristic and model checking techniques in the domains of database and internet applications. The software system developed will allow the integration of these technologies, and the management information needed, from the first phases of the cycle of life in the construction of a software product up to the last ones such as regression tests and maintenance.

The set of technologies that we investigate include the use of *statistical analysis* and of *experimental design* for obtaining *metrics* in the phase of analysis, the application of the

*Email: javier.dolado@ehu.es
†Email: isabel.ramos@lsi.us.es
‡Email: tuya@uniovi.es

bayesian nets to the decision processes, the application of the standards of *process evaluation* and quality models, the utilization of metaheuristics algorithms and technologies of prediction to optimize resources, the technologies of visualization to construct control dashboards, hybrid models for the simulation of processes and others.

Keywords: Software engineering, project management, cost estimation, metrics, experimentation, project simulation, information visualization, software testing, process assessment, metaheuristics algorithms.

1 Project Overview

The IN2GESOFT project is devoted to the study of the different techniques that can be used by project leaders and developers to deal with the management decision problems in Software Engineering. The IN2GESOFT project deals with real problems that appear in software firms. The viewpoint chosen is to focus on the different processes. The basic assumption is that many of the problems in cost, quality and deadlines of software projects can be dealt with in the activities of managing products, processes and resources. The introduction of new methods in those phases of management is of paramount importance for improving the overall performance of the project.

The goal of the project is to explore new methods and technologies in the field of software management and apply them to the assessment and control of the different products, resources, quality and processes which conform the software project. The joint interest of the three IN2GESOFT subgroups is to improve the methods for innovating the methods that software project managers use for getting the project done reliably, on time and within budget.

The project is composed of three subprojects that research into different aspects of the questions about managing software projects. The subproject IN2QUANT is directed towards the quantitative and experimental issues in software engineering. The subproject IN2MED studies different methods for the definition of software management models and the subproject IN2TEST researchs into many of the quantitative and experimental questions of the testing and verification processes. The project began on 13-Dec-2004 and will end on 12-Dec-2007. Next is the brief description of each subproject. Figure 1 plots the milestones of the project with the main tasks for each subproject.

1.1 Subproject 1: IN2QUANT. *Applying Empirical Techniques and Quantitative Tools to the Effective Management of Software Projects*

Among the topics addressed by this subproject are:

- *Statistical and experimental methods in software engineering.* We use new statistical methods in experimental software engineering in order to reduce the amount of information that is used in decision making. Specifically we use *principal component analysis*, factorial analysis for selecting variables in software experiments. [Tasks *iq3.1, iq3.2, iq3.3*] Other area of interest is the use of different data mining techniques for improving the estimations performed over software project data bases. The data bases are split using

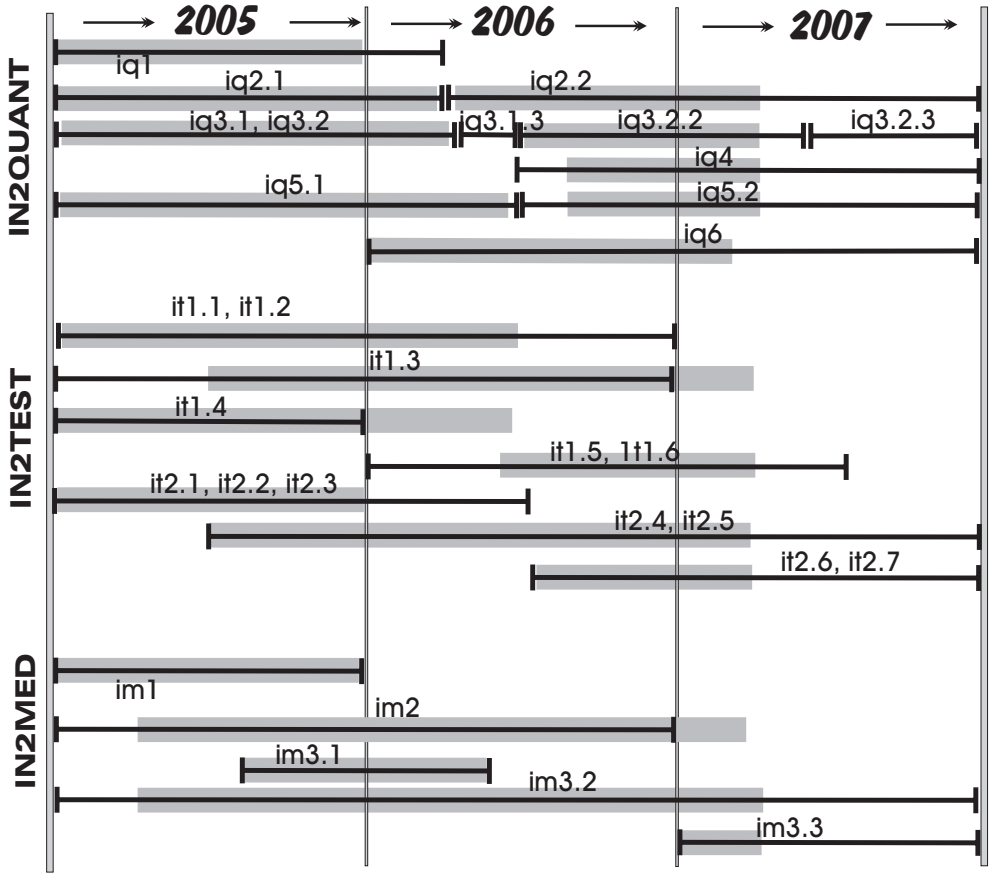


Figure 1: Milestones of the project tasks.

different algorithms in order to increase the predictions capabilities of the data sets [Task *iq3.1*].
Another line of research in this broad area is to characterise a software project data base. In this sense we try to extract the most representative instances of the data base so that the project manager can have a quick overview of the database.

- *Representation and visualization of software projects.* We have a twofold interest in this topic. On one hand there is interest in finding adequate representations for software project management and on the other hand we need to visualize many aspects of the software management process. In the first part of the topic we have worked different new representations for the Gantt charts [Task *iq2.1*]. In the second we are developing different metaphors that may visualize different aspects of the software management tasks [Task *iq2.2*].

- *Ontologies for software projects.* This is a key topic in this subproject. We are developing

ontologies for software engineering and software project management. In the case of the software engineering ontologies, the main question is how to define all terms and how to combine different specific sets of concepts in software engineering [Tasks *iq2.1, iq6*].

- *Dashboard for managing systems.* We build dashboards for the dynamic management of strategic decisions. We apply the system dynamics method for building simulators that may enhance the abilities of the project manager. Dashboards built with this method provide a flexible way of interacting with different potential actual situations [Tasks *iq2.2, iq6*].

- *Software process assessment.* We have studied different models of software process asessment, comparing different aspects of the Capability Maturity Model (CMM) and the standard SPICE (ISO/IEC TR-15504), Software Process Improvement and Capability dEtermination. The latter is used as the reference model for applying it in our research [Task *iq5*].

- *New cost estimation models and improvement of software project databases.* In this aspect we ellaborate new parametric models combining regression models with fuzzy systems and other techniques [Task *iq3*].

1.2 Subproject 2: IN2MED. *Taxonomy of models for software processes measurement and assessment*

The main objectives of the IN2MED subproject are:

- *Taxonomy of software models and definition of the processes for data collection.* Our intent is to analyze the current typology of the mostly common software projects of our industrial environment. From this analysis and by using the processes applied within their development cycle, different categories of software projects will be identified [Task *im1*].

- *Proposal for a methodological and conceptual framework to help in the decision making process regarding the management and assessment of software processes.* The proposal includes the development of machine learning techniques, prediction techniques, dynamic simulation modelling and process modelling according to the international standards of maturity models, total quality and self-evaluation of software processes [Tasks *im2, im3.1*].

- *Conceptual framework and their support tools.* The methodological framework proposed should have, among others, the following features: web application architecture, flexible enough as to accept new modules to add extra functionality regarding control elements, new variables and estimation, management and visualization procedures [Tasks *im3.2, im3.3*].

1.3 Subproject 3: IN2TEST. *Integration of automated testing techniques in the software development & maintenance process*

The topics addressed by this subproject are mainly related to Test Automation:

- *Test case generation with metaheuristic techniques.* Metaheuristic algoritms are a way to find good enough solutions for complex optimization problems. For test case generation the approach consists on represent the cost function to be minimized in terms of the fulfilment of a given coverage criteria. Tabu Search and Scatter Search are used for develop test cases using structural coverage criteria [Tasks *1.1, 1.2, 2.1*].

- *Test case generation with model checking.* Model checkers are able to verify properties in a model. If properties are not satisfied a counterexample is shown. When used for test case development the approach consists on specify the test requirements in terms of a property that is negated, and represent the program or specification under test as a model for the model checker. This approach is used for develop test cases for BPEL specifications and XPath queries [Task *1.3*].

- *Adequacy criteria for database testing.* Although many adequacy criteria have been developed for many programming languages, few work has been conducted in the development of adequacy criteria for database queries. In this case specifically tailored criteria are developed for SQL (structural and mutation based criteria) and for XPath queries [Tasks *1.4, 1.5*].

- *Integration of tools and experimentation.* Several tools to support the above approaches have been developed to support the process of test case generation and evaluation. Additionally, controlled experiments are performed to assess the fault detection ability of the techniques [Tasks *2.2 to 2.7*].

2 Level of Success of the Project

Part of the results obtained are listed in the section References, and here we detail some of the results that we have obtained. The figure 2 plots the topics that the members of the subprojects are researching.

2.1 Topics Developed in the IN2QUANT subproject

- *Characterization of software project databases* In [15] we have proposed a new method for global understanding of the software project database. We have found that it is possible to have a smaller database with less instances of the database, and this helps the project manager to deal with the data interpretation tasks.

- *Improving the parametric methods in software cost estimation.* We have applied different new methods for the problem of software cost estimation, including bayesian nets, fuzzy variables and other algorithms. Clustering has been a tool for finding also new equations for estimation. All results obtained improvements in the estimation process [14][12][13].

- *Managing projects with system dynamics.* A system dynamics model with several feedback loops represents different hypothesis for managing systems [16].

- *Representation in software projects.* The Gantt chart is augmented with new information related to the participants in the project [17].

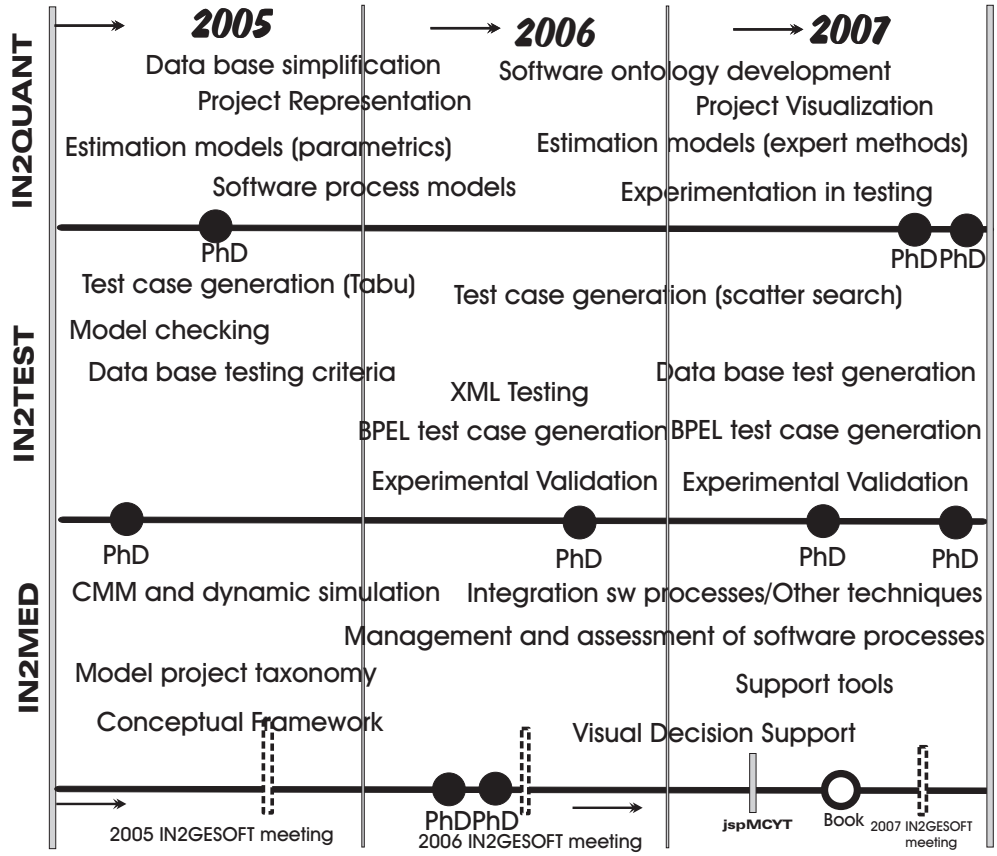


Figure 2: Topics developed and topics in development.

- *New topics in software management: visualization and coordination.* We are developing a new metaphora for the visualization of different activities in management. An initial approach has been proposed after reviewing the related literature. Also, we have presented an initial work dealing with the new problems in coordination in software projects.

- *Experimentation in software engineering.* Different experiments have been carried out and new analysis have been performed [10] [11].

2.2 Topics Developed in the IN2MED subproject

- *Simulation modelling for understanding and improving interactive systems usability.* The developed dynamic model helps visualize the behaviour of User-Centred Design (UCD) activities during the development life cycle of a software system. To model the UCD process the process model described in ISO standard 13407:1999 has been used as a reference [26] [27].

- *Integrating software process simulation models with other techniques.* The integration of continuous modelling with CMM, static and algorithmic models and the design of a metrics collection program have been collected in [22] [25]. In [23] a software project simulator based on System Dynamics is used to analyze the different sources of error in the initial estimates of parametric models.

- *Qualitative modelling and decision making in real environments.* The set of algorithms and tools previously developed have been improved and adapted to obtain qualitative models based on fuzzy rules from data coming from real environments [24].

- *Feature Selection.* We characterize several software engineering databases selecting attributes with the final aim that project managers can have a better global vision of the data they manage. The results show that in general, the smaller databases maintain the prediction capability with a lower number of attributes than the original databases [18] [15].

- *Data Streams.* Software Project Simulation has lead to massive data generation in a data streams manner. In this field, we have developed a decision support system (DDS) that is aimed to provide knowledge models from project management data. Experimental results show an excellent performance of our approach as both a general purpose method and a software project DDS, particularly [19] [21].

- *Visual Decision Support.* Visual data exploration techniques provide the user with graphic views or metaphors that represent potential patterns and data relationships. In this field, we have developed a visual exploration technique – named VETIS - that automatically extracts relevant attributes and displays their interesting measures in order to support two decision support tasks: classification and feature selection [20].

2.3 Topics Developed in the IN2TEST subproject

- *Test case generation with metaheuristic techniques.* We have explored the feasibility of using techniques (alternative to the genetic algorithms) such as Tabu Search and Scatter Search. [3]. The most outstanding result is that the local search performed by Tabu Search improves the finding at the latter stages the most difficult test cases faster than the others. On the other hand, Scatter Search has been developed to perform the same task [4]. Performance of Scatter Search improves the finding of test cases at early stages of the generation. Both approaches are then complementary and its joint use is currently being explored.

- *Test case generation with model checking.* The use of model checking has been previously used for the verification of state-based systems [5]. The research has focused on the generation of conformance test suites for BPEL compositions of web services [1] [9], using the BPEL specification as the only information. In the same line of work, but applied to data repositories, automatic generation of test cases for XML queries specified in XPath have obtained promising results.

- *Adequacy criteria for database testing.* The first line of work is the research on coverage criteria for testing SQL queries based on a structural view of the query, extended with

the tailoring of criteria based on MCDC and equivalence classes [6]. The second approach to the evaluation of coverage of SQL queries is based on mutation analysis. Mutation operators have been developed [2] and integrated in a mutation tool [7]. The Third approach is driven towards the development of test cases for semi-structured data stored in XML documents, based on Category Partition for developing test cases for XPath queries [8].

- *Tool development.* The tools for generating test cases using Tabu Search and Scatter Search, first developed separately, have been refactored to allow the generation using a common framework. A tool named SQLTest has been developed to support training and experimentation in SQL test case development, and SQLMutation to support the generation of the SQL mutants developed [7]. Both tools are publicly available and the latter one is also accessible as a web service. The development of similar tools for testing of XML and BPEL is an ongoing work.

- *Experimental validation.* A series of four controlled experiments with students have been conducted to evaluate the fault detection capability of database test cases written using different criteria. The first analysis of data collected has shown that test cases developed for SQL using the criteria above mentioned are most effective than conventional approaches such as equivalence portioning.

3 Outcomes of the project and main results

Table 1 summarizes the results, so far, of the project. The figures present the works and activities in which the members of the project have taken part in.

One of the most important results of the project, from the perspective of the heads of the subprojects, is the book **Técnicas Cuantitativas para la Gestión en la Ingeniería del Software**, that has 16 chapters, around 400 pages and has a participation of 35 authors. It will be published during 2007. The book is the expression of the many techniques in which the members of the project are experts at. The content of the book describes topics such as: *software process assessment, CMM, SPICE, process simulation, project representation, cost estimation by machine learning, system support for project management, principal component analysis, experimental software engineering, bayesian nets, software estimation by fuzzy methods, data mining for decision support, software test methods and software test management.*

3.1 Publications in Journals and Books

The References show part of the articles that have been published by the members of the project. The journals targeted have been, among others: Journal of Systems and Software, Information and Software Technology, Computers and Operations Research, International Journal of Engineering Intelligent Systems for Electrical Engineering and Communications, SIGPLAN Notices, IEEE Latin America Transactions, Procesos y Métricas de las Tecnologías de la Información, etc.

3.2 Conferences

The conferences, workshops and seminars in which the results have been presented are, among others:

[2005] ADIS 2005, CAISE05, IWANN05, SACACM05, WIST05, WASVWS05, IDA05, PROSIM05, IWSM05, GECCO05, IMWA05, IADIS 2005, MIC 2005, JISBD 2005

[2006] ADIS 2006, JISBD 2006, ICEIS06, ICEISDC06, ICWIST06, WSTV06, KES06, SACACM06, ICCI06, CISTI06, DS06, SQM06, Mutation 2006, ICSEA 2006, TAIC-PART 2006, WS-MaTe 2006, ICKEDS 2006, STV 2006, RedBD 2006, PRIS 2006, JSWEB 2006, ONTOSE 2006.

3.3 Events organized

ADIS 2005, PRIS 2006, ADIS 2006, the Session *Ontologies for Software Engineering* (SEKE06), and the International Conference on Software Processs and Product Measurement MENSURA 2006.

3.4 Research Grants

The subproject IN2QUANT has a researcher under the FPI-MEC grant associated to the project.

3.5 Ph.D. Dissertations

The following researchers, members of IN2GESOFT, have defended their corresponding Ph.D.:

- María José Suárez Cabal, Mejora de casos de prueba en aplicaciones con bases de datos utilizando medidas de cobertura de sentencias SQL (2006)
- María Eugenia Díaz Fernández, Generación automática de pruebas estructurales de software mediante Búsqueda Tabú (2005).
- Roberto Ruiz Sánchez, Heurísticas de Selección de Atributos para Datos de Gran Dimensión (European Thesis) 2006.
- Francisco Ferrer Troyano, Aprendizaje Incremental y Modelado Visual de Reglas de Decisión (2006).
- Antònia Mas Pichaco, Un Nuevo Modelo para la Implantación de un Sistema de Gestión de Calidad en Pymes de Desarrollo de Software basado en SPICE (ISO/IEC 15504) (2005).

Two members obtained during 2005 the special prize for PhDs at their corresponding universities

- Claudio de la Riva, Premio Extraordinario de Doctorado (2005).
- Mari Carmen Otero Vidal, Premio Extraordinario de Doctorado (2005).

3.6 Collaborations with groups and other projects

There have been several collaborations with researchers at other universities: Université du Québec at Montréal and University of Reading at UK.

In 2005 a Marie Curie EU International Incoming Fellowship was submittend and it was approved in 2006. The recipient is *Prof. Alain Abran* from the Université du Québec à Montréal who will visit the University of the Basque Country for the next two years. The members of the project have promoted the following activities: RePRIS (a network for software testing), and projects OPTISOFT and HEMPIC2.

Type of contribution	**Number**
International Journals and book chapters	32
Book chapters	7
National Journals	12
Ph.D. Dissertations	5
International Conferences	55
National Conferences	26
International Program Committees	9
National Program Committees	6
Events organized	5
PhD Students under FPI grants	1

Table 1: Current results of the project.

References

[1] José García-Fanjul, Claudio de la Riva and Javier Tuya, Testing BPEL compositions of web services using SPIN. Journal of Systems and Software, (Under revision, selected from WS-MaTe 2006).

[2] Javier Tuya, Mª José Suárez-Cabal, Claudio de la Riva, Mutating database queries. Information and Software Technology, (In press, July 2006)

[3] Eugenia Díaz, Javier Tuya, Raquel Blanco, Javier Dolado, A tabu search algorithm for Structural Software Testing. Computers and Operations Research (focus issue on Search-Based Software Engineering), (Accepted).

[4] Raquel Blanco, Javier Tuya, Eugenia Diaz, Adenso Diaz, A Scatter Search approach for the Automated Branch Coverage in Software Testing. International Journal of Engineering Intelligent Systems, (Accepted).

[5] Claudio de la Riva, Javier Tuya, Automatic Generation of Assumptions for Modular Verification of Software Specifications. Journal of Systems and Software, 79 (9), pp. 1324-1340, September 2006.

[6] Javier Tuya, Mª José Suárez-Cabal, Claudio de la Riva, A practical guide to SQL white-box testing. SIGPLAN Notices, 41(4), pp. 36-41, April 2006.

[7] Javier Tuya, Mª Jose Suarez-Cabal and Claudio de la Riva, SQLMutation: a Tool to Generate Mutants of SQL Database Queries. Second Workshop on Mutation Analysis (Mutation2006). Raleigh, North Carolina, November 2006.

[8] Claudio de la Riva, José García-Fanjul, Javier Tuya, A Partition-based Approach for XPath Testing.International Conference on Software Engineering Advances. Tahiti, October 2006.

[9] José García-Fanjul, Claudio de la Riva, Javier Tuya, Generation of Conformance Test Suites for Compositions of Web Services Using Model Checking.Testing: Academic & Industrial Conference - Practice and Research Techniques (TAIC-PART). Windsor, Agosto 2006.

[10] M.C. Otero and J.J. Dolado, An Empirical Comparison of the Dynamic Modeling in OML and UML, Journal of Systems and Software, vol. 77, 91-102, 2005.

[11] D Rodríguez, MA Sicilia, JJ Cuadrado, D Pfahl, e-Learning in Project Management using Simulation Models: A Case Study based on the Replication of an Experiment, IEEE Transactions on Education, 2006. IEEE Education Society.

[12] Cuadrado-Gallego, J.J., Fernández, L., Sicilia, M.A.. Enhancing input value selection in parametric software cost estimation models through second level cost drivers. Software Quality Journal, 2006.

[13] Rodríguez, D., Cuadrado-Gallego, J.J., Garre, M., Sicilia, M.A. Multiattribute clustering in software effort estimation parametric models. Journal of Computer Science Technology, 2006.

[14] Cuadrado-Gallego, J.J., Sicilia, M.A., Rodríguez, D., Garre M. An empirical study of process-related attributes in segmented software cost-estimation relationships. Journal of Systems and Software, 2006, 3 (79), 351-361.

[15] Dolado, J., Rodríguez, D., Riquelme, J., Ferrer, F., Cuadrado-Gallego, J.J. A two stage zone regression method for global characterisation of a project database. In Advances in Machine Learning Application in Software Engineering, Idea Group Publishing, 2007, 0-0. ISBN: 1-59140-941-1.

[16] Morlán, I., Nieto, M., Davalillo, A., Dolado, J. A System Dynamic model to facilitate understanding and building stakeholder support for new management strategies in universities 9th Toulon-Verona Conference. Quality in higher Education, Health Care and local Government, Paisley. Escocia (Reino Unido), Septiembre, 2006

[17] Fran J Ruiz-Bertol, Javier Dolado: Human resource representation in Gantt charts, Software Quality Management XIV: Perspectives in Software Quality (SQM 2006), pp. 219-230. 2006.

[18] Giráldez Raúl, JS Aguilar-Ruiz, Riquelme JC. Knowledge-based Fast Evaluation for Evolutionary Learning. IEEE Systems, Man and Cybernetics Part C. Vol. 35, nº 2, pp. 254-261, 2005. Ed. IEEE Press ISSN 0018-9472.

[19] Ferrer-Troyano Francisco, Aguilar-Ruiz JS, Riquelme JC. Incremental Rule Learning and Border Examples Selection from Numerical Data Streams. Journal of Universal Computer Science. Vol. 11, nº 8, pp. 1426-1439, 2005. Ed. Know-Center Springer ISSN 0948-695x.

[20] Ferrer-Troyano Francisco, Aguilar-Ruiz JS, Riquelme JC. Connecting Segments for Visual Data Exploration and Interactive Mining of Decision Rules. Journal of Universal Computer Science. Vol. 11, nº 11, pp. 1835-1848, 2005. Ed. Know-Center Springer ISSN 0948-695x.

[21] Giráldez, R. "Improving the performance of Evolutionary Algorithms for Decision Rules Learning", AI Communications, Vol. 18 (1), pp. 63-65, 2005.

[22] Ruiz M., Ramos I., Toro M. Software Process Dynamics: Modeling, Simulation and Improvement. New Trends in Software Process Modelling. Vol. 18, pp. 21-56, World Scientific Publishing Company, 2006.

[23] Mercedes Ruiz Carreira, JJ Cuadrado Gallego, D Rodríguez García, MA Sicilia. Using Simulation to determine the sensitivity of error sources. Internacional Workshop on Sofware Measurement (IWSM´05), pp. 1-10, 2005, Montreal (Canada).

[24] de Toro F.J, Aroba J, López J.M. Fuzzy-Genetic Methodology for web-based computed-aided diagnosis in medical applications. IX International Conference on Computacional Intelligence. Dortmunt (Alemania), 2006.

[25] de la Villa M, Ruiz-Carreira M, Ramos I. Un estudio crítico comparativo de ISO 9001, CMMI e ISO 15504. Conferência Ibérica de Sistemas e Tecnologias de Informaçao. Ofir (Portugal). Junio 2006.

[26] Nuria Hurtado Rodríguez, Mercedes Ruiz Carreira, Jesús Torres Valderrama. Modelado y simulación como herramienta para la mejora de la usabilidad. Actas del VI Congreso de interacción persona-ordenador, pp. 1-4, 2005, Granada.

[27] Nuria Hurtado Rodríguez, Mercedes Ruiz Carreira, Jesús Torres Valderrama. Towards Interactive Systems Usability improvement through simulation modeling. 6th International Workshop on Software Process Simulation and Modeling (PROSIM´05), pp. 117-125, 2005, San Louis, MO (USA).

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Development of Distributed Systems with Wireless Devices for Dynamic Environments using Agent Technology TIN2004-07033

Emilio Corchado Rodríguez*
Universidad de Burgos

Abstract

The objective of this project is to define an architecture which facilitates the construction of autonomous agents that function efficiently and independently of a machine. Agents and multi-agent systems are used in a wide variety of areas. Many architectures have been proposed for their construction, however, in this case, we are focusing on the development of agents with autonomous decision-making capacities that are based on learning mechanisms and planning and re-planning during execution.

More specifically, our project focuses on the development of agents in distributed wireless systems for industrial control applications, where significant restrictions on time mean that communication problems are likely to occur. Given the limited calculus potential of current mobile devices, such systems have to be lightweight and efficient. Within the framework of this project, interface mechanisms are being developed that allow inter-agent and user-agent interaction, within a multi-agent architecture.

Reasoning models based on the experience that the team has in this field are being developed and applied in the production of various prototypes in order to validate the proposed architecture, in collaboration with those companies that have expressed an interest in the project and its results to date.
Keywords: Reasoning, Communication, Multiagent Systems and Wireless Systems

1 Objectives

The principal objective of the project is the development of a distributed architecture for the construction of wireless systems. Over the last two years, 2005-2006, work has progressed on the development of a deliberative-agent model for wireless devices that allows interaction between the agents. To achieve this aim, the Tasks defined in the project application have been completed:

- T1: Definition/identification of an agent architecture for the construction of systems that work in real time in dynamic environments and that can communicate between wireless devices.

* Email: escorchado@ubu.es

- T2: Development of an architecture that allows autonomous agents to be constructed that are able to carry out planning tasks and have the capacity for wireless communication.

The following activities have been undertaken to develop the above tasks:

T1: Activities

T1.1 Revision of the current proposals
T1.2 Analysis and design of reasoning mechanisms for deliberative agents
T1.3 Identification of a planning and replanning strategy in execution time
T1.4 Implementation and validation of the reasoning mechanism and the planning and replanning strategy
T1.5 Prototype construction

T2: Activities

T2.1 Revision of current proposals
T2.2 Tool design
T2.3 Tool implementation
T2.4 Validation in various test situations

These activities have been successfully developed within the time-frame set out in the project application, with only slight modifications:

	Year 2005											
Months	1	2	3	4	5	6	7	8	9	10	11	12
T1.1 (1,2,3,4)												
T1.2 (4,5,6,7,8)												
T1.3 (6,7.8.9)												
T1.4 (4,5,6,7,8,9,10)												
T1.5 (9,10,11,12)												

	Year 2006											
Months	13	14	15	16	17	18	19	20	21	22	23	24
T2.1 (13,14,15,16)												
T2.2 (16,18,19,20)												
T2.3 (18,19,20,21,22)												
T2.4 (19,2021,22,23,24)												

Technical Resources:

PDAs: 1 TSM400 (on loan from GCI), 1 TSM500 (on loan from GCI)
and 1 QTEK 9090 (on loan from GCI), 1 HP 6515, 4 HP hx2700, 1 PALM TX
3 PCs
1 Printer

Personnel:

Task 1 (T1)
Lead Researcher: E. S. Corchado, Colin Fyfe
Researchers:
Jose Luis Gómez Soto
José Manuel Sáiz Díez
Francisco José Hurtado Ramírez
Juan Carlos Pérez Córdoba
Jacinto Canales del Caso
Jesús Manuel Maúdes Raedo
Juan José Rodríguez Diez
Pedro Burgos Lázaro
Álvaro Herrero Cosio
Bruno Baruque Zanón

Task 2 (T2)
Lead Researcher: Rodolfo Zunino, E. S. Corchado
Researchers:
Raúl Marticorena Sánchez
Carlos López Nozal
Francisco José Hurtado Ramírez
José Manuel Sáiz Díez
Ángel Arroyo Puente
M. Belén Vaquerizo García
Cesar I. García Osorio
David H. Martín Alonso
Dymitr Ruta
Assumpció Pellicer Figueras

Leticia Curiel Herrera. Álvaro Herrero Cosio
Bruno Baruque Zanón

A number of technologists - Assumpció Pellicer Figueras, Álvaro Herrero Cosio, Bruno Baruque Zanón and Leticia Curiel Herrera - have recently joined the GICAP research group, responsible for launching the current project, and are actively collaborating in its development. They are all doctoral students whose theses are closely linked to the results of the research work that is under development in the present project.

2 Project achievement level

The project milestones have successfully been reached within the agreed time-frame. The following milestones have already been or will be achieved over the coming months:

- M1 Reasoning Model for agents - Month 8
- M2 Planning strategy for agents - Month 10
- M3 Prototype - Month 12
- M4 Design of a tool for constructing agents – Month 22 (2 months behind schedule)
- M5 Construction of a tool for developing agents – Month 28 (4 months behind schedule)

We have successfully developed a distributed architecture for constructing deliberative agents in wireless networks, the principal characteristic of which is the use of CBP-BDI (Case-Based Planning – Belief-Desire-Intention) agents (Milestone M1). These deliberative agents incorporate a reasoning CBP (Case-Based Planning) engine, a variant of the CBR (Case-Based Reasoning) [1] system, which allows the agents to learn from initial knowledge, autonomously interact with the environment and users, and adapt through knowledge discovery to environmental changes (Milestone M2). A first prototype has been developed for guiding students at the University of Burgos [P21] (Milestone M3). The models developed to date have been modified for use in wireless environments during the second year. The design of a library for constructing deliberative agents operating in wireless networks was completed two months behind schedule during the second semester of 2006 (June to December). This library should allow the constructions of agents for mobile devices such as PDAs or phones (Milestone M3). Construction of the library is currently underway on the basis of that design and we are developing a tool for constructing deliberative agents that are able to communicate between each other. It is therefore estimated that Milestone 5 will be reached by April 2007.

The option chosen to define an appropriate analysis and design methodology for the problem to be resolved is one that is intended to combine the advantages of both Gaia [3] and (Agent UML) AUML [4, 5, 6]. Through the Gaia analysis, two models are obtained: the role model and the interaction model. In the Gaia design process three models are considered - the agent model, the services model and the acquaintance model [3] – which are combined with a low-level AUML design that is sufficiently detailed to carry out an implementation. The AUML design provides class diagrams for each agent, collaboration or sequence diagrams for each interaction, state and activity diagrams to represent internal states and protocol diagrams to model communicative acts. The agents are constructed with .NET for windows mobile wireless devices and with J2ME and MIDP 2.0 for the Symbian OS. FIPA and KQML are the selected communication platforms. This is one of the prototypes used to validate our work and is being tested in the construction of an agent-based system for use in a Geriatric Residence. Three firms - Tecsidel, Telefónica I+D and GCI -

have been actively involved in the development process. They have provided the required infrastructure for the development of the WI-FI network, as well as PDAs and some RFID mechanisms. The previously mentioned companies are also assisting with two other prototypes that are under construction. One is an intrusion detection agent-based system [P8, P9, P15, P16, P22, P24] and the other is an autonomous agent for drilling [P7, P13].

2.1 Architecture for a reasoning and planning agent

The research team has developed a deliberative agent using a CBR [1] architecture. This allows it to respond to events, take goal-based initiatives, communicate with other agents, interact with users, and identify the best plans to achieve its goals based on past experience. The particular agent in question uses a special type of CBR system which is called a CBP system, especially designed for planning construction, and is also a deliberative agent that works at a high level with the BDI (Belief, Desire and Intention) concepts [2]. It has learning and adaptation capabilities, which facilitate its work in dynamic environments. A CBP-BDI uses case-based reasoning as a reasoning mechanism, which allows it to learn from initial knowledge, to interact autonomously with the environment as well as with users and other agents within the system, providing it with a great capacity for adaptation to the needs of its surroundings.

To test the proposed system, we have developed an agent-based system [P23] that provides efficient working schedules, in execution time, for the staff of geriatric residences, which is designed to improve health care quality and patient supervision. Each of the agents is assigned to either a nurse or a doctor in the residence and provides them with information on patients, locations, case history as well as alerts in case of emergencies. As the members of staff are carrying out their duties (following the work plan indicated by the agent) the latter may require modification due, for example, to delays or alarms, in which case the agent is capable of re-planning in execution time. The internal structure and capabilities of the agents are based on the mental aptitudes of beliefs, desires and intentions. This high-level structure facilitates the incorporation of a CBR system as a deliberative mechanism within BDI agents, facilitating learning and adaptation and providing a greater degree of autonomy than is provided by the BDI architecture alone.

In order to introduce a CBR motor into a BDI agent, the cases used in the CBR system are represented in terms of beliefs, desires and intentions, and a CBR cycle is implemented. A case is a past experience composed of three elements: an initial state or problem description that is represented as a belief; a final state that is represented as a set of goals; and the sequence of actions that makes it possible to move from an initial to a final state. This sequence of actions is represented in terms of intentions or plans. In a planning agent, the reasoning motor generates plans using past experiences and planning strategies, from which we obtain the CBP concept, which consists of four sequential stages: the retrieval stage, in which the past experiences that show most similarity to current one are retrieved; the re-use stage, in which the retrieved solutions are combined to obtain a new optimal solution; the revise stage, in which the solution is evaluated; and the retain stage in which the systems learns from the new experience.

The internal structure of a CBP-BDI [P19, P23] agent is composed of beliefs, desires and intentions. The problem description (initial state) and the solution (situation in which the final state

is achieved) are represented as beliefs, the final state as a goal (or set of goals), and the sequences of actions as plans or intentions. The CBP cycle is implemented through goals and plans. When the goal corresponding to one of the CBP stages is triggered, different plans (algorithms) can be simultaneously executed to achieve the goal. Each plan may trigger new sub-goals and, consequently, cause the execution of new plans. This is an autonomous agent that can survive in a dynamic environment precisely because it incorporates this planning and reasoning mechanism.

Artificial intelligent neuro-symbolic models have also been developed to construct the deliberative agents. These models allow the agents to reason, plan and re-plan in execution time, allowing them to work in time-restricted environments. We have developed a model that incorporates a Connectionist Maximum Likelihood Hebbian Learning Model [P2] within a CBR System. It is a robust replanning system, which is easily used in wireless devices as it is extremely lightweight, the main feature of which is that it allows dynamic learning to take place in a non-supervised and automatic way. We have also been analysing different alternatives that involve the use of Maximum Likelihood Hebbian Learning applied to the Scale Invariant Map [P1], Maximum Likelihood Topology Preserving Ensembles [P20] and Outlier Resistant PCA Ensembles [P25] in the construction of light-weight reasoning algorithms for wireless agents.

The main difficulties that have been encountered relate to the implementation of the physical system, which are due to the architectural diversity of the wireless devices, the poor compatibility between environments, and the high price of the technology required to construct the testing environments; all of which are elements that have delayed the achievements of Milestones 4 and 5.

3 Indicators of results

Four new doctoral students, who joined the research group following the approval of the project, now form part of the project team:

Name	Transfer	Title of Thesis	Estimated date of presentation
Álvaro Herrero Cosio	Yes	Mobile Connectionist Agent Based Agent for Intrusion Detection Systems	2008
Bruno Baruque Zanón	Yes	To be determined	2008
Leticia Curiel Herrera.	Yes	Autoagent systems to Automate Drilling Tasks	2008
Assumpció Pellicer Figueres	Yes	To be determined	2008

This project has contributed to strengthening the relations between GICAP and the following international research groups:

Group	Lead Researcher	URL
SEALab, labor. of the Depart. of Biophysical & Electro. Engine-DIBE) Genoa University.	Rodolfo Zunino	http://www.sealab.dibe.unige.it/
Computational Intelligence Research Group.	Bogdan Gabrys	http://dec.bournemouth.ac.uk/cgi-bin/staff.pl?User=bgabrys

Work proceeds within the framework of these collaborative relations, on the development of neuro-symbolic systems to improve the reasoning capability and autonomy of intelligent agents and multiagents.

We are also working in collaboration with the following Spanish research groups:

Group	Lead Researcher	URL
Bisite. Univ. de Salamanca	Juan Manuel Corchado	http://bisite.usal.es/index.php?lang=en
Ingeniería de Organización. Universidad de Burgos	R. del Olmo Martínez Lourdes Sáiz Bárcena	http://ubu.es/investig/grupos/OE-1.htm
Grupo Integridad Estructural (GIE) de la Univ. de Burgos.	Jesús Manuel Alegre Pedro Bravo Díez	http://ubu.es/investig/grupos/ing_const/mmc-2/index.html
Grupo de Tecnología de los Alimentos. Univ. de Burgos	Jordi Rovira Carballidos	http://www.ubu.es/investig/grupos/TA-1.htm

Collaborative research with the first group is focused on the development of reasoning systems, agents and multiagent systems, while collaboration with the other three groups is oriented towards the application of multiagent architectures and neuro-connectionist systems to improve the reasoning capability and the autonomy of intelligent agents and multiagent systems, and the possibility of applying our systems to different real-life problems, making it possible to work in a multidisciplinary environment.

It is important to emphasise that GICAP is a member of the following thematic networks:

Nature-inspired Smart Information Systems. NISIS. Funded by the European Commission. http://www.nisis.risk-technologies.com/Default.aspx?AspxAutoDetectCookieSupport=1
AgentCities.es, Creación de un Entorno Innovador para la Comunicación de Agentes Inteligentes. http://grusma2.etse.urv.es/AgCitES/
Red Española de Minería de Datos y Aprendizaje Automático. http://www.lsi.us.es/redmidas/

As a result of all these collaborative networks, GICAP organized and planned the following scientific conferences and workshops:

7th International Conference on Intelligent Data Engineering and Automated Learning. September, 20-23; 2006. Burgos, Spain. Proceedings. Lecture Notes in Computer Science 4224 Springer 2006, ISBN 3-540-45485-3 http://www2.ubu.es/ideal2006/
Hybrid Artificial Intelligence Systems Workshop (HAIS'06) within the framework of: IBERAMIA/SBIA/SBRN 2006. http://www.icmc.usp.br/iarn2006/coevents/hais.php

The following workshops and special sessions are planned and have been accepted for 2007:

KES 2007 Invited Session: **Hybrid Artificial Intelligence Systems Workshop (HAIS'07.-KES2007)** Organizers: E. Corchado (University of Burgos),J. Corchado (University of Salamanca). Within the framework of the KES-2007 Conference; 11th International Conference on Knowledge-Based & Intelligent Information & Engineering Systems; Italy; 12-14 Sept. 2007. http://ubu.es/investig/grupos/ing_const/lsi-1/HAIS07v2.html
KES 2007 Invited Session: **Artificial Intelligence Applications in Security, 2007; AIAS** 2007 Organizers: Emilio Corchado (University of Burgos, E), Rodolfo Zunino (University of Genoa). Within the framework of the KES-2007 Conference; 11th Inter. Conference on Knowledge-Based & Intelligent Information & Engineering Systems; Italy; 12-14 Sept. 2007 http://ubu.es/investig/grupos/ing_const/lsi-1/AIAS07.html
IWANN 2007 Invited Session **Nature-inspired intelligent methods and applications** Organizers: Emilio Corchado (Univ. of Burgos, E), Bogdan Gabrys (Bournemouth Univ., UK) Within the framework of the 9th International Work-Conference on Artificial Neural Networks (IWANN'2007) San Sebastián (Spain) June 20-22, 2007: http://iwann2005.ugr.es/2007/special.html

The publications obtained under the support of this project are the followings:

P1. Corchado E., Fyfe C. Application of Maximum Likelihood Hebbian to the Scale Invariant Map. Computing and Information Systems Journal. UK. ISSN: 1352-9404. May 2005.Vol 9, No. 2. Pag: 46–51.

P2. Juan M. Corchado, Jim Aiken, Emilio S. Corchado, Florentino Fdez-Riverola. Evaluating air-sea interactions and fluxes using an instance-based reasoning system. AI COMMUNICATIONS (Journal Citation Report 2005 0.612). 2005. Volume 18, Number 4 / 2005; pp. 247 - 256

P3. Corchado E., Herrero A., Baruque B., Saiz J. M. Intrusion Detection System Based on a Cooperative Topology Preserving Method. 7th International Conference on Adaptive and Natural Computing Algorithms; ICANNGA 2005. Proceedings of the 7th International Conference on Adaptive and Natural Computing Algorithms; ICANNGA 2005. ISBN = 3-211-24934-6. Springer Verlag. Coimbra (Portugal). March 2005.

P4. Corchado E., Curiel L., Bravo P. A Cooperative Unsupervised Connectionist Model to Identify the Optimal Conditions of a Pneumatic Drill. Proce. of the 4th IEEE international workshop on Soft Computing as Transdiscip. Science and Technology. Springer Verlag. ISBN-10 3-540-25055-7. Japan. May; 2005.

P5. Emilio Corchado, Álvaro Herrero, José Manuel Sáiz. A cooperative connectionist IDS model to identify independent anomalous SNMP situations. Proceedings of the "III TALLER DE MINERÍA DE DATOS Y APRENDIZAJE (TAMIDA'2005); I CONGRESO ESPAÑOL DE INFORMÁTICA (CEDI'2005)". Ed. Thomson ISBN 84-9732-449-8. Eds: R. Ruiz, J. C. Riquelme, J. S. Aguilar-Ruiz. Granada-Spain, Sept. 2005.

P6. Corchado E., Curiel L., Bravo P. An Application of an Unsupervised Projection Cooperative Model to the Detection of the Best Drilling Conditions. Proce. of the "I Simposio De Inteligencia Computacional; I Congreso Español De Informática (Cedi'2005)". Thomson ISBN: 84-9732-444-7. Granada-Spain, Sept. 2005.

P7. Leticia Curiel, Emilio Corchado, Pedro Bravo, Álvaro Herrero. Collaborative Agents for Drilling Optimisation Tasks Using an Unsupervised Connectionist Model. IWPAAMS'2005. Proceeding of the 4th International Workshop on Practical Applications of Agents and Multiagent Systems (IWPAAMS 2005). ISBN: 84-9773-222-7. Ed.: Universidad de León. León, Spain, 20- 21 October, 2005.

P8. Álvaro Herrero, Emilio Corchado, José Manuel Sáiz, Leticia Curiel. A Proposal: Distributed Agent-Based IDS to Detect Anomalous SNMP Situations Using Unsupervised Learning. IWPAAMS´2005. Proceeding of the 4th International Workshop on Practical Applications of Agents and Multiagent Systems (IWPAAMS 2005). ISBN: 84-9773-222-7. Ed.: Universidad de León. León, Spain, 20- 21 October, 2005.

P9. Emilio Corchado, Álvaro Herrero, José Manuel Sáiz. A Feature Selection Agent-based IDs. NiSIS 2005, 1st European Symposium on Nature-inspired Smart Information Systems. Portugal (Algarve)

P10. Álvaro Herrero, Emilio Corchado, José Manuel Sáiz. Identification of Anomalous SNMP Situations Using a Cooperative Connectionist Exploratory Projection Pursuit Model. Proceeding of the Sixth International Conference on Intelligent Data Engineering and Automated Learning; IDEAL 2005. Springer, Series: Lecture Notes in Computer Science , Vol. 3578. ISBN-10: 3-540-26972-X. ISBN-13: 978-3-540-26972-4. Brisbane, Australia, July 6-8, 2005. ISI 2005. (Journal Citation Report 2005: 0.402.)

P11. E. Corchado, A. Herrero, J. M. Sáiz. An Unsupervised Cooperative Pattern Recognition Model to Identify Anomalous Massive SNMP Data Sending. First International Conference on Natural Computation (ICNC'05) actas del congreso. Lecture Notes in Computer Science (LNCS), Springer Verlag, ISBN-10: 354028320X, ISBN-13: 978-3540283201. August 27-29, 2005, Changsha, China. ISI 2005, JCR: 0.402.

P12. E. Corchado, A. Herrero, J. M. Sáiz. Detecting Compounded Anomalous SNMP Situations Using Unsupervised Pattern Recognition. Procee. of the 15th International Conference On Artificial Neural Networks. ICANN 2005. Lecture Notes in Computer Science (LNCS), Springer Verlag. ISBN-10: 3540287523 ISBN-13: 978-3540287520 Warsaw, Poland. Sept.11-15,2005. ISI 2005 (J.CR. 2005: 0.402).

P13. Curiel L., Corchado E., Bravo P. A MultiAgent System to Automate Drilling Tasks. Proceedings of the V International Workshop on Practical Applications of Agents and Multiagent Systems. IWPAAMS'2006. ED. Universidad de Valladolid, ISBN: 84-934181-7-X. Univ. de Valladolid. Segovia, Spain. October 19-20; 2006.

P14. Bruno Baruque, Bogdan Gabrys, Emilio Corchado, Álvaro Herrero, Jordi Rovira, Javier González. Outlier overcoming using re-sampling techniques. Proceedings of the V International Workshop on Practical Applications of Agents and Multiagent Systems. IWPAAMS'2006. ED. Universidad de Valladolid, ISBN: 84-934181-7-X. Universidad de Valladolid. Segovia, Spain. October 19-20; 2006.

P15. A. Herrero, B. Baruque y E. Corchado A Multiagent System for ID: MOVICAB-IDS. Proce. of the V Internat. Workshop on Practical Applications of Agents and Multiagent Systems. IWPAAMS'2006. ED. Univ. de Valladolid, ISBN: 84-934181-7-X. Univ. de Valladolid. Segovia, Spain. Oct.19-20; 2006.

P16. A. Herrero y E. Corchado. Mobile Hybrid Intrusion Detection System. Proce. of the Hybrid Artificial Intellig. Systems Workshop (HAIS'06),IBERAMIA/SBIA/SBRN Int. Joint Conf. 2006. ISBN 85-87837-11-7.
P17. A. Herrero, E. Corchado, Bruno Baruque. The Importance of Time in the Identification of Anomalous Situations by Means of MOVICAB-IDS. Proceedings of the 2nd European Symposium on Nature-Inspired Smart Information Systems (NiSIS Symposium 2006). 29th Nov. – 1st December 2006. Tenerife. Spain.
P18. B. Baruque, E. Corchado, B. Gabrys, A. Herrero, J. Rovira y J. González. Unsupervised Ensembles Techniques for Visualization. Proceedings of the 2nd European Symposium on Nature-Inspired Smart Information Systems (NiSIS Symposium 2006). 29th November – 1st December 2006. Tenerife. Spain.
P19. E. Corchado, M. A. Pellicer and J.M. Corchado, A Guiding Agent: Smart Dynamic Technology for Solving Distributed Problems. Proc. of the 2nd European Symposium on Nature-Inspired Smart Information Systems. 29th Nov. – 1st Dec. 2006., Tenerife. Spain.
P20. E. Corchado, B. Baruque, B. Gabrys. Maximum Likelihood Topology Preserving Ensembles. Proceedings of the Intelligent Data Engineering and Automated Learning Conference- IDEAL2006. Springer – LNCS, vol. 4224, 2006. ISBN: 978-3-540-45485-4. ISI 2005, JCR: 0.402. Burgos, Spain. September, 2006
P21. A. Herreo, B. Baruque y E. Corchado. Sistema de guiado inteligente para estudiantes universitarios, Informe Técnico, DIC-2006.09-2, Departamento de Ingeniería Civil. Universidad de Burgos. 2006.
P22. A. Herrero, E. Corchado y J.M. Sáiz. MOVICAB-IDS: Visual Analysis of Network Traffic Data Streams for Intrusion Detection". Proce. of the Intelligent Data Engineering and Automated Learning - IDEAL2006. Springer – LNCS, vol. 4224, 2006. ISBN: 978-3-540-45485-4. ISI 2005, JCR: 0.402. Burgos, Spain. Sep., 2006.
P23. E. Corchado, A. Herrero, L. Curiel. Autonomous Agent for Geriatric Centres. Informe Técnico, DIC-2006.11-1, Departamento de Ingeniería Civil. Universidad de Burgos. 2006.
P24. E. Corchado, A. Herrero y J.M. Sáiz. Testing CAB-IDS through Mutations: on the Identification of Network Scans. Procee. of the 10th Inter. Conf. on Knowledge-Based Intelli. Information and Engin. Systems, KES 2006. Springer – LNAI, vol.4252, ISBN: 978-3-540-46537-9. ISI 2005 JCR: 0.302. UK. Oct., 2006.
P25. B. Gabrys, B. Baruque, E Corchado. "Outlier Resistant PCA Ensambles". Proce. of the 10th Intern. Conference on Knowledge-Based Intelligent Information and Engineering Systems, KES 2006. Springer – LNAI, vol. 4252, 2006. ISBN: 978-3-540-46537-9. ISI 2005 JCR: 0.302. Bournemouth, UK. Oct., 2006.

4 References

[1] A. Aamodt and E. Plaza, Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches, AI Communications 7 (1994) 39-59.
[2] M.E. Bratman, Intentions, Plans and Practical Reason (Harvard University Press, Cambridge, M.A., 1987).
[3] M. Wooldridge, and N.R. Jennings, and D. Kinny. The Gaia Methodology for Agent-Oriented Analysis and Design, Journal of Autonomous Agents and Multi-Agent Systems, 3 (3) (2000) 285-312.
[4] B. Bauer M.P. Huget, FIPA Modelling: Agent Class Diagrams.
[5] J. Odell, R. Levy, and M. Nodine. FIPA Modelling: Agent Class Superstructure Metamodel. FIPA meeting and interim work.
[6] J. Odell and M.P. Huget. FIPA Modelling: Interaction Diagrams.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

GPS: Plataforma de Gestión de Procesos Software: Modelado, Reutilización y Medición TIN2004-07083

Paloma Martínez [*]
Departamento de Informática,
Universidad Carlos III de Madrid

Antonio Amescua[**]
Departamento de Informática,
Universidad Carlos III de Madrid

Abstract

The main project purpose is the design and development of a set of software tools to support Software Engineering Best Practices, defined by international standardization organizations such as ISO, SEI and PMI. These tools will allow the definition of software processes and will assist in the software project management based on Best Practices Models. In order to get the project effectively, it is necessary to develop formal specifications language to represent processes and also an advanced system to store, classify and retrieve assets and artifacts, as work products, project planning, requirement documents, metrics, risks, etc. This document describes the project status after near two years of work as well as the main results obtained till now.
Keywords: software processes, metrics, reuse, software process metamodels, repositories.

1 Project Goals and resources

Different authors argue that a high percentage of failures in software development are not due to bad capabilities of technicians but to deficiencies in used methods. Thereby, to enhance the productivity and quality of software intensive organizations is essential to introduce seamlessly efficient management and engineering practices. As these practices are institutionalized, the organizations will be able to achieve some important incomes.

We have found that the most effective way to introduce improvements in an organization consists of implementing them as efficient practices applied to the organization's projects. One of the difficulties to implement this way of software process improvement is due to the need of a knowledge management framework that permits the definition of effective practices in terms of project and product patterns, their instantiation in organization's projects and the gathering of data to determine the effective practices success and the identification of new ones.

The main objective of GPS project consists of the definition of a knowledge management environment to introduce process improvements into software intensive organizations by means of

[*] Email: pmf@inf.uc3m.es
[**] Email: amescua@inf.uc3m.es

efficient practices applied to organization's projects. Specific objectives defined in the project proposal are:

1. To work in the definition of a process modeling language that allows the representation at different abstraction levels of processes and methods adopted in a company in order to achieve good quality products.
2. From a process re-engineering point of view, a set of tools that allow to index, query, retrieval and reuse of processes considering their metadata and content will be provided.
3. From the Process Improvement point of view, a software tool to define and evaluate software processes will be developed.
4. From the point of view of Software Process Measurement, several algorithms defined in conjunction with the SEI will be developed.
5. From the point of view of repositories management, a software tool will be implemented in order to provide indexation, query and retrieval of multiformat documents using associated metadata and document content.
6. To work in interface design in order to obtain graphical based interfaces.

1.1 Project Resources

To achieve the goals stated above the project was assigned 16 EDP (14,5 from Universidad Carlos III de Madrid and 1,5 from Universidad Politécnica de Madrid). There is also a person with a FPI grant working in the project from September 2006.

Id	Nombre de tarea
1	**GPS: Plataforma de Gestión de Procesos Software: Modelado, Reutilización y Medición**
2	**Platform Definition**
3	Initial Analysis of processes requirements
4	Requirements Analysis of GPS Platform
5	Definition of the initial Architecture
6	Requirements Catalog
7	**Process Definition FOR Reuse**
8	Analysis of software process management
9	Study of interrelation between software process management and other disciplines
10	Definition of Product Patters as a way to define software processes FOR Reuse
11	Implementation of software tools to store and use Product Patterns
12	**Process Adaptation WITH Reuse**
13	Analysis of the state of the art and the practice in relation with process definition reuse
14	Definition of Project Pattern Concept, Metamodel and Provision of Predefined Patterns
15	Development of a software tool to clasify, index, search and reuse Project Patterns
16	Development of Project Patterns interfaces based on SPEM
17	**Development of Process Asset Libraries**
18	Design of database to support PAL and introduction of reference values
19	Design and Development of Requirements Management PAL
20	Design and Development of Project Management PAL
21	**Process Improvement Measurement**
22	Analysis of the state of the art and the practice related to Process Improvement Measurement
23	Definition of a technique and development a support tool to measure Process Improvement
24	Definition of a measurement package based on Product Pattern Improvement Approach
25	Definition of a measurement package based on Project Patterns Improvement Approach

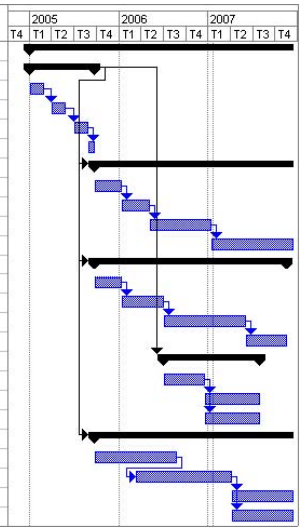


Figure 1: GPS project planning

1.2 Project Planning

Figure 1 shows the current GPS project planning that includes some changes with respect to the proposed one as it is explained in Section 2.

2 Level of achievement

In the project proposal the *Platform Definition* work package had as main objective to establish the functional and technical requirements of the GPS platform. Four tasks have been carried out: *Initial Analysis of processes requirements* in charge of identify the current research and industry about tools for Software Process Improvement, especially the CAKE (Computer Aided Knowledge Engineering) platform developed by several members of the project; *Requirement Analysis of GPS Platform* devoted to identify the ideal requirements; *Definition of the Initial Architecture* to propose components, functionality and interfaces of the platform and *Requirements Catalog* to establish functional and non functional requirements order by priority in different versions.

The result is a complete description of the ideal GPS platform that achieved the project objectives as well as it constituted a powerful and original tool for the users. In order to schedule the next activities of the project, that is, the design and development of the platform we faced to the following problems: the design of GPS platform would be strongly linked to the implementation environment, a great development effort that requires a specialized and exclusively dedicated team for the development task and the budget required to contract this team. After analyzing these problems and considering the characteristics of our project team we decide to rearrange the platform solution to a set of prototypes devoted to show the viability of solutions addressed in this project.

In this way, the new objectives are: to work on common definitions of the concepts to be used in the project (pattern, project pattern, product pattern, etc.); the development of a unified repository of processes; decomposition of project problems and challenges on the research groups; more capability of working in order to produce more research results. To facilitate the work in the project a web site (http://sem.sel.inf.uc3m.es:8400/default.aspx) was developed.

Four research lines grouped in two categories were defined: two reuse oriented lines (Line 1: Process Adaptation with Reuse and Line 2: Process Definition for Reuse) and two process Management oriented lines (Line 3: Development of Process Asset Libraries and Line 4: Process Improvement Measurement)

Research on *Process adaptation with reuse* defines and specifies the Reusable Project Pattern (RPP) concept. The mechanisms for cataloguing, indexing and searching for reusable components of RPP, are defined and implemented in a tool called PM-CAKE. Specific objectives are to define process patterns concept as a way to improve the software development process and project reuse, to determine a metamodel that specifies project patterns, to define in detail procedures to classify, index, search and reuse project patterns, to develop a tool able to apply project patterns issues to management software projects.

Research line 2, *Process Definition for Reuse*, focuses on the development of software projects, reusing experts' knowledge related mainly to project management processes. This knowledge will be encapsulated in the concepts "Product pattern" that will be defined as part of this research line results. The main goal is to provide software development enterprises with mechanisms to improve the deployment and the efficiency of use of software processes and software engineering best practices. Using product patterns as part of projects development, enterprises will improve their productivity as well as their capability to reuse knowledge. This research line is related to goals 1, 2 and 3 of the original proposal.

The third research work, *development of Process Asset Libraries*., is focused mainly on objectives 3, 4 and 5 of the proposal, but their results are oriented to firms that need to institutionalize their project and requirement management processes with a modern approach in order to implement their process improvement. A tool for support the process definition, management and improvement is given. The concept of P.A.L. (Process Assets Library) is defined and implemented. and a fully set of components is defined to be implemented in an organization. The main goal of

this research work is based on the definition and validation of a model for the requirement process area and project management (estimation, planning and monitoring) process area, by using a process asset repository.
The research line 4, named as *Process Improvement Measurement*, is mainly focused to the fourth objective of the proposal. This research line intends to define a framework to measure the quality, effectiveness and efficiency of the software processes and the projects launched to improve them. As part of this goal, an specific technique supported by a software tool, for software process improvement measurement is to be defined. This technique will be based on the measurement of the day-to-day activities performed by a software development organization.

2.4 Process Adaptation with Reuse

The final product will be a software tool based on CAKE technology; helping the user in reuse and development process. At the moment this tool is called SdpCAKE (Software development process, computer aided knowledge engineering). SdpCAKE will be a tool that offers solutions for helping the modelling, reusing and software development processes aiding Software Development Small Settings (SDSS). SdpCAKE offers software development patterns which the user can modify and search for finding the most suitable for a new project. The achieved objectives are:: a Reusable Project Metamodel developed in SPEM and a first version of the SdpCAKE tool. Remaining results are to develop new versions of the SdpCAKE tool for defining, reusing, searching and tracking project patterns and to prepare a set of RPPs for using with SdpCAKE.

2.1 Process Definition for reuse

Up to now results obtained are: management processes analysis from the point of view of software engineering discipline; study of the relationship among disciplines like process improvement, knowledge management and integrated development environments (IDEs); study of the pattern concept in the field of software engineering and the definition of the Product Pattern concept as the element to encapsulate the experts knowledge in process development.
The reminder of year 2007 will be dedicated to the implementation of the mechanisms to store and recovery product patterns. In order to achieve this goal different kind of IDEs will be analysed with the aim to select the one that can work appropriately with defined product patterns. The mechanisms to manage the storage and recovery of patterns will be defined and implemented. The extensions to the selected IDE will be implemented and finally the results obtained during the use of product patterns in real projects will be disseminated.

2.3 Development of Process Asset Libraries

The main goal of this research work is the development of a metamodel to evaluate, implement and upgrade the requirement process and project management process (Project Estimation, Planning and Monitoring) by using a process assets repository as support media. The accomplishment obtained to date is: a process definition of requirement, estimation, planning and monitoring processes by using as a main reference the Capability Maturity Model Integration (CMMI) process model; a prototype of defined processes and a definition of assets library.
The remaining work will be devoted to develop the requirement management prototype, to define a shared area of process assets library, to define a specific area for project management process (estimation, planning and monitoring) and, finally, to define a process dashboard using a measurement repository.

2.2 Process Improvement Measurement

The results obtained are: a consolidated version of the process improvement measurement and strategic management; a procedure and a questionnaires to determine the state of the practices in Spanish software companies related to process improvement measurement and strategic management; a technique, named "Sandwich Model" to measure the effectiveness, efficiency and quality of software processes and their improvement has been defined; and, finally, the first version of the software tool to support "Sandwich Model" has been developed. This software tool is based on SOA. Remaining results are to prepare two pre-defined packages to apply the "Sandwich Model" and second version of the software tool supporting "Sandwich Model".

3 GPS mid-term results

This section summarizes the main results of the GPS project.

3.1 Personnel in training

Five PhD students are preparing their theses in issues related to the project, section 3.2 describes these works. Also some undergraduate students are working in their master theses:

- F. Javier González Ortega: SPEM metamodel
- Mª José Jiménez Calvo-Manzano: Diseño de un sistema de gestión de proyectos software en equipo.
- Pedro de las Heras García : Análisis de Mejora de Equipos usando TSP.
- María Illera Bermejo: Análisis del estado de la investigación en mejora de proceso software.
- Luis Miguel Pozo: Análisis sobre herramientas software que permitan elegir proyectos o metodologías en función de las características de los proyectos a desarrollar.
- Carlos Benito Rodríguez: Creación de patrones de producto para la gestión de proyectos mediante metodologías ágiles
- Mª Ángeles González Velázquez Creación de patrones de producto para la gestión de proyectos siguiendo la metodología Craig Largman.
- Alberto Collantes Desarrollo de un repositorio en la plataforma SQL-Server 2005 que modele la información de los patrones de producto
- Luis Miguel Pozo: Desarrollo de herramienta software que permita elegir metodologías de procesos.

3.2 Publications

Table 2 summarizes the main publications derived from this project up to now.

	Journals	**Book Chapters**	**Conferences**	**Thesis**
2005			[17]	1
2006	[5,13,14]	[24]	[2,3,4,6,8,10,11,18,19]	5
2007	[9,12,23]	[20,21]	[1,7,15,16,22,25]	

Table 2: GPS Mid-term results

The members of the group are also supervising six theses (one is finished and five under development):

- HarithAl -Jumaily: Aplicación de Técnicas Activas par el Control de Restricciones en el desarrollo de Bases de Datos (september 2006). Supervisors: Paloma Martínez and Dolores Cuadra
- Fuensanta Medina: Aproximación basada en patrones de producto para facilitar la implantación de la mejora de procesos utilizando como soporte sistemas colaborativos. Supervisor: Maribel Sanchez Segura
- Diego Martín: Reutilización de patrones de SPD. Supervisor: Juan Llorens.
- Hugo Mitre: Medición para la gestión estratégica y operativa de la mejora de procesos en organizaciones intensivas en software. Supervisor: Antonio Amescua
- Ivan Antonio García Pacheco: Metamodelo para la definición, implantación y mejora de los pocesos de gestión de proyectos. Supervisors: José A. Calvo-Manzano and Tomás San Feliu.
- Edgar Ariel Serrano Rico: Metamodelo para la definición, implantación y mejora de los procesos de gestión de requisitos. Supervisors: Gonzalo Cuevas y José A. Calvo-Manzano.

3.3 Technology Transfer

The technology transfer activities performed in the scope of GPS project has been performed by means of the collaboration with several companies.

PROGRESION SMP (http://www.progresion.net) is a spin-off company owned by Carlos III University researchers that provides consulting services to software intensive organizations to improve their development and service provision processes. PROGRESION is providing a set of data to be used in the analysis of the practical viability of the product and project patterns concepts. This collaboration has produces several research papers, being [13] the most valuable contribution.

The Reuse Company (http://www.reusecompany.com) is a spin-off company owned by researchers coming from several universities, including Carlos III University. The collaboration activities are based in the Project Patterns Management Tool that is integrated in the Knowledge Engineering Framework created by The Reuse Company. This software tool is providing the capabilities mentioned in Section 2.4.

3.4 Cooperation with other research groups

Some GPS researchers are participating actively in other research groups, disseminating the GPS research results. The conclusions and contributions from these other research groups have been used to reorganize and improve the first versions of project's results.

- Latin-American Cathedra on Software Process Improvement (MPSEI) which is composed of researchers from several Spanish universities and companies. The collaboration activities between GPS and MPSEI researchers have been centred in the interchange of research results related to product, projects and process patterns concepts and the fundamentals of Process Asset Libraries (PAL), [3].

- RePRIS – Spanish Software Testing Network - The collaboration activities between GPS and MPSEI researchers are related to the publication of the GPS research results related to verification, validation and testing processes using the concept of product patterns, [12]

Finally, several GPS researchers have been collaborated with researchers from Software Engineering Institute (SEI) at Carnegie Mellon University, Pittsburgh (EEUU), with similar research lines to the project objectives, concretely, with those related to software processes definition.

4 References

[1] Al-Jumaily Harith, Cuadra Dolores, Martínez Paloma. MDD Approach for Maintaining Integrity Constraints in Databases. CAISE 2007, 19th Conference on Advanced Information Systems Engineering (submitted).

[2] Al-Jumaily Harith, De Pablo César, Cuadra Dolores, Martinez Paloma, Using UML's Sequence Diagrams as Termination Analyzer for Triggers-Based Execution, BNCOD 2006, 23rd British National Conference on Databases, July 2006, Northern Ireland.

[3] Amescua Antonio, Cuevas Gonzalo, García Javier, Llorens Juan, Martinez Paloma, and Martín Diego. A Pattern-Based Approach to Deploy Process Improvements in Small Settings. Proceedings of the First International Research Workshop for Process Improvement in Small Settings, 2005, January 2006.

[4] Amescua Antonio, García Javier, Sánchez-Segura Maria-Isabel, Medina Fuensanta, Software process improvement for practitioners based on knowledge management tools. 5th WSEAS International Conference on Software Engineering, Parallel And Distributed Systems, February 15, 2006 Alcala de Henares

[5] Amescua Antonio, García Javier, Sánchez-Segura Maria-Isabel, Medina Fuensanta. Approaching Software Process Improvement to Organizations. WSEAS Transactions on computers, Issue 3, Volume 5, March 2006, pp 507-514.

[6] Amescua Antonio, García Javier, Sánchez-Segura Maria-Isabel, Medina-Domínguez Fuensanta. A Pattern-based solution to bridge the gap between theory and practice in using process models. PROSIM 2006, Software Process Simulation and Modelling Conference, 20-21 May 2006, Shanghai, China.

[7] Bermón Leonardo, Amescua Antonio, García Guzmán Javier and Martínez Paloma. Process Asset Library in Software Process Support Technology: A Review of the literature. 2007 IRMA International Conference. Mayo 2007, Vancouver, Canadá.

[8] Camps Rafael y Cuadra Dolores. Transforming Ternary Associations to Database Schemas, 9º Workshop Iberoamericano de Ingeniería de Requisitos y Ambientes de Software (IDEAS) 2006, La Plata, Buenos Aires, Argentina, 24 al 28 de abril de 2006

[9] Cuadra Dolores, Gascueña Concepción M., Martínez Paloma. A multidimensional approach to the representation of the spatio-temporal multi-granularity. Data and Knowledge Engineering Journal (conditional acceptance).

[10] De Pablo-Sánchez César, González-Ledesma Ana, Moreno Antonio, Martínez-Fernández José Luis, Martínez Paloma. MIRACLE at the Spanish CLEF@QA 2006 Track, Working Notes for the CLEF 2006 Workshop, 20-22 September, Alicante, Spain.

[11] De Pablo-Sánchez César, Martínez-Fernández José Luis and Martínez Paloma. Named Entity Processing for Cross-lingual and Multilingual IR Applications, Workshop on New Directions in Multilingual Information Access, 29th International Conference on Research and Development in Information Retrieval, 2006, Seattle, USA.

[12] García Guzmán Javier, Amescua Seco Antonio, Velasco de Diego Manuel. TOP 10 de factores que obstaculizan la mejora de los procesos de verificación y validación en

organizaciones intensivas en software. Revista española de innovación, calidad e ingeniería del software (REICIS). To appear in 2007.
[13] García Guzmán Javier, López-Cortijo García Román, Amescua Seco Antonio, Cuevas Agustín Gonzalo. CASE STUDY: a practical approach for SPI in large Spanish companies. Process: Improvement and Practice, Volume 11, Issue 3. May/June 2006.
[14] García-Serrano Ana M., Martínez-Fernández José L., Martinez Paloma, Experiences in evaluating multilingual and text-image information retrieval, International Journal of Intelligent Systems, Volume 21, Issue 7, pages 655-677, July 2006.
[15] Gascueña Concepción M., Cuadra Dolores, Martínez Paloma. A formal proposal for incorporating spatio-temporal multigranularity in a conceptual multidimensional model. PODS 2007, Symposium on Principles of Databases Systems (submitted).
[16] Martín Diego, García Javier, Amescua Antonio, Llorens Juan. Reusable Project Patterns to enhance SPI. The Software Engineering Process Group Conference (SPEG) 2007 (submitted).
[17] Martínez Paloma, Amescua Antonio, García Javier, Cuadra Dolores, Llorens Juan, Fuentes J.M., Martín Diego, Cuevas Gonzalo, Calvo-Manzano José Antonio, San Feliu Tomás, Requirements for a knowledge management framework to be used in software intensive organizations. 2005 IEEE International Conference on Information Reuse and Integration, Las Vegas, Nevada, USA, 15 al 17 de Agosto de 2005.
[18] Martínez-Fernández José Luis, Villena Julio, García-Serrano Ana, Martínez Paloma. MIRACLE Team Report for ImageCLEF IR in CLEF 2006. Working Notes for the CLEF 2006 Workshop, 20-22 September, Alicante, Spain
[19] Medina Fuensanta, Amescua Antonio, Sánchez-Segura Maria-Isabel, García Javier. An Experience Providing University Students with Skills to Work with Collaborative Tools. IEEE International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering. 4-14 Diciembre 2006. Bridgeport, USA.
[20] Medina Fuensanta, Sánchez-Segura Maria Isabel, Amescua Antonio, García Javier. Teaching Collaborative Web Portals Technology at University. Encyclopedia of Portal Technologies and applications. Febrero 2007 Idea Group Publishers.
[21] Medina Fuensanta, Sánchez-Segura Maria-Isabel, Amescua Antonio, García Javier. Patterns in the field of software engineering. Encyclopedia of Information Science and Technology, Febrero 2008, Idea Group Publishers
[22] Medina-Domínguez Fuensanta, Sánchez-Segura Maria-Isabel, Amescua Antonio, García Javier. Extending Microsoft Team Foundation Server Architecture to support Collaborative Product Patterns. Internacional Conference on Software Process. Minneapolis. 19-20 Mayo 2007.
[23] Sánchez-Segura Maria-Isabel, García Javier, Amescua Antonio, Medina Fuensanta A Study on How Software Engineering Supports Projects Management. Software Quality Journal (conditional acceptance).
[24] Sánchez-Segura María-Isabel, New trends in Software Process Modelling. Febrero de 2006. World Scientific. ISBN. 981-256-619-8.
[25] Santos Ignacio, Cuadra Dolores, Martínez Paloma. On the semi-automatic validation and decomposition of ternary relationships with optional elements. ICEIS 2007, (submitted).

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Técnicas para el Robustecimiento de Procesos en Visión Artificial para la Interacción (*Strengthening of Computer Vision Processes for Interaction*) TIN2004-07087

José Javier Lorenzo Navarro *
Universidad de Las Palmas de Gran Canaria
Inst. Univ. Sistemas Inteligentes y Aplic. Numéricas en Ingeniería
Edif. Central Parque Tecnológico. 35017 Las Palmas

Abstract

The proposal of this project falls into the field of artificial vision for interaction, and particularly of Perceptual User Interfaces (PUIs). The objective is to define, build and evaluate a framework for devising and developing solutions for Natural Communication Interfaces, especially for interaction with people in environments with changing conditions. The design of the system is based on Active Vision techniques combined with Homeostatic Adaptation, which facilitate communication in changing illumination conditions, aiding in the exchange of visual information. Also, it can study and model processes of interaction perception-action through gesture recognition. The general problems that must be solved include detection of presence, tracking and identification of people, tracking and analysis of facial, hand or body gestures.
Keywords: Computer Vision, Human-computer interaction, Homeostatic adaptation.

1 Goal of the project

The objective of the project was to define, build and evaluate a framework for devising and developing solutions for Natural Communication Interfaces, especially for interaction with people in environments with changing conditions. The design of the system will be based on Active Vision techniques combined with Homeostatic Adaptation, which facilitate communication in changing illumination conditions, aiding in the exchange of visual information. It was also proposed to study and model processes of perception-action through gesture recognition.

The previous general goal was divided into more specific ones which are the following:

*Email: jlorenzo@iusiani.ulpgc.es

1. To investigate the use of homeostatic models with the aim of stabilizing the parameters that define the quality of the images acquired by cameras. Those parameters must be kept into the homeostatic regime which is the one where the system exhibits its maximum performance.
2. To identify the essential variables and to study the homeostatic ranges where the system keeps an acceptable level of performance and the possible interrelation among those essential variables.
3. To determine and include methods to measure the quality of the images in the homeostatic regulation loop, along with the study of the feedback in the vision system and the global system feedback.
4. To keep on researching on methods and techniques to develop robust computer vision modules and their application in human computer interaction. To do that, it is necessary to integrate some modules previously developed in the research group like a real time precategorical template-based traking technique and another one based on contours, the ENCARA module for face detection and hand gesture detection and tracking.
5. To study new methodological considerations, including an analysis of which models and methods are more suitable, in the development of practical applications. Also we were interested in modelling how the quality gets worse, and in the estimation of the parameters.
6. To develop prototypes in Matlab and OpenCV that allow us to test and evaluate them within the research group, and to share them with others research groups.
7. To develop a final practical application with the aim of integrating some of the methods studied in the project. This prototype will be evaluate by users to test the proposal in a real human-computer interaction application.

2 Level of sucess achieved in the project

The first two years of the project have been devoted to study different methods and techniques in order to make more robust existing methods or to develop new ones. The members of the research group has worked in different but complementary areas with the aim of exploring several approaches. Last year will be devoted to the development of a demonstrator which integrates the proposed methods with a funcionality and to test them in a real application of human-computer interaction (HCI). Below, the results obtained in the investigated areas are presented.

2.1 Homeostatic control

Among the previous detailed objetives, some of them make reference to the study of the homeostatic control for stabilizing the parameters which influence the image quality. We have proposed and implemented a initial model based on the definition of a set of homeostatic variables like brightness, contrast and white balance. To test the proposal we introduced the homeostatic control in a typical human-computer interaction application as face detection. In this application the introduction of homeostatic control allows to keep the performance of the face detector at the same level even when illumination conditions changed. This application allowed us to test not only the low level homeostatic control but also its integration with

the application. It was possible because another essential variable, the size of the face, was introduced and the face detector was able to maintain its detection rate even if the subject moved futher away [8].

In objetive 3 we planned to test the homeostatic control feedback in a complete system. A robotic application gives us the opportunity to test it. A simple line follower tracker robot and a circuit which included a dark tunnel was implemented. The feedback of the homeostatic regulation with the whole system was carried out with the introduction of a essential variable that affects the transactional speed of the robot influenced by the brightness of the image. [7, 32]

Also related to the homeostatic control is the study of different essential variables. One of those variable is the contrast of the image that can be corrected when it is not in the homeostatic region using autofocus techniques. Autofocus techniques are based on contrast measures but most of them are not oriented to human computer interaction so a comparison of some of them was made but in a HCI environment. Interestingly, more recent measures based on wavelets do not yield better results, as it is shown in [27].

2.2 Gesture detection and recognition

Another of the objetives proposed in the project is the development of robust computer vision modules and their application to human computer interaction. In this project some techniques for hand gesture tracking and recognition have been developed. In this respect a method based on the modification of the object of interest was developed, so the number of detected gestures is increased. Instead of detecting different individual hand gestures, firstly the wrist is detected and then the gesture is detected based on a cascade classifier. This approach introduces two advantages: an efficient segmentation that is independent of the gesture and a faster training phase [9]. To improve the approach, temporal coherence was introduced getting in this manner three adittional contributions: an initialization process where the user does not give an initial segmentation of the hand, a higher number of detected gestures and a reduction in the processing time giving a real time module to track and identify several gestures [9]. Finally to strengthen the method, a model of skin color is introduced by means of a face detection and then comparing the color with a measure based on the Hausdorff distance. This modification allows to identify different poses of the same gesture [28].

2.3 Object tracking

Several tasks in computer vision are based on the necessity of tracking objects in real time. As part of the goal of studying the use of tracking methods in HCI, it has been proposed a technique based on an heuristic algorithm that detects the movement of several templates in dynamics sequences. The heuristic algorihm reduces the transformation space to explore similar to a search problem which means a lower processing time [11, 12]. In order to update the template after a time of tracking the object, it was introduced a strategy based on the quality of the matching measure along with a visual memory that stores different views of the same object [13]. This approach was tested in a surveillance task with acceptable results [14].

At it was pointed out, in the template based tracking approaches a problem that arises is when to update the template of the tracked object because it changes as the time passes. To minimize this effect an incremental updating strategy is proposed that is based on the context

using a second order isomorfic representation space. This concept emphasizes the influence of the environment into the location of objects in the scene making possible to introduce it in traditional tracking methods in order to automate different aspects.[31].

Edges detected in video sequences using classic Canny operator are usually unstable, prone to appear or dissappear in consecutive frames if Canny's parameters are not properly adjusted. This blinking effect may affect successive stages in the vision process, so it is neccesary to soothe its effect. We propose a modification of classic Canny's algorithm where the last stage of the algorithm, hysteresys, is replaced with a dynamic threshold, thus reducing the amount of hand-tuned parameters. Also the gaussian diffussion is replaced by an anisotropic diffusion in order to avoid the blurring effect and thus increasing edge stability in video sequences. [25]

Another technique that has been proposed in the project is an object tracking method based on the Predator Swarm metaphor to be used in video sequences. A group of particles is designed to fly in a Boid-like fashing (i.e. mimiquing social insects' moving behaviour) while looking for preys that satisfy their taste. Using combined cues from local patches of a given frame, the swarm is able to recognice color, gradient, movement and salient points in order to track its prey, which is translated into an effective real-time and robust object tracking activity. [24]

2.4 Face and people detection

An important competence in any HCI application is face detection. ENCARA2 is an improvement and extension of the previous face detector developed in the reserarch group, ENCARA, an one objetive of this project has been to improve it. One the one hand, making use of the explicit knowledge of the color, movement or face geometry we get a reduction in the processing time in constrained environments. On the other hand, making use of the implicit knowledge obtained from a supervised learning phase we get a more robust method although slower [15, 16, 17, 3].

The use of color information to guide the face detection process produces a real-time face detector but not very robust when ilumination changes. The introduction of implicit detectors based on the the Viola-Jones detector when there is not a recent face detection allow ENCARA2 to be more robust. While ENCARA2 detects faces continuously, it makes use of facial features like color, eyes, face geometry or the last face detected, increasing its rate detection. The combination of the two approaches in an opportunistic way yields a face detector that when it is tested in video sequences, gets better results than the Viola-Jones and Rowley face detectors, even giving information about the eyes position in most of the images.

People detection has been studied in the project in the context of low resolution images. This context is very typical in surveillance imaages where people is represented by a low number of pixels. With the combination of some classifiers based on the Viola-Jones detector and some geometrical constrains corresponding to the human body, a method was developed with better performance than the people detector included in the OpenCV library [29].

2.5 Face identification

As a video sequence is processed by ENCARA2, the visual information provided is highly redundant because many of the detected faces correspond to the same person in a similar pose. To extract automatically exemplars or significant images from the video sequence, a method

has been proposed that detects sharp changes in the aspect which are associated to a lost in the tracking of the eyes. Those exemplars are used to train a gender classifier that starts from scratch (*tabula rasa*) and then progressively learns in a supervised way. The performance obtained with this approach for the gender classification problem is similar to the one obtained with a classifier trained on-line with positive and negative examples [18, 20, 21]. The same approach has been applied to identity recognition with promising results [21, 2] The accuracy obtained with this approach is similar to the one obtained with a classifier trained off-line with positive and negative examples [18, 2].

In face identification one of the most used approach is the one based on principal component analysis (PCA) in order to reduce the representation space. In this project, a study about the number of components necessary to get a classification of human faces [19]. Different descriptors of the individual are considered as gender, glasses, moustache or race. The results for each of the previous descriptors allow us to determine the best configuration of components in a PCA approach. To test the previous approach, a comparison with a method based on Gabor filters obtained iteratively for each descriptor was carried out. Although in some cases the time is greater in our approach, better results are obtained [30].

2.6 Final practical application and methodological considerations

The face robot CASIMIRO developed within the research group also implements a number of techniques that allow more robust vision-based interaction [4, 5]. One of the techniques, for example, deals with face detection. Facial detectors are commonly based on skin colour information. However, they tend to consider wooden furtniture as skin colour patches. To alleviate this, stereo information (obtained using a pair of cameras under the nose of the robot), is used to filter out those zones of the image that are at greater depth, generally furniture and doors. So CASIMIRO has been considered as a first demonstrator for implementing the techniques developed in this project. However, being also part of another research project, the robot deals from its inception with more cognitive issues, and it is at present fully implemented.

Now we are involved in the design of a more practical and feasible application specially for this project. This practical application is intended to integrate some method studies during the execution of the project. The implementation of this application will be during the last year of the project. The demonstrator consists of the development of a kiosk to be used in an optician to allow the client select different types of glasses and test how they look like. The user can interact with the kiosk with gestures to select the model or to adjust the virtual glasses once they appear in the screen on his or her face. A desirable feature of this application is the possibility of taking off the client glasses in the screen and to superimpose the virtual ones. In this way the user can appreciate the look. The development of this demonstrator will allow to study the better methodological considerations in this kind of applications.

3 Result indicators

During the development of this project one member of the group has obtained the PhD degree, Oscar Déniz Suárez [5] and the PhD student Luís Antón Canalís associated to the project obtained the Advanced Studies Degree (DEA) with honors. The scientific production of the

group has been the following one: 1 patent, 3 articles in international journals and 27 articles in national and international conference proceedings (one of them with best application award).

The research group has organized last year the VII Workshop de Agentes Físicos, WAF'2006, April 27-28, Las Palmas de Gran Canaria [6].

In relation to the training activities of the research group, it has renewed in 2006 the *Mención de calidad* of the Spanish Ministry of Education for the PhD program "Sistemas Inteligentes y Aplicaciones Numéricas en Ingeniería". This Phd program has been approved by the University of Las Palmas de Gran Canaria as one of the seven official postgraduate programs of the university.

We have continued with our research collaborations which have resulted in joint publications:

- Dr. M. Castrillón was from september to december of 2006 in the departament of Applied Computer Science of Bielefeld University working with professor Gerhard Sagerer in facial description with incremental PCA and hand detection for interaction.
- Dr. O. Déniz collaborates in the PhD program "Mecatrónica" (*Mención de Calidad* of the Spanish Ministry of Education) of the University of Castilla La Mancha teaching the course "Herramientas de Visión por Computador en Inteligencia Artificial".
- L. Antón is visiting in 2007 the Computer Vision Laboratory of ETH Zurich the next four mounths and will work with Dr. Bastian Liebe in codebooks and object recognition in complex scenes.

References

[1] Hernández-Tejera, F. et al., *Sistema de información interativo inteligente (Kiosko Inteligente).* Patente nº 200201371. Fecha publicación junio, 2006.

[2] M. Castrillón-Santana, O. Déniz Suárez, J. Lorenzo-Navarro, and M. Hernández-Tejera, *Learning to recognize faces by successive meetings.* Journal of Multimedia, ISSN 1796-2048, vol 1, Issue 7, pp 1-8, 2006

[3] Castrillón Santana, M., Déniz Suárez, O., Hernández Tejera, M. y Guerra Artal, C., *Encara2: Real-time detection of multiple faces at different resolutions in video streams* Journal of Visual Communication and Image Representation, 2007 *(in press)*

[4] O. Déniz, M. Hernández, J. Lorenzo, M. Castrillón, *An engineering approach to sociable robots.* Journal of Experimental and Theoretical Artificial Intelligence. *(accepted)*

[5] Óscar Déniz Suárez, *An Engineering Approach to Sociable Robotics.* PhD thesis. Inst. de Sistemas Inteligentes y Aplic. Numéricas en Ingeniería, Univ. de Las Palmas de Gran Canaria, julio, 2006.

[6] Jorge Cabrera-Gámez, Vicente Matellán Olivera (Eds.), *Actas del VII Workshop de Agentes Físicos WAF'2006*, Las Palmas de Gran Canaria, 27-28 de Abril de 2006. Universidad de Las Palmas de Gran Canaria 2006, ISBN 84-689-8115-X.

[7] J. Daniel Hernández Sosa, J. Lorenzo, Antonio C. Domínguez Brito and M. Castrillón-Santana, *A Homeostatic-Adaptive Approach for Controlling Robotic Systems.* Eurocast 2005, Tenth International Conference on Computer Aided Systems Theory, February 7-11, Las Palmas de Gran Canaria, Spain.

[8] Javier Lorenzo-Navarro, Daniel Hernández, Cayetano Guerra and José Isern-González, *A Proposal For A Homeostasis Based Adaptive Vision System*. Lecture Notes in Computer Science, Pattern Recognition and Image Analysis: Second Iberian Conference, IbPRIA 2005, Estoril, Portugal, June 7-9, 2005, vol 3522/2005, 2005.

[9] L. Antón-Canalis, E. Sánchez-Nielsen and M. Castrillón-Santana, *Hand Pose Detection for Vision-based Gesture Interfaces*. Proceedings IAPR MVA 2005, Tsukuba Science City, May 2005, Japan.

[10] Luis Antón-Canalís, Elena Sánchez-Nielsen, C. Guerra-Artal and M. Castrillón-Santana, *Fast and Accurate Hand Pose Detection for Human-Robot Interaction"*. 2nd Iberian Conference on Pattern Recognition and Image Analysis, Ibpria 2005, Estoril, Portugal.

[11] Elena Sánchez-Nielsen, Mario Hernández-Tejera, *Heuristic Algorithm for computing fast template motion problem in video streams.*7th Int. Conf. on Advanced concepts for intelligent vision systems (ACIVS 2005), Antwerp, Belgium, Lecture Notes in Computer Science, Vol: 3704/2005, pp. 507-516, September 2005.

[12] Elena Sánchez Nielsen, Mario Hernández Tejera, *A heuristic search based approach for moving objects tracking*. Procs of 19 Int. Joint Conference on Artificial Intelligence (IJCAI'05), 2005, pp. 1736-1737.

[13] Elena Sánchez-Nielsen, Mario Hernández-Tejera, *Heuristic Algorithms for Fast and Accurate Tracking of Moving Objects in Unrestricted Environments*. 1st International Symposium on Brain, Vision and Artificial intelligence BV&AI 2005, 19-21 October, Naples, Italy. Lecture Notes in Computer Science, vol. 3708/2005, pp. 547-554, Oct 2005.

[14] Elena Sánchez Nielsen, Mario Hernández Tejera, *A fast and accurate tracking approach for automated visual surveillance*. Procs. of IEEE Int. Carnahan Conference on Security Technology, 2005, pp. 113 -116.

[15] M. Castrillón Santana, Oscar Déniz Suárez, Cayetano Guerra Artal y Josep Isern González, *Cue Combination for Robust Real-Time Multiple Face Detection at Different Resolutions*. Eurocast 2005, Tenth International Conference on Computer Aided Systems Theory, February 7-11, Las Palmas de Gran Canaria, Spain. Extended version in LNCS vol. 3643/2005 pp. 398-403

[16] M. Castrillón-Santana, C. Guerra-Artal and M.Hernández-Tejera, *Real-time Detection of Faces in Video*. Second Workshop on Face Processing in Video (FPiV'05) in Proceedings of Second Canadian Conference on Computer and Robot Vision (CRV'05), 9-11 May 2005, Victoria, BC, Canada. ISBN 0-7695-2319-6. pp. 298-305. Victoria, Canada.

[17] Modesto Castrillón-Santana, Javier Lorenzo-Navarro, Oscar Déniz-Suárez, José Isern-González and Antonio Falcón-Martel, *Multiple Face Detection at Different Resolutions for Perceptual User Interfaces*. Lecture Notes in Computer Science, Pattern Recognition and Image Analysis: Second Iberian Conference, IbPRIA 2005, Estoril, Portugal, June 7-9, 2005, Proceedings, vol. 3522/2005, 2005.

[18] M.Castrillón-Santana, O. Déniz-Suárez, J. Lorenzo-Navarro and M.Hernández-Tejera, *Face Exemplars Selection from Video Streams for Online Learning*. Second Workshop on Face Processing in Video (FPiV'05) in Proceedings of Second Canadian Conference on Computer and Robot Vision (CRV'05), 9-11 May 2005, Victoria, BC, Canada. ISBN 0-7695-2319-6. pp. 314-321

[19] Modesto Castrillón-Santana, Javier Lorenzo-Navarro, Daniel Hernández-Sosa and Yeray Rodríguez-Domínguez, *An Analysis of Facial Description in Static Images and Video Streams*. Lecture Notes in Computer Science, Pattern Recognition and Image Analysis: Second Iberian Conference, IbPRIA 2005, Estoril, Portugal, June 7-9, 2005, vol. 3522/2005, 2005.

[20] M. Castrillón-Santana, O. Déniz Suárez, J. Lorenzo-Navarro, and M. Hernández-Tejera, *Gender and Identity Classification for a Naive and Evolving System*. Second Workshop on Multimodal User Authentication MMUA 2006, Toulouse, France.

[21] M. Castrillón Santana, O. Déniz-Suárez, J. Lorenzo-Navarro and M. Hernández-Tejera, *Face recognition from a tabula rasa perspective*. 7th International Conference on Automatic Face and Gesture Recognition, FG2006, April 10-12 2006, Southampton, UK.

[22] L. Anton-Canalís, E. Sanchez-Nielsen, *Hand Posture Dataset Creation For Gesture Recognition*. VISAPP'06, Setúbal, Portugal, Feb. 25-28 2006. Vol.2. pp 197-200. ISBN: 972-8865-40-6.Setúbal, Portugal.

[23] L. Anton-Canalís, M. Hernández-Tejera and E. Sanchez-Nielsen, *Swarmtrack: A Particle Swarm Approach to Visual Tracking*. VISAPP'06, Setúbal, Portugal, Feb. 25-28 2006. Vol.2. pp 221-228. ISBN: 972-8865-40-6.Setúbal, Portugal.

[24] Luis Antón-Canalís, Mario Hernández-Tejera,Elena Sánchez-Nielsen, *Particle Swarms as Video Sequence Inhabitants For Object Tracking in Computer Vision*. Sixth Internacional Conference on Intelligent System Design and Applications, Jinan, Shandong, China, 2006.

[25] Luis Antón-Canalís, Mario Hernández-Tejera, Elena Sánchez-Nielsen, *AddCanny:: Edge Detector For Video Processing*. Advanced Concepts for Intelligent Vision Systems (Acivs-06), Antwerp, Belgium, 2006.

[26] Luis Antón-Canalís, Mario Hernández-Tejera and Elena Sánchez-Nielsen, *Analysis of Relevant Maxima in Distance Transform. An Application To Fast Coarse Image Segmentation*. 3nd Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA 2007), Gerona, 2007. *(accepted)*

[27] J. Lorenzo, O. Déniz, M. Castrillón and C. Guerra, *Comparison of focus measures in a face detection environment*. 4rd International Conference on Informatics in Control, Automation and Robotics (ICINCO'07), Angers, France, 2007. *(submitted)*

[28] E. Sánchez-Nielsen, L. Antón-Canalis, C. Guerra-Artal, *An autonomous and User-Independent Hand Posture Recognition System for Vision-based Interface Tasks*. CAEPIA'05 - Conferencia de la Asociación Española para la Inteligencia Artificial , November 16-18 2005, Santiago de Compostela, Spain.

[29] M. Castrillón-Santana, Q. Vuong, *Combining Human Perception and Geometric Restrictions for Automatic Pedestrian Detection*. CAEPIA'05 - Conferencia de la Asociación Española para la Inteligencia Artificial , November 16-18 2005, Santiago de Compostela, Spain. *CAEPIA'05´s Best Application Award)*

[30] M. Castrillón-Santana, J. Lorenzo-Navarro, D. Hernández-Sosa, J. Isern-González, *Face Description for Perceptual User Interfaces*. CAEPIA'05 - Conferencia de la Asociación Española para la Inteligencia Artificial, November 16-18 2005, Santiago de Compostela, Spain.

[31] Cayetano Guerra, José Isern, Antonio Carlos Domínguez, and Daniel Hernández, *Template Updating Based on Context*. I Workshop sobre Reconocimiento de Formas y Análisis de Imágenes, I Congreso Español de Informática (CEDI'2005), September 13-16, Granada, Spain.

[32] Daniel Hernández, Javier Lorenzo, Antonio C. Domínguez, Josep Isern, *A Proposal of a Homeostatic-Adaptive Control for a Robotic System*. VII Workshop de Agentes Físicos, WAF'2006, April 27-28, Las Palmas de Gran Canaria, Spain.

[33] Oscar Déniz, Luis Antón, Modesto Castrillón, Mario Hernández, *Expressive Robotic Face for Interaction*. VII Workshop de Agentes Físicos, WAF'2006, April 27-28, Las Palmas de Gran Canaria, Spain.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Intelligent Systems to Assist in Cytological Diagnosis and to Train Cytotechnics TIN2004-07236

Jose Manuel Zurita López*
Main Researcher
Dept. Computer Science and Artificial Intelligence
University of Granada

Abstract

This project aims at the application of Soft Computing techniques to the development of intelligent systems which assist in the cytological diagnosis and the training of cytotechnics. As an important previous task, suitable knowledge representation must be defined for the Cytology area. Then knowledge extraction methods must be defined for both acquiring knowledge from the experts and deriving knowledge from a set of clinical data. To achieve all of this different components from Soft Computing will be used (Artificial neural networks, Fuzzy Systems, Evolutionary systems). Later on, two intelligent systems to aid in cytological diagnosis will be developed: one for the thyroid and another for cervix-vaginal cytology. These systems will not only allow for the management and exploitation of knowledge, but also will be able to learn and hence increase their knowledge base. They will also support the diagnoses with bibliographical references relative to them and retrieved through bibliography database queries. As a specific case of exploitation of these systems, an intelligent tutoring system to assist in the training of future cytotechnics will be developed.

Keywords: Intelligent systems, cytological diagnosis, Soft Computing, knowledge acquisition.

1 Objetivos del proyecto

El diagnóstico citológico es un problema de enorme interés en la detección de múltiples patologías (como por ejemplo, cáncer). Es una tarea básica que requiere de un conjunto amplio de conocimientos y no poca experiencia para desarrollarla de forma efectiva. La mayoría de las observaciones y decisiones diagnósticas en citología están mediatizadas por vaguedades e incertidumbres. La interpretación de los rasgos morfológicos observados con ayuda del microscopio es central en la construcción de un diagnóstico. A su vez, la dificultad en el diagnóstico varía según el tipo de tejidos que se analizan, desde algunas fáciles a otras en las que no se dispone de conocimiento completo. Además, la práctica diagnóstica diaria enfrenta al citólogo a casos en los que necesita consultar y ha de acudir a fuentes bibliográficas.

* Email: zurita@decsai.ugr.es

Este campo representa una área clara para el diseño de sistemas inteligentes que incluyan como parte de su funcionalidad el manejo del conocimiento (como en los sistemas expertos) pero que se complemente con otras técnicas que permitan extraer conocimiento de forma automática, que amplíe el conocimiento mediante técnicas de Aprendizaje Automático o que proporcionen soporte mediante recuperación de información en bases de datos bibliográficas.
Los conocimientos necesarios para formar un buen citólogo-citotécnico son amplios. Formar un buen patólogo requiere un esfuerzo importante. Se puede contribuir notablemente a su formación mediante un sistema inteligente que ilustre la resolución de casos reales, explique los diagnósticos obtenidos y ofrezca referencias para justificar y ampliar la respuesta. Un sistema de este tipo entra dentro de la categoría de los "Intelligent Tutoring Systems", que constituyen un área activa de la investigación en aras de facilitar y mejorar la formación y el aprendizaje en seres humanos.

Teniendo en cuenta las hipótesis de partida anteriores, los objetivos generales que nos proponemos en el proyecto son:

a) Estudiar la problemática de la extracción del conocimiento para el diagnóstico citológico y proponer soluciones adaptadas al caso, tanto para el conocimiento adquirido de expertos como para la extracción automática a partir de datos de historiales clínicos (casos de la práctica diaria).
b) Diseño de un sistema basado en técnicas de Computación Flexible para ayudar a los profesionales en el diagnóstico citológico. El sistema permitirá la incorporación/actualización fácil del conocimiento y será accesible a través del Web, facilitando su uso y la colaboración de expertos de todo el mundo. El sistema incorporará motores de búsqueda bibliográfica que se activen automáticamente a partir de los diagnósticos. Se tratará de un sistema general adaptable a diversas patologías. En particular se implementarán dos: uno sobre citología de tiroides y otro sobre citología cérvico-vaginal.
c) Construir un sistema tutor asociado al sistema inteligente que facilite la formación de citotécnicos. A partir del conocimiento explotado a través del sistema inteligente y de casos de prueba diseñados adecuadamente, se proporcionará un sistema de consulta y entrenamiento para los futuros citotécnicos.

Estos objetivos se conseguirán con el siguiente plan de trabajo que se divide en siete tareas y cuyo cronograma queda detallado a continuación.

Tarea 1. Análisis del problema de diagnóstico citológico.
Tarea 2. Diseño del modelo de sistema inteligente.
Tarea 3. Adquisición del conocimiento.
Tarea 4. Desarrollo de los sistemas inteligentes.
Tarea 5. Análisis del problema de la formación de citotécnicos.
Tarea 6. Diseño del sistema tutor.
Tarea 7. Desarrollo del sistema tutor.

Actividades/Tareas	Primer año (*)	Segundo año (*)	Tercer año (*)
Tarea 1.1 Análisis general del problema			
Tarea 1.2 Análisis de requisitos del sistema inteligente			
Tarea 2.1 Diseño arquitectural			
Tarea 2.2 Módulo de representación del conocimiento			
Tarea 2.3 Módulo de inferencia			
Tarea 2.4 Módulo de justificación de resultados			
Tarea 2.5 Módulo de aprendizaje			
Tarea 2.6 Módulo de consultas bibliográficas			
Tarea 3.1 Estudio, diseño y aplicación de técnicas de elicitación de conocimiento exp			
Tarea 3.2 Aplicación de métodos de aprendizaje a datos sobre bases de casos clínic			
Tarea 3.3 Elección de los diagnósticos y rasgos (Tiroides)			
Tarea 3.3 Elección de los diagnósticos y rasgos (Cérvico-vaginal)			
Tarea 3.4 Selección de las palabras clave (Tiroides)			
Tarea 3.4 Selección de las palabras clave (Cérvico-vaginal)			
Tarea 3.5 Creación de base de imágenes (Tiroides)			
Tarea 3.5 Creación de base de imágenes (Cérvico-vaginal)			
Tarea 4.1 Codificación sistemas inteligentes			
Tarea 4.2 Integración y prueba			
Tarea 4.3 Validación			
Tarea 5.1 Análisis de las necesidades docentes (ITS)			
Tarea 5.2 Análisis de funcionalidad y requisitos (ITS)			
Tarea 6.1 Diseño arquitectural (ITS)			
Tarea 6.2 Definición de casos.			
Tarea 6.3 Módulo evaluador			
Tarea 6.4 Módulo gestor de alumnos			
Tarea 7.1 Codificación (ITS)			
Tarea 7.2 Integración y prueba (ITS)			
Tarea 7.3 Validación (ITS)			

2 Nivel de éxito alcanzado en el proyecto

Atendiendo a las tareas planificadas en el cronograma anterior, el grado de ejecución del proyecto es el siguiente:

Tarea 1.1 Análisis general del problema. Meses 1 al 4 inclusive.

Durante estos meses se tuvieron diversas reuniones en el Hospital Virgen de las Nieves en Granada, en las que se hizo un análisis general del problema desde las distintas perspectivas y componentes del grupo: expertos en ingeniería del conocimiento, médicos patólogos y expertos en fondos bibliográficos. Aquí pusimos de manifiesto el gran esfuerzo que había que hacer para compaginar los distintos puntos de vista que unos y otros teníamos sobre el problema que queríamos resolver. Estas reuniones fueron exitosas en la medida en que adquirimos una visión del problema general a partir de la cual nos sería más fácil la comunicación con el resto de los componentes del equipo. Naturalmente a lo largo de estos 4 meses se analizaron trabajos anteriores en esta misma línea de investigación realizados por otros investigadores.

Tarea 1.2 Análisis de requisitos del sistema inteligente. Meses 4 al 6 inclusive.

Simultáneamente con la Tarea 1.1, se fue haciendo un análisis de los requisitos que el sistema inteligente debería cumplir. Durante estos tres meses se fueron planteando características deseables para el sistema, teniendo en cuenta el posible grado de cumplimiento de estas con las tecnologías que podíamos desarrollar. Por ejemplo, un requisito importante desde el punto de vista de su utilización, era que este sistema inteligente fuera accesible vía Web, de forma que se utilizara de forma distribuida en lugar de individual. También se requirió, entre otras características, el uso de técnicas de ingeniería del conocimiento que pudieran manejar la incertidumbre siempre presente en este tipo de problemas. Finalmente se hizo un documento con el total de requisitos que el sistema inteligente debería verificar.

Tarea 2.1 Diseño arquitectural. Meses 7 al 8 inclusive.

En esta tarea se hizo un diseño de la arquitectura del sistema inteligente. Se enumeraron los distintos módulos que contempla el sistema así como la interrelación entre los mismos.

Tarea 2.2 Módulo de representación del conocimiento. Meses 9 a 10 inclusive.

Aquí se planteó la forma de representar el conocimiento en nuestro sistema inteligente, y se optó por la forma más operativa y que mejor se acomodaba a los objetivos del proyecto: mediante el uso de *sistemas basados en reglas*, dando por tanto al sistema inteligente la arquitectura de un sistema de producción. Puesto que uno de los objetivos importantes del proyecto es la capacidad para manejar la incertidumbre, se optó por los *sistemas basados en reglas difusas*, cuyo mecanismo de inferencia basado en el *razonamiento aproximado* hace factible este fin.

Tarea 2.3 Módulo de inferencia. Meses 11 al 12 inclusive.

Como hemos dicho en la Tarea 2.2, el mecanismo de inferencia utilizado será el *razonamiento aproximado*, mecanismo de inferencia que utiliza la lógica difusa y que permite manejar la incertidumbre e imprecisión que aparece en gran parte de los problemas de la vida real.

Tarea 2.4 Módulo de justificación de resultados. Meses 9 al 10 inclusive.

En todo sistema experto es de gran importancia la justificación de la información proporcionada al usuario tras cualquier consulta por parte de éste, ya que la confianza en el sistema viene avalada en gran parte por el tipo de razonamiento y por tanto la justificación que le haga sobre la conclusión a la que ha llegado. Con el tipo de representación del conocimiento llevada a cabo en la Tarea 2.2 va

a resultar más viable este módulo de justificación gracias a la acción de causalidad (causa-acción) que rigen los sistemas basados en reglas.

Tarea 2.5 Módulo de aprendizaje. Meses 10 al 12 inclusive.

Este módulo se diseñó con el objetivo de conseguir un ciclo de vida del sistema inteligente en continuo desarrollo. Se pensó que el sistema debía incorporar nuevo conocimiento que pudiera aparecer en cualquier momento de la ejecución del sistema experto. Sobre todo se hizo especial hincapié en la incorporación de nuevo conocimiento resultante del proceso de inferencia y de la posterior validación del mismo. Por ejemplo, el sistema puede inferir un determinado diagnóstico para el problema del tiroides en base a unos determinados rasgos suministrados; tras la justificación del sistema de este diagnóstico, es posible que una validación efectuada por un experto patólogo sobre el mismo concluya que la respuesta no es la adecuada. En cuyo caso, el sistema, a través de su módulo de aprendizaje es capaz de incorporar el nuevo conocimiento a su propia base.

Tarea 2.6. Módulo de consultas bibliográficas. Meses 12 al 14 inclusive.

En el artículo "Motores de búsqueda automáticos a partir de un diagnóstico determinado", [7], nos aproximamos al método de consulta bibliográfica, estableciendo las bases y los requisitos para el mismo. Tras un arduo trabajo de definición de las principales consultas bibliográficas de interés para los anátomo-patólogos y de expresión de las mismas en los lenguajes de consulta de distintos motores de búsqueda, se implementó un sistema que automatiza la realización de dichas consultas.

El sistema adapta la forma general de la consulta al aspecto particular de interés para el facultativo, ejecuta las consultas mediante conexión directa con los motores de búsqueda e integra los resultados.

Tarea 3.1 Estudio, diseño y aplicación de técnicas de extracción de conocimiento experto. Meses 7 al 10 inclusive.

La extracción del conocimiento ha seguido tres vertientes:

a) Mediante entrevistas estructuradas. Donde los ingenieros del conocimiento a través de una entrevista previamente preparada solicitaban información a expertos patólogos.
b) Vía web. Con la finalidad de obtener el mayor número de rasgos y diagnósticos posibles para la citología de tiroides y cérvico-vaginal, se preparó una aplicación web donde se alentaba a los especialistas a colaborar en el proyecto aportando información relevante. Esta página se publicitó a través de los distintos congresos específicos que había sobre el tema.
c) Haciendo uso de la herramienta ADAUCO. ADAUCO es un prototipo software resultado de la tesis doctoral titulada "Diseño e implementación de una herramienta para la adquisición automática de conocimiento" de un miembro del equipo investigador, el Dr. José Jesús Castro (ésta tesis fue codirigida por el investigador principal del proyecto Dr. José Manuel Zurita). Los fundamentos teóricos en los que se basa esta herramienta quedaron recogidos, entre otras, en las publicaciones [2],[3].

Tarea 3.2 Aplicación de métodos de aprendizaje a datos sobre bases de casos clínicos. Meses 7 al 13 inclusive.

Esta tarea se fue realizando en paralelo con la tarea 3.1. En este caso, nos centramos en la utilización de métodos de aprendizaje automáticos a partir de ejemplos. Utilizamos distintos métodos de aprendizaje inductivo para generar reglas lingüísticas a partir de casos clínicos. Entre ellos caben destacar [1], el método de Wang y Mendel, y otros propuestos en [8] que emplean distintos modelos de redes neuronales para la construcción o derivación de reglas difusas que representen el conocimiento implícito en un conjunto de datos. El resultado de estas ejecuciones proporciona bases de reglas, cuya agregación amplía la base de conocimiento. Sin embargo, para la elaboración de la base final, son necesarias aún etapas de sincretización, refinamiento y validación.

Tarea 3.3.a Elección de los diagnósticos y rasgos (Tiroides). Meses 10 al 11 inclusive.
Esta tarea se realizó utilizando las técnicas de extracción de conocimiento previamente comentadas y un pequeño resumen queda reflejado en el trabajo "El diagnóstico en citología: Extracción de rasgos y cuantificación de los mismos" [4]. También, debido a la importancia de hacer portable el conocimiento adquirido para que otros investigadores pudieran hacer uso de él, se pensó, que una de las formas más actuales y novedosas de hacer esto era con el uso de las *ontologías*. Aunque, si bien, el uso de esta técnica no era una especificación inicial en el proyecto, se decidió intentar ampliar la capacidad de nuestro sistema en la medida de lo posible con esta nueva forma de representación del conocimiento en la web. De aquí surgió el trabajo "Representación del conocimiento mediante ontologías: Ensayo sobre citología tiroidea" [5].
Tarea 3.3.b Elección de los diagnósticos y rasgos (Cérvico-vaginal). Meses 11 al 14 inclusive.
En la publicación "Bases para la construcción de un sistema tutor inteligente para el aprendizaje de la citología cérvico-vaginal" [6] se hace una descripción de los rasgos y los diagnósticos para la citología cérvico-vaginal. Si bien, como ya dijimos, el sistema siempre está abierto a la incorporación de nuevo conocimiento.
Tarea 3.4 Selección de palabras clave. Meses 12 al 16 inclusive.
El resultado de esta tarea son sendas colecciones de palabras clave para las dos patologías de interés. Las palabras clave son esenciales para la elaboración de consultas y en la descripción del conocimiento.
Tarea 3.5 Creación de base de imágenes. Meses 14 al 22 inclusive.
Los sistemas inteligentes a desarrollar tienen una forma de interacción con el usuario fundamentalmente visual. Las imágenes de las muestras al microscopio son la principal fuente de información. La captación de un conjunto de imágenes de calidad, amplio y representativo es una tarea esencial para el desarrollo del sistema. Las imágenes son componente esencial de la base de conocimiento. Esta tarea es ardua y ciertos contratiempos relativos a la disponibilidad de muestras y problemas con los equipos de captación han implicado un retraso en su ejecución que aún no ha finalizado completamente, aunque no está muy lejana.
Tarea 4.1. Codificación de los sistemas inteligentes. Meses 13 al 20 inclusive.
El sistema preliminar orientado a la ayuda en el diagnóstico de patologías del tiroides está completamente finalizado (y accesible a través de la web). Se desarrolló usando exclusivamente herramientas de código abierto ("Open Source"). Fundamentalmente, Apache, MySQL, PHP y Javascript. El trabajo realizado ha puesto de manifiesto algunas carencias del método de desarrollo y de las herramientas empleadas. Para el segundo sistema, hemos cambiado la filosofía de desarrollo hacia algo más cercano al Ciclo de vida basado en prototipos y una herramienta que permite desarrollos de código más productivos. En particular, hemos migrado a programación en python junto con el marco de desarrollo web, Django. Esto ha supuesto un ligero retraso en el inicio de la implementación del sistema inteligente para patologías cérvico-vaginales pero avanza a un ritmo superior al proceso anterior, por lo que en los primeros meses del tercer año estará completamente disponible.
Tarea 4.2 Integración y Prueba. Meses 21 a 22 inclusive.
Esta tarea está finalizada íntegramente para el primero de los dos sistemas inteligentes (dedicado al tiroides). Está retrasada respecto al segundo sistema debido a las razones indicadas en tarea previa. El propio cambio de ciclo de vida y metodología de desarrollo también acelerará esta prueba porque la integración y pruebas van integradas en el proceso de desarrollo (heredados de la metodología "Xtreme Programming".)
Tarea 4.3 Validación. Meses 23 a 24 inclusive.
Estas dos tareas están retrasadas debido a su dependencia de las dos previas.

Tarea 5.1 Análisis de las necesidades docentes (del Sistema inteligente tutor). Meses 17 a 18 inclusive.
Se elaboró un informe por parte de los expertos en formación de citotécnicos acerca de las necesidades docentes de estos profesionales. El informe incluye una descripción de los aspectos carenciales de los métodos y procedimientos en uso actualmente implicando una posible mejora mediante su automatización.
Tarea 5.2 Análisis de funcionalidad y requisitos no funcionales. Meses 19 a 20 inclusive.
Aplicando la metodología de análisis combinada del ciclo de vida tipo "cascada" y la del ciclo de vida basado en prototipos, se extrajeron los requisitos funcionales y no funcionales que debería cumplir la aplicación ITS. Esta aplicación tiene dos componentes especiales frente al proyecto de desarrollo de software "típico": su orientación docente y la utilización de conocimiento experto para realizar la labor de guía en el aprendizaje del alumno (usuario final del sistema). Estas dos particularidades han afectado a la forma de realizar el proceso de captación de requisitos funcionales y no funcionales. La adaptación a estas circunstancias ha hecho demorar la finalización de esta tarea y de la anterior, arrastrando el retraso el inicio de las tareas de desarrollo.
Tareas 6.1 a 6.3. Desarrollo del ITS.
Todas estas tareas han demorado su inicio por la necesidad de particularización de la metodología de captación de requisitos a las características especiales de la aplicación ITS.

3 Indicadores de resultados

Los indicadores que muestran la relevancia de los resultados obtenidos hasta ahora son prometedores. En cuanto a las **publicaciones realizadas** podemos destacar:

a) J. Esquivias, J. Alonso, J.M. Benítez, M. Parra, F. Jiménez, C. Prieto, M. Merino. El diagnóstico en citología: Extracción de rasgos y cuantificación de los mismos. Ensayo sobre citología tiroidea, 2005. VII Congreso Virtual Hispanoamericano de Anatomía Patológica.
b) J. Esquivias, J.M. Zurita, M. Gómez, E. Villar, J.M. Sánchez, J. Alonso. Representación del conocimiento mediante ontologías: Ensayo sobre citología tiroidea, 2005. VII Congreso Virtual Hispanoamericano de Anatomía Patológica.
c) J. Esquivias, M.M. Serrano, J. Alonso, M. Parra, C. Nieves, C. Prieto, M. Merino, J.M. Benítez, C. Herrera. Bases para la construcción de un sistema tutor inteligente para aprendizaje de la citología cérvico-vaginal, 2005. VII Congreso Virtual Hispanoamericano de Anatomía Patológica.
d) M. Rodíguez del Castillo, J.M. Benítez, J.M. Fernández-Luna, J.M. Zurita, C. Herrera, F. Jiménez Burgos, M.M. Serrano, J. Merino, J. Esquivias**.** Motores de búsqueda automáticos a partir de un diagnóstico determinado, 2005. VII Congreso Virtual Hispanoamericano de Anatomía Patológica.

En cuanto a desarrollos:

a) El sistema de ayuda al diagnóstico de patologías de tiroides está totalmente desarrollado, pendiente de finalizar su validación. Esta aplicación ya ha sido probada por varios expertos nacionales que han emitido su opinión favorable sobre el nivel de resolución de la herramienta. La presentación del sistema en diversos congresos del área ha sido bien recibida.

b) El sistema de ayuda al diagnóstico de patologías cérvico-vaginales está en la fase final de implementación-integración-prueba, todo conjuntamente debido a la metodología de desarrollo empleada en este caso. El retraso en el inicio de la implementación se compensa con un ritmo más acelerado (y también más seguro). En breve estará disponible para su validación por un nuevo grupo de expertos.
c) El sistema de ayuda en la formación de citotécnicos ha superado ya la fase de especificación de requisitos y se espera su conclusión para el tercer trimestre del año final del proyecto.

Con respecto a la **transferencia tecnológica y colaboración con otros grupos de trabajo**, hay que destacar que la vertiente práctico-investigadora de este proyecto ha determinado una arquitectura de sistema inteligente con capacidad de implementación en otras áreas de la medicina. A partir de este proyecto y mediante colaboraciones con otros profesionales médicos surgió la realización de otro gran proyecto de transferencia tecnológica, **El Paciente Virtual.** Este proyecto lo están desarrollando un grupo de investigadores de la Universidad de Granada (dentro del cual están tres de los miembros del proyecto que nos ocupa), la Fundación IAVANTE (Fundación para el Avance Tecnológico y Entrenamiento Profesional) de la Consejería de Salud de la Junta de Andalucía y CITIC (Centro Andaluz de Innovación y Tecnologías de la Información y las Comunicaciones).

4 Bibliografía

[1] J.L. Castro, J.J. Castro-Schez, J.M. Zurita. Use of a fuzzy machine learning technique in the knowledge acquisition process, 2001. Fuzzy Sets and Systems, 123 , pp. 307-320.
[2] J.L. Castro, J.J. Castro-Schez, J.M. Zurita. Combining both a fuzzy inductive learning and a fuzzy repertory gris method, 2002, en Technologies for constructing intelligent systems 1, Task, B. Bouchon-Meunier, J. Gutiérrez-Ríos, L. Magdalena, R.R. Yager (editores).
[3] J.J. Castro-Schez, J.L. Castro, J.M. Zurita. Fuzzy repertory table: A method for acquiring knowledge about input variables to machina learning algorithm, 2004. IEEE Transactions on Fuzzy Systems, vol. 12, nº 1, pp. 123-139.
[4] J. Esquivias, J. Alonso, J.M. Benítez, M. Parra, F. Jiménez, C. Prieto, M. Merino. El diagnóstico en citología: Extracción de rasgos y cuantificación de los mismos. Ensayo sobre citología tiroidea, 2005. VII Congreso Virtual Hispanoamericano de Anatomía Patológica.
[5] J. Esquivias, J.M. Zurita, M. Gómez, E. Villar, J.M. Sánchez, J. Alonso. Representación del conocimiento mediante ontologías: Ensayo sobre citología tiroidea, 2005. VII Congreso Virtual Hispanoamericano de Anatomía Patológica.
[6] J. Esquivias, M.M. Serrano, J. Alonso, M. Parra, C. Nieves, C. Prieto, M. Merino, J.M. Benítez, C. Herrera. Bases para la construcción de un sistema tutor inteligente para aprendizaje de la citología cervico-vaginal, 2005. VII Congreso Virtual Hispanoamericano de Anatomía Patológica.
[7] M. Rodíguez del Castillo, J.M. Benítez, J.M. Fernández-Luna, J.M. Zurita, C. Herrera, F. Jiménez Burgos, M.M. Serrano, J. Merino, J. Esquivias**.** Motores de búsqueda automáticos a partir de un diagnóstico determinado, 2005. VII Congreso Virtual Hispanoamericano de Anatomía Patológica.
[8] J.M. Benítez. "Extracción de reglas difusas de y con redes neuronales artificiales", Tesis Doctoral, Universidad de Granada, 1998.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

TIEMPO: IT Handling of Multilingual Financial and Economic Information TIN2004-07246-C03

M. Vilares Ferro [*] Univ. of Vigo
M.A. Alonso Pardo [†] Univ. of A Coruña
V.J. Díaz Madrigal [‡] Univ. of Sevilla

Abstract

The phenomenon of globalisation in the access to information forces us to face the challenge of adapting to it. There are four main reasons for this. Firstly, as a result of the overwhelming avalanche of relevant information now available. Secondly, due to the often critical nature of the decision-making process. Thirdly, as a consequence of a specific work environment in terms of the terminology, syntax and semantics employed. Finally, caused by the need to manage information in various languages. In this context, financial markets are one of the most sensitive sectors.
Keywords: Information recovery, knowledge acquisition, named entity recognition, natural language processing, multi-lingual access, parsing, tagging, term extraction, text alignment.

1 Objectives and project aim

Our aim is the development of an environment for extracting, managing and evaluating information from texts concerning financial markets in multi-lingual environments. As a complement, we contemplate a multi-lingual search for information structured around a natural language interface, and the generation of linguistic resources.

Such a proposal translates into a complex task requiring the conjunction of efforts in a number of specialized domains with particular goals that we can classify in three different levels: lexical, syntactic, and semantic. More in detail, our strategy includes a lexical phase whose aim is to identify named entities, terms and collocations in the text. Here, we have considered both available free tools and others developed by the research groups themselves.

The syntactic phase follows the lexical one, and its goal is to identify relevant components and relations between them, in the text. In order to deal with ungrammatical inputs, commonly found in the field of *natural language processing* (NLP), we consider shallow and robust parsing techniques. Resulting partial trees will be the basis for the semantical phase, in particular in order to allow automatic knowledge acquisition.

[*]Email: vilares@uvigo.es
[†]Email: alonso@udc.es
[‡]Email: vjdiaz@lsi.us.es

The multi-lingual character of our proposal leads us to provide strategies to pass efficiently from one language to another in the domain of *information retrieval* (IR) facilities. Having discarded the use of complex machine translation tools, we are exploring other alternatives, such as text alignment, which we could consider not only in the treatment of multi-lingual IR activities on parallel texts, but also as a basis for generating glossaries and dictionaries of interest to index generation in this context. Finally, the lack of multi-lingual corpora in the languages and domain in question, in particular in dealing with parallel texts, has obliged us to generate these linguistic resources ourselves.

2 Working tasks and success level

2.1 Lexical level

Tagging Tools. Tagging provides most of the information needed to identify and classify tokens from documents, which implies disambiguating lexical forms. We have used a variety of taggers in our tests. MrTagoo, an Hmm-based tagger, currently available for Spanish and Galician, has proved to be an adequate experimental platform to deal with IR related tools [1, 26, 49, 50], additionally including new operational capabilities such as spelling correction [17, 16, 20, 22, 21] and synonymy management facilities [45]. In order to develop generic strategies, we are also using TreeTagger and TnT, both free tools.

Automatic corpus annotation. The availability of large tagged data corpora is an essential aspect in most NLP tasks. The effort required to manually tag this number of phrases has encouraged many researchers like ourselves to create automatic applications for this issue. Our approach [28, 32] represents a completely automatic method for enlarging an already existing corpus, so it acquires the desired number of tagged phrases. The extra content of the corpus will be obtained from any knowledge source, such as the web, from which we extract untagged sentences to be analyzed. Considering the initial small corpus as the seed, our method makes it evolve until it reaches the target size in a totally automatic way.

Named entity recognition (NER). We are interested in identifying and labeling semantic information in a text, in such a way as to allow repeatable semantic patterns to emerge. So, we seek to locate and classify atomic elements in text into predefined categories such as the names of persons, organizations, locations, expressions of time, quantities, monetary values or percentages. The more complex the ontology, the less accurate the classification, thus originating semantic and performance problems [2]. To alleviate this problem one can increase the size of the annotated corpus, but this is a very expensive task so we have designed a system [34] which provides a number of features which facilitate the visualization and tagging of annotated text. In the development of our NER system [7, 8], we have used the Spanish corpus distributed for the CoNLL02 shared task, and the tagger generator TnT. In order to improve performance, we have defined three transformations that give us additional versions of the training corpus, and we have trained TnT with them to obtain different taggers. Then we have applied a stacking scheme [9] to combine the results of the models.

Term extraction. In dealing with French we have considered Fastr[1] and Acabit[2]. The former is a term analyzer and a variant recognizer. Given that Fastr does not implement term

[1] http://www.limsi.fr/Individu/jacquemi/FASTR/
[2] http://www.sciences.univ-nantes.fr/info/perso/permanents/daille/index.html

extraction in Spanish, we have adapted the grammatical meta-rules and the set of tags used in French to Spanish, profiting from the syntactic similarity between these two languages. ACABIT is a term acquisition tool that takes as its input a formated text and returns an ordinated list of candidate terms. Given that ACABIT uses the FLEMM tagger, we have adapted the input interface in order to allow inputs from TREETAGGER. In the case of Spanish, we have also considered two different approaches. The first is a tool developed *ad-hoc* at the SYNTAX research group in the Univ. of Santiago de Compostela and it uses a TREETAGGER entry.

Spelling correction. We have also worked on the correction of the errors[3] present in the document collection, in order to make the error rate of the parser as low as possible. In this context, we avoid techniques traditionally implemented in commercial applications, based on a hand-made correction according to a list of candidate replacement options. We propose to apply fully automatic techniques, searching for repairs in a limited context which is dynamically determined for each error detected. We baptize this kind of error recovery technique as *regional* [17, 16, 20, 22, 21].

2.2 Syntactic level

Shallow parsing. This is a kind of "light parsing" where the analysis of a sentence seeks to identify the main constituents, also called chunks, but does not specify their internal structure, nor their role in the main sentence. We have developed a shallow parser, implemented by means of finite-state transducers with a 5-layer architecture [11, 12]. Once the parsing has finished, the syntactic function of each phrase detected is identified. Then, the dependencies between the heads of phrases are extracted [11, 12]. These dependencies are general enough to be present in any of the target languages of the project [10, 60, 59], taking into account that our work focuses on IR applications [62, 47, 61, 13]. Trying to make the technique as general as possible, we have also developed a technique for the compilation of parsing schemata [29, 66, 30], a declarative, high-level formalism for the description of parsing algorithms that can be used for any grammar in the Chomsky hierarchy [53, 54]. The goal is their application in the NLP domain [31] and, in particular, in the case of IR tools [55].

Robust parsing. This is the ability to find partial parses in an ungrammatical input, in such a way as to enable parsing to continue in the presence of parts of the input unexplained by the language grammar. Typically this involves producing a parse tree even in the presence of syntax errors, corrupted or unknown sentences. Our proposal is a regional least-cost strategy which applies a dynamic validation in order to avoid cascaded errors, gathering all relevant information in the context of the error location. The system guarantees the asymptotic equivalence with global repair strategies [15, 18].

2.3 Semantic level

Knowledge acquisition. We focus [52] on extracting and then connecting terms in order to detect pertinent relations and eliminate non-deterministic interpretations. To deal with, two principles are considered: the *distributional semantics model* establishing that words whose meaning is close often appear in similar syntactic contexts; and the assumption that terms shared by these contexts are usually nouns and adjectives. The parse takes the form of a graph

[3] typos, conversion errors, OCR errors, etc.

whose arcs represent relations of the type *governor/governed*, which permits the lexicon to be concentrated around *pivot terms* and even establishes similarity measures between these. Term extraction is organized around the recognition of generic lexical and/or syntactic patterns from these pivot terms. We profit from this topological information to apply automatic learning techniques in order to locate those dependencies that are more frequent and less ambiguous, focusing the meaning of the text on what we baptize as *strong dependencies*. At this point, we can infer a number of semantic tags that we use for text indexing. Also, once basic syntagmas and properties have emerged from text, we focus on more sophisticated patterns connecting them in order to derive more complex semantical relations as, for example, *hyperonymy*.

Parallel text alignment. Alignment is usually done by finding correspondence points. Early works used punctuation signs as alignment points, dividing texts into parallel sentences. Later, homograph tokens in both texts were used as correspondence points and enabled the identification of segments smaller than the sentence. Other methods have relied on using cognates, i.e. similar words having the same meaning as additional correspondence points. In particular, it is our aim to evaluate how cognates affect the quality of the results. For this purpose we used a simple measure of similarity which accepts as possible cognates all pairs whose similarity is above a certain threshold. Then, we studied [51] how the alignment is affected by changing this threshold. We have developed a methodology to assess the quality of resulting alignments, determining the most probable causes of misalignment and evaluating how these errors are affected by changes in our cognates' thresholds.

Ranking documents. We have also investigated [3, 27] how to adapt the TextRank method to make it work in a supervised way. TextRank is a graph based method that applies the ideas of the ranking algorithm used in Google (PageRank) to nlp tasks, such as text summarization, keyword extraction or word sense disambiguation. TextRank operates in an unsupervised way, without using any training corpus. Our main contribution is the definition of a method that allows us to apply TextRank to a graph that includes information generated from a training tagged corpus.

2.4 Indexing models

A syntactic-dependency based model. A model based on the use of the syntactic dependencies extracted by the shallow parsing, jointly with single terms, has been developed. We have considered the use of syntactic dependency pairs obtained from the topic and the use of syntactic dependency pairs obtained from top documents retrieved. Our experiments [10] indicate that the use of syntactic information is useful for refining the results obtained through single-terms, but improvement is limited by the noise introduced by syntactic dependencies when they are not accurately selected.

A pseudo-syntactic model. A pseudo-syntactic approach was also tested [11, 12, 62]: the distance between terms appears as a practical alternative that avoids problems arising from a grammar or a parser, and is also language-independent. In addition, the information about the occurrence of individual words can be integrated in a consistent way with the information about proximity to other terms, which in turn is often related with the existence of syntactic relations between such terms. Then, we have applied a locality-based model that considers the collection to be indexed not as a set of documents, but as a sequence of words where each occurrence of a query term has an influence on the surrounding terms. Such influences are additive, thus, the contributions of different occurrences of query terms are summed, yielding

a similarity measure. As a result, those areas of the texts with a higher density of query terms, or with important query terms, show peaks in the resulting graph, highlighting those positions of the text which are potentially relevant with respect to the query. It is worth noting that relevant portions are identified without the need to perform any kind of splitting in the documents, as is done in passage retrieval. Our experiments on this approach did not improve the results obtained by the syntactic-dependency based model.

A data-fusion based model. We have also merged the two previous approaches, using data fusion [61, 47, 13], a technique of combination of evidences that consists of combining the results retrieved by different representations of queries or documents, or by different retrieval techniques. In our approach, we have opted for using the following boolean criterion: First, we retrieve the documents contained in the intersection of the top documents retrieved by both approaches. Next, the documents retrieved in the top documents by only one of the approaches are retrieved. Finally, the rest of documents retrieved using single-terms. In our experiments, the improvements attained with this new approach were general, particularly in the case of precision at N documents, without penalizations for non-interpolated and R-precision.

Cross-lingual retrieval. We have developed [60, 59] an n-gram model that takes as input previously existing aligned word lists and obtains as output aligned n-gram lists. It can also handle word translation probabilities, as in the case of statistical word alignments. This solution avoids the need for word normalization during indexing or translation, and it can also deal with out-of-vocabulary words. Since it does not rely on language-specific processing, it can be applied to very different languages, even when linguistic information and resources are scarce or unavailable. Our proposal adds a high speed during the n-gram alignment process to these characteristics.

A sentence matching based model. We exploit the meaning of single-terms by integrating them into an edit distance construction. So, we extend a matching algorithm in a similarity parse tree measure taking into account the semantic proximity between words [14, 5, 6, 56]. This allows us to deal with similarity problems where we can take advantage of the use of semantic information in pattern-matching processes. We have chosen to compute a modified edit distance where the data tree can be simplified by removing some irrelevant subtrees with no associated cost. We also support the use of *variable length don't care* (VLDC) symbols in the pattern tree, which allow us to omit structural details and manage more general patterns. To formalize semantic similarity at word level, we use the WordNet taxonomy. On this basis we compute the semantic cost associated with the edit operations applied to the words in the sentences we are comparing. Semantic and syntactic distances are computed in parallel, propagating the similarity measure at word level through the nodes in accordance with the syntactic distances computed by tree matching.

3 Result Indicators

3.1 Formation Activities

During the project, J. Vilares (Udc) has finished the Ph.D. Thesis entitled *Application of Natural Language Processing to Spanish Information Retrieval*. Co-directed by M.A. Alonso (Udc) and J.L. Freire (Udc), 2005. The thesis was awarded the highest mark of "Sobresaliente Cum Laude por Unanimidad", the European Doctor Mention and the UDC Doctorate Prize.

The examination committee was formed by Gabriel Pereira Lopes (New University of Lisbon, Portugal), John I. Tait (University of Sunderland, UK), Éric de la Clergerie (INRIA, France), C. Martín Vide (Rovira i Virgili University, Spain) and J. Graña Gil (UDC, Spain).

The following Ph.D. Thesis are expected to be finished towards the end of the project:

1. F.M. Barcala (UVIGO). Co-directed by J. Graña (UDC) and M. Vilares (UVIGO), he works on improving the management of linguistic information for practical applications by developing specific IR algorithms and techniques for structured collections.
2. C. Gómez (UDC), recipient of a FPU fellowship from the MEC since May 2006, is also working on his Ph.D. thesis, with the aim of designing a general technique for generating robust parsers from declarative specifications in order to build practical natural language tools applicable to the fields of IR and QA. He has joined the project in May 30, 2006. The thesis is being directed by M.A. Alonso (UDC) and M. Vilares (UVIGO).
3. J. Otero (UVIGO), recipient of a FPI fellowship from the MEC since 2005, is working on his Ph.D. thesis under the direction of J. Graña (UDC) and M. Vilares (UVIGO), with the aim of designing a general technique for generating robust spelling correction techniques.
4. M.A. Molinero (UVIGO) has presented his DEA (UDC), and is recipient of a pre-doctoral fellowship from the Xunta de Galicia. He is working on the application of complex lexical and syntactic information to IR. This thesis is being directed by M. Vilares (UVIGO).

Additionally, M. Fernández (UVIGO), C. Gómez (UVIGO, currently FPU recipient in UDC) and S. Carrera (UVIGO) have been hired to work on the project and they will develop their Ph.D. Theses on subjects related to it. Finally, F. Enríquez (USE), F. Cruz (USE) and C. Gómez (UDC) will present their DEAs this year from the results of the project.

3.2 Publications

Articles in journals indexed by ISI JCR

[1] F.M. Barcala, M.A. Molinero, and E. Domínguez. Information retrieval and large text structured corpora. *Lecture Notes in Computer Science*, 3643:91–100, 2005.

[2] J.M. Cañete and F.J. Galán. Towards a theory on the role of ontologies in software engineering problem solving. *Lecture Notes in Computer Science*, 3442:205–219, 2005.

[3] F. Cruz, J.A. Troyano, and F. Enríquez. Supervised textrank. *Lecture Notes in Artificial Intelligence*, 4139:632–639, 2006.

[4] E. Méndez, J. Vilares, and D. Cabrero. COLE experiments at QACLEF 2004 spanish monolingual track. *Lecture Notes in Computer Science*, 3491:544–551, 2005.

[5] F.J. Ribadas, M. Vilares, and J. Vilares. Semantic similarity between sentences through approximate tree matching. *Lecture Notes in Computer Science*, 3523:638–646, 2005.

[6] F.J. Ribadas, J. Vilares, and M.A. Alonso. Integrating syntactic information by means of data fusion techniques. *Lecture Notes in Computer Science*, 3643:169–178, 2005.

[7] J.A. Troyano, V. Carrillo, F. Enríquez, and F.J. Galán. Named entity recognition through corpus transformation and system combination. *Lecture Notes in Computer Science*, 3230:255–266, 2004.

[8] J.A. Troyano, V.J. Díaz, F. Enríquez, and L. Romero. Improving the performance of a named entity extractor by applying a stacking scheme. *Lecture Notes in Computer Science*, 3315:295–304, 2004.

[9] J.A. Troyano, V.J. Díaz, F. Enríquez, and V. Carrillo. Applying stacking and corpus transformation to a chuncking task. *Lecture Notes in Computer Science*, 3643:150–158, 2005.

[10] J. Vilares, M.A. Alonso, and F.J. Ribadas. COLE experiments at CLEF 2003 Spanish monolingual track. *Lecture Notes in Computer Science*, 3237:345–357, 2004.

[11] J. Vilares and M.A. Alonso. Dealing with syntactic variation through a locality-based approach. *Lecture Notes in Computer Science*, 3246:255–266, 2004.

[12] J. Vilares, M.A. Alonso, and M. Vilares. Morphological and syntactic processing for text retrieval. *Lecture Notes in Computer Science*, 3180:371–380, 2004.

[13] J. Vilares, M.P. Oakes, and J.I. Tait. A first approach to CLIR using character n-grams alignment. *Lecture Notes in Computer Science*, 2006.

[14] M. Vilares, F.J. Ribadas, and J. Vilares. Phrase similarity through the edit distance. *Lecture Notes in Computer Science*, 3180:306–317, 2004.

[15] M. Vilares, V.M. Darriba, and J. Vilares. Parsing incomplete sentences revisited. *Lecture Notes in Computer Science*, 3945:102–111, 2004.

[16] M. Vilares, J. Otero, and J. Graña. On asymptotic finite-state error repair. *Lecture Notes in Computer Science*, 3246:271–272, 2004.

[17] M. Vilares, J. Otero, and J. Graña. Automatic Spelling Correction in Galician. *Lecture Notes in Computer Science*, 3230:51–57, 2004.

[18] M. Vilares, V.M. Darriba, J. Vilares, and F.J. Ribadas. A formal frame for robust parsing. *Theoretical Computer Science*, 328:171–186, 2004.

[19] M. Vilares, J. Otero, and J. Graña. Spelling correction on technical documents. *Lecture Notes in Computer Science*, 3643:131–139, 2005.

[20] M. Vilares, J. Otero, and J. Graña. Regional finite-state error repair. *Lecture Notes in Computer Science*, 3317:269–280, 2005.

[21] M. Vilares, J. Otero, and J. Graña. Robust spelling correction. *Lecture Notes in Computer Science*, 3845:319–328, 2006.

[22] M. Vilares, J. Otero, and J. Graña. Regional vs. global finite-state error repair. *Lecture Notes in Computer Science*, 3317:120–131, 2005.

[23] M. Vilares, J. Otero, and V.M. Darriba. Regional vs. global robust spelling correction. *Lecture Notes in Computer Science*, 3878:575–586, 2006.

Articles in other journals and books

[24] M.A. Alonso, E. de la Clergerie, V.J. Díaz, and M. Vilares. Relating tabular parsing algorithms for LIG and TAG. in Harry Bunt, John Carrol and Giorgio Satta (eds.), *New Developments in Parsing Technology* volume 23 in *Text, Speech and Language Technology Series*, chapter 8, pp. 157–184. Kluwer Academic Publishers, 2004.

[25] M.A. Alonso, J. Vilares, and F.J. Ribadas. Experiencias del grupo COLE en la aplicación de técnicas de procesamiento del lenguaje natural a la recuperación de información en español. *Inteligencia Artificial*, 22:123–134, 2004.

[26] F.M. Barcala, M.A. Molinero, and E. Domínguez. Construcción de sistemas de recuperación de información sobre córpora textuales estructurados de grandes dimensiones. *Procesamiento del Languaje Natural*, 34:41–48, 2005.

[27] F. Cruz, J.A. Troyano, F. Enríquez, and F.J. Ortega. Textrank como motor de aprendizaje en tareas de etiquetado. *Procesamiento del Lenguaje Natural*, 37:33–40, 2006.

[28] F. Cruz, J.A. Troyano, F. Enríquez, and F.J. Ortega. Ampliación automática de corpus mediante la colaboración de varios etiquetadores. *Procesamiento del Lenguaje Natural*, 37:11–18, 2006.

[29] C. Gómez-Rodríguez, J. Vilares, and M.A. Alonso. Generación automática de analizadores sintácticos a partir de esquemas de análisis. *Procesamiento del Lenguaje Natural*, 35:401–408, 2005.

[30] C. Gómez-Rodríguez, M.A. Alonso, and M. Vilares. Estudio comparativo del rendimiento de analizadores sintácticos para gramáticas de adjunción de árboles. *Procesamiento del Lenguaje Natural*, 37:179–186, 2006.

[31] C. Gómez-Rodríguez, J. Vilares, and M.A. Alonso. Automatic generation of natural language parsers from declarative specifications. *Frontiers in Artificial Intelligence and Applications*, 112:259–260, 2006.

[32] J. González, D. González, and J. A. Troyano. Adaptación del método de etiquetado no supervisado tbl. *Procesamiento del Lenguaje Natural*, 37:5–10, 2006.

[33] D. Guiziou, L. Félix, J. A. Gallegos, E. Sánchez, and Cr. Valderrey. *Fisioterapia: Glosario de convenciones textuales*. 2004.

[34] F. Ortega, V. Díaz, and L. Romero. Named: Una herramienta para la edición y manipulación de corpus. *Procesamiento del Lenguaje Natural*, 37:365–366, 2006.

[35] L. Rodríguez and C. García. Application of fusion techniques to speaker authentication over IP networks. IEE *Proceedings - Vision, Image and Signal Processing, I*, 2004.

[36] E. Sánchez. *Teoría de la traducción: convergencias y divergencias*. Servicio de Publicaciones de la Univ. de Vigo, 2004.

[37] E. Sánchez. Traducción de textos médicos entre el francés y el español: creación y explotación de corpus electrónicos. *Anales de Filología Francesa*, 12:395–412, 2004.

[38] E. Sánchez. Traducción de textos médicos: elaboración de un corpus electrónico francés-español. *Le français face aux défis actuels. Histoire, langue et culture*, 7:425-438, ISBN: 84-338-3238-7. 2004.

[39] E. Sánchez. Aproximación traductológica al análisis de corpus para el estudio de las convenciones textuales. In *Panorama actual de la investigación sobre traducción e interpretación*, pp. 113–127. Editorial Atrio, Granada, Spain, 2004.

[40] E. Sánchez. Corpus electrónicos y traducción: aplicación de las nuevas tecnologías a la traducción de textos médicos (francés-español). In *Insights into Scientific and Technical Translation*, pp. 267–273. Univ. Pompeu Fabre, Granada, Spain, 2004.

[41] E. Sánchez. Investigación traductológica en la traducción científica y técnica. *Revista de traductología*, 9:131–150, 2005.

[42] E. Sánchez. Traducción: de la teoría a las aplicaciones prácticas. In *Estudios de traducción, lingüística y filología dedicados a Valentín García Yebra*, pp. 203–218. 2005.

[43] E. Sánchez. Multidimensionalidad de la traducción científica y técnica entre el francés y el español. In *Manual de Traducción científica y técnica*. 2006.

[44] E. Sánchez. MYOCOR: creación y explotación de un corpus bilingüe (FR-ES) sobre enfermedades neuromusculares. *Revista de Traduçao Científica e Técnica*, 4:67–83, 2006.

[45] A. Sobrino, S. Fernández, and J. Graña. Access to a large dictionary of spanish synonyms: a tool for fuzzy information retrieval. *Studies in Fuzziness and Soft Computing*, 197:299–316, 2006.

[46] J. Vilares, C. Gómez-Rodríguez, and M.A. Alonso. Enfoques sintáctico y pseudo-sintáctico para la recuperación de información en español. In *Recuperación de información textual: aspectos lógicos y ecológicos — Text Information Retrieval: Soft-Computing and Ecological Aspects*, pp. 97–104. Servizo de Publicacións e Intercambio Científico, Univ. de Santiago, Santiago de Compostela, Spain, 2006.

[47] J. Vilares and M.A. Alonso. Tratamiento de la variación sintáctica mediante un modelo de recuperación basado en localidad. *Procesamiento del Lenguaje Natural*, 36, 2006.

Proceedings of international conferences

[49] F.M. Barcala, M.A. Molinero, and E. Domínguez. Information retrieval and large text structured corpora. In *Proc. of Tenth Int. Conf. on Computer Aided Systems Theory*, pp. 55-57, 2005.

[50] F.M. Barcala, M.A. Molinero, and E. Domínguez. XML rules for enclitic segmentation. In *Proc. of Eleventh Int. Conf. on Computer Aided Systems Theory*, 2007.

[51] G.P. Lopes, V.M. Darriba and T. Ildefonso. Measuring the impact of cognates in parallel text alignment. In *Proc. of 12th Portuguese Conf. on Artificial Intelligence*, pp. 334–343, Covilha, Portugal, 2005. Universidade de Beira Interior.

[52] M. Fernández, E. de la Clergerie and M. Vilares. From text to knowledge. In *Proc. of Eleventh Int. Conf. on Computer Aided Systems Theory*, Spain, 2007.

[53] C. Gómez-Rodríguez, M.A. Alonso, and M. Vilares. Generating XTAG parsers from algebraic specifications. In *Proceedings of the 8th Int. Workshop on Tree Adjoining Grammar and Related Formalisms*, pp. 103–108, East Stroudsburg, PA, 2006. ACL.

[54] C. Gómez-Rodríguez, M.A. Alonso, and M. Vilares. On theoretical and practical complexity of TAG parsers. In *Proc. of the 11th conference on Formal Grammar*, pp. 61–75. Center for the Study of Language and Information, Stanford, 2006.

[55] C. Gómez-Rodríguez, M.A. Alonso, and M. Vilares. Generation of indexes for compiling efficient parsers from formal specifications. In *Proc. of Eleventh Int. Conf. on Computer Aided Systems Theory*, 2007.

[56] F.J. Ribadas, J. Vilares, and M.A. Alonso. Integrating syntactic information by means of data fusion techniques. In *Proc. of Tenth Int. Conf. on Computer Aided Systems Theory*, pp. 96–99, 2005.

[57] F.J. Ribadas, E. Lloves, and V.M. Darriba. Multiple label text categorization on a hierarchical thesaurus. In *Proc. of Eleventh Int. Conf. on Computer Aided Systems Theory*, 2007.

[58] E. Sánchez. Le corpus électronique MYOCOR: Application des nouvelles technologies à la traduction des textes médicaux. In *Actes du XVIIe Congrès mondial Fédération Int. des Traducteurs*, pp. 155–157, 2005.

[59] J. Vilares, M.P. Oakes, and J.I. Tait. CoLesIR at CLEF 2006: Rapid prototyping of a n-gram-based CLIR system. In Carol Peters Alessandro Nardi and José Luis Vicedo, editors, *Results of the CLEF 2006 Cross-Language System Evaluation Campaign, Abstracts of the* CLEF *2006 Workshop*, Alicante, Spain, 2006.

[60] J. Vilares, M.P. Oakes, and J.I. Tait. CoLesIR at CLEF 2006: Rapid prototyping of a n-gram-based CLIR system (abstract). In Carol Peters Alessandro Nardi and José Luis Vicedo, editors, *Results of the CLEF 2006 Cross-Language System Evaluation Campaign, Abstracts of the* CLEF *2006 Workshop*, Alicante, Spain, 2006.

[61] J. Vilares, C. Gómez-Rodríguez, and M.A. Alonso. Syntactic and pseudo-syntactic approaches for text retrieval. In Vicente P. Guerrero-Bote, editor, *First Int. Conf. on Multidisciplinary Information Sciences and Technologies*, pp. 104–108, Badajoz, Spain, 2006. Open Institute of Knowledge.

[62] J. Vilares, C. Gómez-Rodríguez, and M.A. Alonso. Managing syntactic variation in text retrieval. In Peter R.King, editor, *Proc. of the 2005* ACM *Symposium on Document Engineering*, pp. 162–164, New York, USA, 2005. ACM Press.

[63] M. Vilares, J. Otero and J. Graña. Spelling Correction on Technical Documents In *Proc. of Tenth Int. Conf. on Computer Aided Systems Theory*, pp. 73–76, 2005.

[64] M. Vilares, J. Otero and J. Vilares. Robust spelling correction. In *Proc. of Tenth Int. Conf. on Implementation and Application of Automata*, pp. 151-159, 2005.

Proceedings of national conferences

[65] E. Sánchez. Creación y explotación de recursos documentales sobre enfermedades neuromusculares. In *Actas* XXII *Congreso de* ASEM, pages 85–94, 2005.

[66] C. Gómez-Rodríguez, J. Vilares, and M.A. Alonso. Compilación eficiente de esquemas de análisis sintáctico. In Francisco Javier López Fraguas, editor, *Actas de* PROLE'05, pages 151–159, Madrid, Spain, 2005. Thomson Paraninfo.

3.3 Collaboration with other research groups

1. **The** ATOLL **research group,** INRIA **(France)**: Leading group in human language technology, directed by E. de la Clergerie.

   (a) E. de la Clergerie visited UDC and UVIGO from March 13 to 20, 2006. He gave two talks about "From meta-grammars to factorized grammars" and "Error mining in parsing output of large corpora". The aim was to create a larger consortium in order to present a project proposal for the VII Framework Programme.

   (b) In May 2005, E. de la Clergerie visited UDC to participate in the examination committee of the Ph.D. of J. Vilares.

   (c) E. de la Clergerie visited UVIGO from November 14 to November 16, 2004 to give the talk "DYALOG: A frame for robust treatment in NLP. Generation of meta-grammars. Knowledge acquisition from texts written in natural language".

(d) M. Fernández (UVIGO) has done a 5-month post-graduate stay (from February 27 to July 31, 2006) and a visit (from November 22 to December 19, 2006) to work on methods and techniques for automatic knowledge acquisition from texts.

(e) F.J. Ribadas (UVIGO) has done several post-doctoral stays (from May 31 to June 30, 2004; from November 18 to December 23, 2004; from November 23, 2005 to March 4, 2006) to work on the integration of linguistic knowledge in IR tools.

(f) In September 2004 and August/September 2005 M. Vilares (UVIGO) visited the group with the aim of studying future collaborations in the domains of automatic knowledge acquisition and metagrammar-based parsers, respectively.

2. **The GLINT research group, New Univ. of Lisbon (Portugal)**: Leading group in human language technology, directed by G. Pereira Lopes.

   (a) In May 2005, G. Pereira visited UDC to participate in the examination committee of the Ph.D. of J. Vilares.

   (b) G. Pereira visited UVIGO from November 14 to November 16, 2004. He gave a talk about "Text alignment and its application to the development of machine translation tools".

   (c) V.M. Darriba (UVIGO) has done a 5-month post-graduate stay (June 23-November 23, 2005) and a visit (January 7-February 18, 2005), to work on parallel-texts alignment.

   (d) J. Otero (UVIGO) has done two 3-month post-graduate stay (March 1-May 31, 2006; September 16-December 16, 2005) to work on the integration of syntactic information in spelling correction tools.

3. **The Information Retrieval Group of the School of Computing and Technology, Univ. of Sunderland (UK)**: Leading group in human language technology, directed by Prof. J.I. Tait.

   (a) J. Vilares (UDC) has done a 5-month post-doctoral stay (from September 14, 2005 to February 15, 2006) with this group, working on methods and techniques for cross-language IR. He also gave a talk about "Managing Syntactic Variation in IR: Spanish as a Case in Point" in October 5, 2005. In late 2006, he again visited Tait's group for a 3-month post-doctotal stay.

   (b) In May 2005, J.I. Tait visited UDC to participate in the examination committee of the Ph.D. of J. Vilares.

   (c) M. Oakes visited UDC and UVIGO from March 13 to 20, 2006. He gave a talk about "Regular sound-changes for Cross-Language Information Retrieval". The aim was to create a larger consortium in order to present a project proposal for the VII Framework Programme.

4. **The I3S Laboratory, CNRS/Univ. of Nice**: Leading laboratory in computer languages technology. In September, 2006, M. Vilares (UVIGO) visited the I3S Laboratory invited by Dr. Jacques Farre. The goal of this visit was the study of future collaborations with I3S and INRIA in the domain of tree-adjoining grammar parsers.

5. **The Dept. of Computer Languages and Systems, Univ. of Alicante**: F. Cruz (USE) has started a collaboration with Paloma Moreda in the area of Semantic Role Labelling (SRL). We have already planned a visit for her to our department to help us in the beginning of our investigations.
6. **The Univ. of Rochester, NY (USA)**: F. Enríquez de Salamanca (USE) has received an invitation from Mr. James F. Allen (URCS Faculty Member) to spend an unspecified period of time working with his research group, in order to share knowledge and find tasks in which we could collaborate in the near future.
7. **The Group of Spanish Syntax, Univ. of Santiago de Compostela**: Leading group in computational linguistics, directed by Prof. Guillermo Rojo:
   (a) From September 11 to 12, 2006, C. Gómez (UDC) participated in a meeting with G. Pereira Lopes (New Univ. of Lisbon, Portugal) and Gaël Dias (Univ. of Beira Interior, Portugal). He gave a talk on "Compilation of parsing schemata"
   (b) This group participates with our groups in UDC and UVIGO, and also other research groups in the universities of Galicia in the thematic grid *Galician grid for language processing and information retrieval*; led by Prof. M. Vilares (UVIGO), 2006-2009.
   (c) F.M. Barcala has collaborated in the MEC project "Multilingual creation and integration of linguistic resources by terminological and discursive control strategies in specialized communication domains" (HUM2004-05658-C02-02).
8. **The Natural Language and Computational Linguistics Group, Univ. of Sussex (UK)**: Leading group in parsing and generation, directed by Prof. John Carrol. A 4-month stay of C. Gómez (UDC) has been planned to work on the development of QA strategies on TAGs.
9. **The Dept. of Computer Science and Engineering, Univ. of North Texas (USA)**: J.A. Troyano (USE) has been in touch with Rada Mihalcea, who is one of the original authors of the TEXTRANK algorithm. We have sent our main paper about an extension of this algorithm and her response has been very encouraging, and we are sure that her advice and revisions will be of great help in our research.
10. **Laboratory of Agent Modelling-LabMAg, Dept. of Computer Science, Univ. of Lisbon (Portugal)**: J. Vilares (UDC) visited this laboratory from May 29 to June 5, 2005. He gave a talk about "Dealing with Syntactic Variation in Spanish IR".
11. **The Information Retrieval Group, Dept. of Information Studies, Univ. of Sheffield (UK)**: On October 18, 2005, J. Vilares (UDC) visited it at the invitation of Profs. Mark Sanderson and Paul Clough, who have a great experience in the field of cross-lingual IR and take part in the organization of CLEF campaigns. J. Vilares gave a talk about "Parsing-based vs. locality-based approaches for European Text Retrieval".
12. **The Information Retrieval Group, Dept. of Computing Science, Univ. of Glasgow (UK)**: On November 14, 2005, J. Vilares (UDC) visited this group, invited by Prof. Keith van Rijsbergen. It is a leading group in IR, being the creators of the TERRIER system and winners of the 2005 edition of TREC. During his visit, Vilares gave a talk on the issue "From multiwords to distances: dealing with syntax in IR".

13. **The Question Answering Group, Univ. of Edinburgh (UK)**: On February 10, 2006, J. Vilares (UDC) visited this group, invited by its coordinator Prof. Bonnie Webber. This group has created the QED (Edinburgh's QA system) and the WEE/QAAM, a shallow web QA system based on information fusion techniques.
14. **The Research Group in Computational Linguistics, School of Humanities, Languages and Social Studies; Univ. of Wolverhampton (UK)**: On December 12, 2006, J. Vilares (UDC) visited this group, invited by its director Prof. Ruslan Mitkov, a leading researcher in anaphora resolution, whose other current research interests include centering, automatic abstracting and term extraction.
15. **Dept. of Translation at the Univ. of Granada**, directed by Prof. P. Faber who has visited UVIGO from March 13 to 20, 2006. He gave a talk about "Terminology oriented to process in coastal and port engineering".

3.4 Collaboration with companies and institutions

1. MISTER DOC, a Portuguese company in the domain of computer science technology applied to the development of information systems, has established a pre-protocol of collaboration with the research group of UVIGO in order to design IR applications.
2. TELÉMACO, an Spanish company in the domain of IR systems, has collaborated in the projects "Generating, extracting and structuring legal information by means of artificial intelligence techniques" (Xunta de Galicia PGIDIT05SIN044E, 2005-2006, led by F.J. Ribadas from UVIGO and J. Vilares from UDC) and "Tools for the automatic analysis and classification of documents in the legal domain" (Diputación de Ourense 2005-INOU-09, 2006, led by V.M. Darriba from UVIGO). A pre-protocol of collaboration with the groups of USE, UDC and UVIGO in order to design QA applications has been established.
3. 3.14 FINANCIAL CONTENTS, a Spanish company in the domain of accessing economic information, has collaborated in the projects "Who, what, where, when, how many, how much, how and why? NLP tools, machine learning and Bayesian networks to build question-answering robots in financial markets" (Xunta de Galicia PGIDIT05SI059E, 2005-2008, led by V.M. Darriba from UVIGO) and "Application of Artificial Intelligence for extracting cognitive and qualitative information from financial markets" (Xunta de Galicia PGIDIT02SIN01E, 2002-2005, led by M. Vilares from UVIGO).
4. INDISYS, an Spanish company in the domain of intelligent dialogue systems, has established a pre-protocol of collaboration with the research groups of USE, UDC and UVIGO in order to design QA applications.
5. LA VOZ DE GALICIA, a Spanish newspaper, has established a pre-protocol of collaboration with the research groups of USE, UDC and UVIGO in order to design QA applications.
6. ASEM, Spanish National Association for Neuromuscular Illness, has collaborated in the project "Creation and application of documentary resources about neuromuscular illness" (Xunta de Galicia PGIDIT04SIN065E, 2004-2007, led by E. Sánchez from UVIGO).
7. The Ramón Piñeiro Center for Research on Humanities (Xunta de Galicia) has been involved in the development of the project "Tagger-Lemmatizer for Current Galician" since 1996. The latest contract covered the period 2004 to 2006 with UVIGO (led by M. Vilares) and UDC (led by J. Graña).

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Parallel Architectures and Applications (Cooperative Project) TIN2004-07440-C02 (01 and 02)

Ramón Beivide[*]
Universidad de Cantabria

José Miguel-Alonso[**]
Universidad del País Vasco

Abstract

This research project deals with interconnection technologies for parallel and distributed computing systems. Nowadays, interconnection networks are ubiquitous in the computing field. At the chip level, networks are present in clustered microprocessors, on-chip multiprocessors and systems on chip. Some workstations and servers begin to use switched networks to interconnect their different modules. Clusters, cc-NUMAs, MPPs and SANs rely on their underlying interconnection subsystem for obtaining high performance and robustness. Some designs for IP switches begin to consider internal structures based on interconnection networks. Finally, Grid Computing is hot topic pursuing efficient designs for wide area networks. This project pursues new research contributions in the fields mentioned above including architectural and technological proposals and software development.

Keywords: Interconnection networks, performance evaluation, parallel and distributed computing.

1. Project Goals

The main outcomes of a cooperative research project are summarized in this report. Two coordinated research teams belonging to the University of Cantabria (subproject 1) and to the University of the Basque Country (subproject 2) support this project. The group in Cantabria is officially composed of five regular professors, one system manager and three PhD students. The group in the Basque Country is composed of three professors (and two PhD students that did not form part of the grant-requesting team).

The research project begun in September 2004 and ends in September 2007. The subproject 1 achieved a direct funding of 157800 € (40800 € in manpower) plus a grant for one PhD student. The subproject 2 achieved a direct funding of 34400€.

The goals of this research project were organized around five main research lines:

[*] Email: mon@atc.unican.es
[**] Email: j.miguel@ehu.es

1. Bibliographic study and state-of-the-art in parallel computing. As the research project deals with parallel architectures and applications, a permanent goal is to acquire in some cases, or to maintain in others, enough know-how in the research field as to propose new ideas and solutions. The emphasis of the project when it was proposed was directed towards interconnection subsystems. Then, a continuous bibliographic analysis had to be done in the broad spectrum of networks present today in different platforms: multi-core on-chip architectures, multiprocessors, clusters, supercomputers and grids.

2. Simulation Tools. The design, development and maintenance of accurate simulation tools constitute key aspects for a research group in interconnection networks. The objectives of this project included the development of simulation tools designed by our research groups and the adaptation of external ones to our simulation platforms. In the same way, the design and analysis of adequate benchmarks and load models were also targets to be considered.

3. Topological Issues. Many of the design problems of interconnection networks have a topological nature. Some of the targets programmed in this project pivoted around topological issues including the proposal of new network geometries and lay-outs. Routing, fault-tolerance, partitioning and support for collective communication were also targets of our research project. Special emphasis was declared on very large networks, as those used by the "top ten" supercomputers. In addition, contributions to the design of algebraic error-correcting codes were envisaged due to their intrinsic topological nature.

4. Architectural proposals. The proposals of complete network and router architectures constitute another goal of this project. First, the topic of fault-tolerance was considered as a critical one for forthcoming network technologies. In addition, routing and deadlock avoidance were included as research topics. The design of scalable high-performance routers was considered as another important issue. Special emphasis was also declared on small size networks for current multi-core on-chip architectures and big networks for high-end supercomputers. Finally, the architecture of network interfaces was also considered as another research goal of the project.

5. Parallel Applications. The targets around this research line were related to the parallelization of scientific codes which requires high computational power. Two application suites were considered. The first one is based on a set of bio-inspired algorithms denoted as *Estimation of Distribution Algorithms*. The second one is a physics application devoted to the search of the *Higgs boson* based on artificial neural networks. Clusters and Grid computing were declared as target platforms to run the resulting parallel codes. In addition, load monitoring and load modeling were announced as other research topics of the project.

2. Project Achievements

2.1. Simulation Tools

Supporting the advances in the other project goals, a complete evaluation framework has been developed and maintained. Currently, we are working in two main simulation platforms, SICOSYS (**Si**mulator of **Co**mmunicating **Sys**tems) and INSEE (**I**nterconnection **N**etwork **S**imulation and

Evaluation **E**nvironment), and around these tools we have a complex infrastructure that allows us to evaluate different interconnection network usage scenarios.

SICOSYS was initially planned, almost 10 years ago, trying to achieve the precision of hardware simulators with much lower computational effort. In order to be able to analyze current proposals, the simulator is under continuous development. The tool has more than 80.000 of C++ code. Its flexible software design has made possible to use a coherent tool for very different scenarios by different researchers without problems.

The main new features added to the tool support the analysis of fault-tolerant techniques in the interconnection network; in addition, the simulator is able to report power consumption measures. Besides, different architectural router proposals have been added to the simulator.

SICOSYS has been optimized to expand its scalability to current computing infrastructures. Nowadays, commercial multi-core processors have made possible to acquire affordable small scale SMP systems. Our computing infrastructure follows this trend. In order to evaluate different architectural proposals employing HTC (High Throughput Computing), large amounts of DRAM are required which can increase the infrastructure cost dramatically. To mitigate this drawback, we have parallelized SICOSYS. Now, it is possible to run efficiently one simulation with 2 to 4 threads per computing node. We have improved the answer time for multi-billion cycle simulations being able to simulate very large networks without reduce the original tool accuracy [1].

SICOSYS has been designed with a functional interface to external simulators. In the past, that interface was used with execution-driven simulators like Rsim and SimOS. Now, we have added GEMS (General Execution-driven Multiprocessor Simulator). GEMS is able to run unmodified Solaris®8/9/10 software on top of simulated state-of-the art superscalar processor with a complex memory hierarchy and using SICOSYS as the interconnection network simulator. The tool relies in SIMICS® to simulate the full system activity, including I/O. Consequently, now we have a complete framework to perform full-system simulation with precise processor/memory/network details. Using the complete tool, we are able to evaluate OS dependent workloads such as transactional or commercial applications whit a high level of confidence. Moreover, the development of new benchmarks has been greatly simplified. GEMS is an actively supported tool by the group of Prof. Mark D.Hill at the University of Wisconsin-Madison, with more than one hundred of registered users. Many multiprocessor researches use this tool in their recent papers. The knowledge and later integration of GEMS was possible through a direct collaboration with the original developers. One of the members of the U. Cantabria (Valentin Puente) spent one sabbatical year at UW-Madison collaborating with the group of Prof. Mark D.Hill.

Although not as expensive, in terms of memory and CPU utilization, as a hardware simulator, SICOSYS is still a heavyweight tool. It is not possible, using current hardware, to run this tool to simulate a network with thousands of routers. For this reason, we are working on a complementary, lightweight simulation environment called INSEE. It is less precise, in terms of timing, than SICOSYS, but allows us to experiment with very large networks.

The main two modules of INSEE are a network simulator (FSIN) and a traffic generation module (TrGen). Regarding improvements to FSIN, the most relevant one has been the extension of capabilities to deal with many different topologies, including *k*-ary *n*-cubes and trees. Important re-

design has allowed us to describe all its behavior in a single configuration file, so a single binary can simulate many different architectural proposals. Output capabilities have also been enhanced, generating files that can be imported directly by data analysis tools.

The traffic generation capabilities of TrGen have also been greatly improved. Originally, it was only able to generate traffic using a small selection of synthetic traffic patterns. In the current version, the selection is wider. Also, we have included the capability of injecting traffic to the FSIN simulator using traces obtained from actual applications. This has allowed us to perform more realistic evaluations of performance. Finally, TrGen allows FSIN to interact with a full-system simulator, Simics®, allowing simulations with workloads generated by actual, running applications.

The details of these tools are described in [2, 3, 4, 5]. The availability of these tools has allowed us to perform experiments in the fields of topologies for interconnection networks [6] and congestion control [7, 8, 9]. Also, the experimentation with different ways of injecting traffic in the network have allowed us to detect some pitfalls in some "classic" approaches; that is the reason we propose new ways of generating traffic in [10, 11].

Note that the development of the INSEE tools have been made at the UPV/EHU, but inputs for their design, and utilization, has been done in cooperation with the Group at the U. of Cantabria, and with Dr. Cruz Izu of the University of Adelaide.

Recently we have started a line of collaboration with the BSC/CNS and IBM Research Laboratories–Zurich. One of our immediate objectives is to share our knowledge on simulation tools and application characterization using traces. Later, we will work on topologies and scheduling strategies for large-scale high-performance supercomputers. We have held several meetings hosted by the BSC/CNS, and one of the members of the UPV/EHU team (Javier Navaridas, a PhD student under the supervision of Dr. Miguel-Alonso) has been visiting the Zurich laboratories from November 2006 to February 2007.

Concerning parallel workloads, we are actively using NAS Parallel Benchmarks (OpenMP or MPI versions), older benchmarks such us SPLASH 1&2 and some transactional benchmarks such as static web applications (SURGE) or Online Transaction Processing (SPECjbb)

Regarding traffic modeling, we have been working in different ways of providing realistic traffic to our network simulators. This has been done using three different approaches: i) *Full-system simulation*: A collection of machines are simulated using Simics®, and run unmodified MPI applications. The network simulator performs the interchange of messages; ii) *Trace-based simulation*: Instrumented versions of MPI applications are run in actual machines (our own clusters, or machines such as BSC/CNS Mare Nostrum). Detailed traces of the interchange messages are obtained. These traces are used to feed simulators. The challenge here is to respect causal relationships; iii) *Synthetic traffic*: The usual way of working is considering a collection of independent sources of traffic that use some probability distribution functions to compute message sizes, destinations, and inter-generation times. We have introduced a "burst" mode in which traffic sources are not totally independent, which reflect the coupled nature of most parallel applications. A recent paper submitted to EuroPar 2007 [12] compares the last two options. Paper [2] evaluates the first one.

2.2. Topological Issues

The objectives achieved within this workpackage can be grouped around two topics: distance-reduced practicable topologies and algebraic codes.

In respect to topologies, dozens of parallel computers of different size have been designed around Torus interconnection networks. Typically, a 2D Torus arranges its N nodes in a square Mesh with wraparound links. Above few thousands nodes, it has been shown that parallel computers should use 3D topologies, being a cubic 3D Torus the most desirable solution. Parallel machines, such as the HP GS1280 based on Alpha 21364 and the Cray X1E vector computer, have used two-dimensional Tori. Others, such as the Cray T3D and T3E have used three dimensional Tori by piling up several 2D Tori planes. The IBM BlueGene is a notable example of a massively parallel computer that joins $2^6 \times 2^5 \times 2^5$ nodes in a mixed-radix 3D Torus. In addition, each dimension of a Torus network may have a different number of nodes, leading to rectangular and prismatic topologies for two and three dimensions respectively. These topologies are denoted as mixed-radix networks. Mixed-radix Tori are often built for practical reasons of packaging and modularity. The HP GS1280 employs a 2D rectangular network and the IBM Blue-Gene a 3D prismatic one. However, mixed-radix Tori have two important drawbacks: first, they are no longer edge-symmetric and second, the distance-related network parameters (diameter and average distance) are quite far from the optimum values of square and cubic topologies. The edge asymmetry introduces load imbalance in these networks, and for many traffic patterns the load on the longer dimensions is larger than the load on the shorter ones. In addition, maximum and average delays are too long as they depend on the poor values of diameter and average distance exhibited by these networks.

During this project we have analyzed and introduced a family of 2D toroidal networks which includes standard and twisted Tori configured as square or rectangular networks. We have proposed the use of the Gaussian integers as the appropriated tool for defining, analyzing and exploiting this set of networks. By using this algebraic tool we have solved some interesting applications over this type of networks such as unicast and broadcast optimal routing and the perfect placement of resources on them. Some of the results of this research line have been published in several papers and others have been submitted for publication. Most of the material generated on this topic has been collected in [38]. In [13], we have considered practicable lay-outs for this kind of networks. Also, in [14] we have explored a particular network case as a suitable topology for on-chip networks. In [47, 48] the routing problem over circulant graphs has been revisited.

Moreover, we have proposed and analyze alternative mixed-radix 2D and 3D Torus topologies that avoid the two above mentioned problems by adequately twisting the wraparound links of one or two network dimensions. The performance exhibited by these twisted Tori is notably higher than the one obtained when using their standard mixed-radix counterparts. Some of the results of this research line have been published in a recently accepted paper [6].

In respect to the field of Algebraic Codes, the design of error-correcting codes for two-dimensional signal spaces has been recently considered in the technical literature. Hamming and Lee distances have been proved to be inappropriate metrics to deal with QAM signal sets and other related constellations. Up to our knowledge, the first author who modeled certain QAM constellations

with quotient rings of Gaussian integers was Klaus Huber. In his papers, Huber introduced a new distance for using in QAM-like constellations denoted as Mannheim metric. The rational behind this metric is to consider the Manhattan (or taxicab) metric modulo a two-dimensional grid.

Based on this seminal work, we have proposed perfect codes for different multidimensional signal spaces. To solve these problems, we have introduced an original relationship among the fields of Graph Theory, Number Theory and Coding Theory. One of our main findings is the proposal of a suitable metric over quadratic, hexagonal and four-dimensional constellations of signal points. This metric is the distance among vertices of a new class of Cayley graphs defined over integer rings which include Gaussian, Eisenstein-Jacobi and Lipschitz Graphs. Hence, such graphs represent mathematical models of the multidimensional constellations under study. A problem in Graph Theory known as the perfect dominating set calculation has been solved over these graphs. A sufficient condition for obtaining such a set has been given for each case. Obtaining these sets of domination directly yields to the construction of perfect codes for the alphabets under consideration.

Papers concerning codes over Gaussian integers have been published in notable Information Theory conferences' proceedings or are in reviewing process in journals. Specifically, in [39] and [40] perfect codes over quotient rings of Gaussian integers are considered. The metric applied to these codes is the distance induced by a circulant Gaussian graph. Also, the weight distribution of these circulants has been presented in [41]. In addition, perfect codes over any quotient ring of Gaussian integers are presented in [42]. In this paper we have shown, as well, that Lee perfect codes are a particular case of our Gaussian perfect codes. In [39], 1-perfect error correcting codes over hexagonal constellations modeled by quotient rings of Eisenstein-Jacobi integers have been introduced. Recently, in [43], a method for finding certain perfect t-dominating sets over degree six circulant graphs has been considered. A preliminary approach to degree eight graphs and codes over them is being considered. In this case, the graphs will be built over certain subsets of the integer quaternions. Perfect 1-dominating sets are obtained and perfect codes over these sets are compared, in some cases, with perfect Lee codes, which we have considered in [44]. Most of this work corresponds to the PhD dissertation of C. Martínez [50].

2.3. Architectural Proposals

Among the different tasks that were raised in this objective, most effort has been directed towards the improvement of packet routers, and to the increase of the fault-tolerant capabilities of the network. Thus, very significant results were obtained in [15] (not completely explored yet) giving rise to a new strategy to implement adaptive routing in irregular networks. This method relies on the use of a deadlock avoidance mechanism developed by our groups in previous projects.

Other aspect in which our work has given results of impact has been the development of a new fault-tolerant mechanism for medium and large size k-ary n-cube networks [16, 17]. This mechanism makes possible to tolerate, at network level, every time-space combination of failures provided that the network topology remains connected. The application of this new mechanism to other network structures can give new interesting results. In [45], we present an exhaustive evaluation of the proposal under different realistic scenarios.

Many other issues, corresponding to the hardware support for improving the communication operations inside the router have been carried out. Thus, new solutions for solving several congestion problems, included in [18, 19, 20], have been proposed, with special emphasis on the aspect discovered (at least published for the first time) for our groups, about the instability of channel occupation in networks whose diameter surpasses the twenty of nodes [21]. Additionally and with respect to the network interfaces, a deep analysis of the main characteristics of the network interfaces present at the moment in the market (and with options of future) has been made in collaboration with the BSC/CNS.

It is worth to remark that an important part of the challenges of the interconnection networks has been transferred inside the chip. The capacity of integrating several processors inside the same chip has converted the interconnection network in a key element for the performance of new multi-core architectures. For this reason, our groups are making an effort in the acquisition of the necessary expertise and tools as to make contributions in interconnection mechanisms suitable for this new environment. As a consequence of this effort, a line of collaboration with the BSC/CNS has produced significant results [22, 23] about kilo-instruction processors as a way for removing bottlenecks in the access to shared resources. Specifically, a new router structure for this type of systems has been developed whose performance surpasses the one of the architectures used until the moment. These results have been submitted for publication [46].

Finally, we want to mention that, in cooperation with Dr. Izu of the U. of Adelaide, we have studied some congestion-control mechanism that can be incorporated in large-scale interconnection networks. In particular, we have studied the mechanism included in IBM's BlueGene/L (that gives priority to in-transit traffic, even at the cost of delaying new injections) and proposed and studied LBR (Local Buffer Restriction), an extension of the Bubble routing mechanism to adaptive virtual channels. Results of these studies are summarized in [7, 8, 9].

2.4. Parallel Applications

One of the objectives of this project was to perform, in cooperation with the Intelligent Systems Group of the UPV/EHU, the parallelization of a particular class of bio-inspired algorithm called EDAs (Estimation of Distribution Algorithms). These algorithms have proven very successful when solving complex optimization problems, but are very CPU-consuming. If we were able to accelerate them, via parallelization, we could reach better solutions in shorter times, and/or explore more solutions. This is precisely what we have done. The details of the way EDAs work, and how parallelization has been done, are in [24]. This paper also includes a preliminary performance evaluation of the parallel solutions, extended in [25].

We have collaborated with the group of Prof. Ubide at the Faculty of Chemistry of the UPV/EHU, in order to apply our fast, parallel version of EDAs to a complex problem in quantitative chemistry: the creation of multivariate calibration models for chemical applications. The problem consists of obtaining data (light spectra) of the concentration of species taking part in controlled chemical reactions, selecting the most relevant data, and creating a prediction model. This model can be feed with measured data from new reactions with unknown concentrations of the same species; the model should predict those concentrations. The application of EDAs has

proven to be a very successful approach, because of the quality of the obtained models. In [26, 27, 28] we discuss the approaches to the problem used traditionally by chemists, and compare those with new approaches based on the "intelligent" search in the space of solutions performed by an EDA. In [29] we extend this work evaluating different mechanisms of input-data reduction, and different EDAs.

As a result of this work, Alexander Mendiburu presented his PhD dissertation in January 2006 [30] under the supervision of Dr. Jose Miguel-Alonso and Dr. Jose Antonio Lozano.

In the field of grid computing, we have been working in economic models to schedule applications in a grid environment. The results of this work are in [31]. What we have done is an extension to the grid architecture, introducing a collection of modules that, when launching an application, select resources taking into account not only availability and adequacy, but also a cost model. Implementation has been done using a Globus 4 environment [www.globus.org/toolkit], and the GridWay [www.gridway.org] meta-scheduler.

Another objective of this project was to perform, in cooperation with the IFCA (Instituto de Física de Cantabria - Institute of Physics of Cantabria), the parallelization of an Artificial Neural Nets algorithm. Currently, to search for the Higgs boson at CERN (European Nuclear Research Centre) a multi-layer perceptron (MLP) is used. MLPfit is a sequential application for the design, implementation and use of multilayer perceptrons. It has been developed by DAPNIA which depends on the French Commission of Atomic Energy and adopted by CERN for the implementation of artificial neural nets for High-Energy Physics. The details of how parallelization has been carried out are in [32]. This work also has produced several Master theses [33, 34, 35, 36]. One of these students has joined our group as a PhD student. In addition, a formal collaboration with IFCA has been established under a new contract [37] for participating in the Crossgrid initiative. Crossgrid is a European project, whose objective is the creation, management and exploitation of a Europe-wide Grid computation environment, permitting the interactive use of applications that are extremely intensive in terms of calculation and data. The applications making use of the Crossgrid project's infrastructure are related with the fields of biomedicine, meteorology or High-Energy Physics. Also, in [49] we have presented an evaluation of OpenMoxis environments.

3. Indicators of results

Collaboration with research groups

- Dr. Cruz Izu – Dep. of Computer Science, U. of Adelaide, Australia.
- Prof. Ernst Gabidulin. Moscow Institute of Physics and Technology. Russia.
- Prof. Mark D. Hill, UW-Madison. USA.
- Prof. Mateo Valero and Jesús Labarta. BSC/UPC.
- Dr. Carlos Ubide – Faculty of Chemistry, UPV/EHU.
- Intelligent Systems Group – Computer Science Sc., UPV/EHU (www.sc.ehu.es/isg)
- BSC / CNS (www.bsc.es)
- IBM Research Laboratories–Zurich (www.zurich.ibm.com)
- I2Bask – Research Network of the Basque Country (www.i2bask.net)
- IFCA (CSIC/U. Cantabria, (www.ifca.unican.es)

Patents: "Mecanismo de encaminamiento tolerante a fallos altamente escalable". Inventores: Valentín Puente Varona, José Ángel Gregorio Manasterio, Fernando Vallejo Alonso, Ramón Beivide Palacio. U. Cantabria. Request #: P200500530. Request date: 01-03-05.

PhD dissertations:

- Dr. Alex Mendiburu (supervised by Dr. Miguel-Alonso) UPV/EHU
- Dr. Carmen Martínez (supervised by Dr. Ramón Beivide) U. Cantabria

Current PhD students: 2 at the UPV/EHU, 3 at U. of Cantabria, 1 at U. of Burgos

Master's thesis: 1 at the UPV/EHU, 4 at U. of Cantabria

Participation in other projects: HiPEAC Network of Excellence (European Project IST-004408)

Articles in international journals: 7 accepted plus 2 submitted
Papers in international conferences: 21 accepted and 3 submitted
Papers in national conferences: 7 accepted
Other publications (technical reports): 3

4. References

1. P.Prieto, V.Puente, P.Abad y J.Á. Gregorio, "Paralelización de un Entorno de Evaluación de Redes de Interconexión", XVII Jornadas de Paralelismo, 2006.
2. Javier Navaridas, Fco. Javier Ridruejo, J. Miguel-Alonso. "Evaluation of Interconnection Networks Using Full-System Simulators: Lessons Learned". Proc. 40th Annual Simulation Symposium, Norfolk, VA, March 26-28, 2007.
3. F.J. Ridruejo, J. Miguel. "Simulación de redes de interconexión utilizando tráfico real". Actas de las XVI Jornadas de Paralelismo. Thomson, 2005 (ISBN 84-9732-430-7). Pages: 109 - 116.
4. F.J. Ridruejo, J. Miguel-Alonso. "INSEE: an Interconnection Network Simulation and Evaluation Environment". Lecture Notes in Computer Science, Volume 3648 / 2005 (Proc. Euro-Par 2005), Pages 1014 - 1023.
5. F.J. Ridruejo, A. Gonzalez, J. Miguel-Alonso. "TrGen: a Traffic Generation System for Interconnection Network Simulators". International Conference on Parallel Processing, 2005. 1st. Int. Workshop on Performance Evaluation of Networks for Parallel, Cluster and Grid Computing Systems (PEN-PCGCS'05). ICPP 2005 Workshops. 14-17 June 2005 Page(s): 547 – 553
6. Jose M. Camara, Miquel Moreto, Enrique Vallejo, Ramon Beivide, Jose Miguel-Alonso, Carmen Martinez, Javier Navaridas. "Mixed-radix Twisted Torus Interconnection Networks". Proc. 21st IEEE International Parallel & Distributed Processing Symposium - IPDPS '07, Long Beach, CA, March 26-30, 2007.
7. F.J. Ridruejo, J. Navaridas, J. Miguel-Alonso, C. Izu. "Evaluation of Congestion Control Mechanisms using Realistic Loads". Actas de las XVII Jornadas de Paralelismo. Albacete, Septiembre 2006.

8. C. Izu, J. Miguel-Alonso, J.A. Gregorio. "Effects of injection pressure on network throughput". Proc. PDP'06 - 14th Euromicro Conference on Parallel, Distributed and Network-based Processing. Montbéliard-Sochaux, 15-17 Feb. 2006, France. Pp. 91-98
9. J. Miguel-Alonso, C. Izu, J.A. Gregorio. "Improving the Performance of Large Interconnection Networks using Congestion-Control Mechanisms". Technical report EHU-KAT-IK-06-05. Department of Computer Architecture and Technology, The University of the Basque Country.
10. C. Izu, J. Miguel-Alonso, J.A. Gregorio. "Evaluation of Interconnection Network Performance under Heavy Non-uniform Loads". Lecture Notes in Computer Science, Volume 3719 / 2005 (Proc. ICA3PP 2005), Pages 396 - 405.
11. J. Miguel-Alonso, F.J. Ridruejo, C. Izu, J.A. Gregorio and V. Puente. "Are independent traffic sources suitable for the evaluation of interconnection networks?". Technical report EHU-KAT-IK-07-05. Department of Computer Architecture and Technology, The University of the Basque Country.
12. F.J. Ridruejo, J. Navaridas, J. Miguel-Alonso and Cruz Izu. "Realistic Evaluation of Interconnection Network Performance at High Loads". Submitted to EuroPar 2007.
13. E. Vallejo, R. Beivide and C. Martinez, «Practicable Lay-outs for Optimal Circulant Graphs» The 13th Euromicro Conference on Parallel, Distributed and Network-Based Processing (PDP'05)
14. C. Martínez, E. Vallejo, R. Beivide, C. Izu and M. Moretó, "Dense Gaussian Networks: Suitable Topologies for On-Chip Multiprocessors", International Journal of Parallel Programming, Vol: 34, pp. 193-21, Junio 2006.
15. V. Puente, J.A. Gregorio, F. Vallejo, R. Beivide and C. Izu, "High-performance adaptive routing for networks with arbitrary topology", Journal on System Architecture, Vol 52, Issue 6, June 2006, pp.345-358
16. V.Puente, J.A.Gregorio, F.Vallejo y R.Beivide, "Immunet: A Cheap and Robust Fault-Tolerant Packet Routing Mechanism", International Symposium on Computer Architecture (ISCA-2004).
17. V. Puente, J.A. Gregorio, "Immucube: Scalable Fault Tolerant Routing for k-ary n-cube Networks", Aceptado para publicación en IEEE Transactions on Parallel and Distributed Systems.
18. C.Izu, R.Beivide y J.A. Gregorio, "The Case of Chaotic Routing Revisited", WDDD Workshop. In conjunction with the International Symposium on Computer Architecture (ISCA-2004)
19. C.Izu y R.Beivide, "Understanding Buffer Management for Cut-through 1D Rings", EURO-PAR 2004.
20. Miguel-Alonso, J.A. Gregorio, V. Puente, F. Vallejo, R. Beivide, P. Abad, « Desequilibrio de Carga en Redes k-ary n-cube », XV Jornadas de Paralelismo (2004)
21. J.Miguel-Alonso, J.A.Gregorio, V.Puente, F.Vallejo y Ramón Beivide, "Load unbalance in k-ary n-cube Networks", EURO-PAR 2004.
22. E.Vallejo, M. Galluzzi, A. Cristal, F. Vallejo, R. Beivide, P. Stenström, J. E. Smith and M. Valero, "Implementing Kilo-Instruction Multiprocessors", IEEE International Conference on Pervasive Services 2005 (ICPS'05); Santorini, Greece.
23. E.Vallejo, F.Vallejo, R.Beivide, M.Galluzzi, A.Cristal, P.Stenström, J.E. Smith, M.Valero, "KIMP: Multicheckpointing Multiprocessors", XVII Jornadas de Paralelismo, 2005.

24. A. Mendiburu, J.A. Lozano, J. Miguel-Alonso. "Parallel implementation of EDAs based on probabilistic graphical models". IEEE Transactions on Evolutionary Computation, Aug. 2005 Vol. 9, N. 4, Pages 406 - 423.
25. A. Mendiburu, J. Miguel-Alonso and J.A. Lozano. "Implementation and performance evaluation of a parallelization of Estimation of Bayesian Network Algorithms". Parallel Processing Letters, Vol. 16, No. 1 (2006) 133-148.
26. A. Mendiburu, J. Miguel-Alonso, J.A. Lozano, "Evaluation of Parallel EDAs to Create Chemical Calibration Model", e-science, p. 118, Second IEEE International Conference on e-Science and Grid Computing (e-Science'06), 2006.
27. A. Mendiburu, J. Miguel-Alonso, J.A. Lozano, M. Ostra and C. Ubide, "Parallel EDAs to create multivariate calibration models for quantitative chemical applications", Journal of Parallel and Distributed Computing, Volume 66, Issue 8, Parallel Bioinspired Algorithms, August 2006, Pages 1002-1013.
28. A. Mendiburu, J. Miguel-Alonso, J. A. Lozano, M. Ostra and C. Ubide. "Calibración multivariante de reacciones químicas utilizando EDAs paralelos". Actas del IV Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados (MAEB'2005) II. M.G. Arenas et al. (editors). Thomson, 2005 (ISBN 84-9732-467-6). Pages: 881 - 888.
29. A. Mendiburu, J. Miguel-Alonso, J. A. Lozano, M. Ostra and C. Ubide. "Parallel and Multi-objective EDAs to create multivariate calibration models for quantitative chemical applications". International Conference on Parallel Processing, 2005. 1st. Int. Workshop on Parallel Bioinspired Algorithms. ICPP 2005 Workshops. 14-17 June 2005 Page(s): 596-603
30. Alexander Mendiburu Alberro. "Parallel implementation of Estimation of Distribution Algorithms based on probabilistic graphical models. Application to chemical calibration models". PhD thesis, Department of Computer Architecture and Technology, UPV/EHU, January 2006.
31. Jose Antonio Pascual Saiz. Sistema de compraventa de recursos para Grid Computing. Proyecto Fin de Carrera dirigido por J. Miguel-Alonso. Facultad de Informática UPV/EHU, 2006.
32. Gutierrez, Alejandro, Cavero, Francisco, Menéndez de Llano, Rafael. Parallelization of a Neural Net Training Program in a Grid Environment. Proceedings of PDP 2004 by IEEE Computer Society. 2004.
33. Pablo Abad. Análisis Y Evaluación De Sistemas De Colas En Entornos De Computación De Alto Rendimiento. Proyecto Fin de Carrera dirigido por Rafael Menéndez de Llano. Universidad de Cantabria. 2004.
34. Alejandro Gutierrez. Implementación Y Evaluación De La Paralelización De Una Aplicación De Física De Altas Energías Basada En Redes Neuronales Artificiales Para La Detección Del Bosón De Higgs. Proyecto Fin de Carrera dirigido por Rafael Menéndez de Llano. Universidad de Cantabria. 2004.
35. David Hernández Sanz. Instalación, Análisis Y Evaluación De Un Cluster Openmosix Para Entornos De Alta Productividad Y De Alto Rendimiento. Proyecto Fin de Carrera dirigido por Rafael Menéndez de Llano. Universidad de Cantabria. 2005
36. Víctor Sancibrián Grijuela. Evaluación De Distintos Bancos De Pruebas Para Diferentes Arquitecturas De Supercomputadores. Proyecto Fin de Carrera dirigido por Rafael Menéndez de Llano. Universidad de Cantabria. 2005.

37. "Development of Grid Environment for Interactive Applications (CROSSGRID)" Subcontrato de IST-2001-32243. CSIC / Universidad de Cantabria. Investigador principal: Rafael Menéndez de Llano Rozas. 2004.
38. C. Martínez, R. Beivide, E. Stafford, M. Moretó and E. Gabidulin. ``Modeling Toroidal Networks with the Gaussian Integers". Submitted to IEEE Transactions on Computers. 2007.
39. C. Martínez, R. Beivide, and E. Gabidulin. "Perfect Codes from Circulant Graphs". Accepted for publication under changes. IEEE Transactions on Information Theory. 2007.
40. C.Martínez, R.Beivide, J. Gutierrez and E.Gabidulin, "On the Perfect t-Dominating Set Problem in Circulant Graphs and Codes over Gaussian Integers", 2005 IEEE International Symposium on Information Theory (ISIT 2005); Adelaide, Australia.
41. E.Gabidulin, C.Martínez, R.Beivide and J.Gutierrez, "On the Weight Distribution of Gaussian Graphs with an Application to Coding Theory", Eighth International Symposium on Communication Theory and Applications (ISCTA '05); Ambleside, Lake District, UK.
42. C. Martínez, M. Moretó, R. Beivide and E. Gabidulin, "A Generalization of Perfect Lee Codes over Gaussian Integers",. 2006 IEEE International Symposium on Information Theory (ISIT'06).
43. C. Martínez, E. Stafford, R. Beivide and E. Gabidulin, "Perfect Codes over Eisenstein-Jacobi Graphs", Lugar de publicación: Proc. of Tenth Int. Workshop on Algebraic and Combinatorial Coding Theory (ACCT-10), 2006.
44. C. Martínez, E. Stafford, R. Beivide and E. Gabidulin. "Perfect Codes over Lipschitz Integers". Submitted to 2007 IEEE International Symposium on Information Theory.
45. V.Puente, J.A.Gregorio, F.Vallejo y R.Beivide, "Dependable Routing in Interconnection Networks of Flexible Topology", Submitted to IEEE Transactions on Computers. 2007.
46. P. Abad, V. Puente, P. Prieto and J.A. Gregorio, "Rotary Router: An Efficient Architecture for CMP Interconnection Networks", Submitted to ISCA 2007, San Diego, California, USA, 2007.
47. D. Gomez-Perez, J. Gutierrez, A. Ibeas, C. Martinez, « On routing on circulant graphs of degree four », International Symposium on Symbolic and Algebraic Computation 2005.
48. D.Gómez, J.Gutierrez, A.Ibeas, C.Martínez and R.Beivide, « On Finding a Shortest Path in Circulant Graphs with Two Jumps", The Eleventh International Computing and Combinatorics Conference (COCOON 2005); Kunming, Yunnan, P.R.China.
49. D.Hernández, R.Menéndez de Llano, P.Abad, C.Izu. "Evaluación de un cluster OpenMosix para entornos de alta productividad y de alto rendimiento", XVI Jornadas de Paralelismo, 2005.
50. C. Martínez. "Codes and Graphs over Complex Integer Rings". PhD Dissertation. To be defended the 12th of March, 2007.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Improvement and Application of the Game Technology in Virtual Reality and web contents TIN2004-07451-C03

Mateu Sbert*	Miguel Chover**	Francisco Ibáñez
Coordinator Main Researcher UdG subproject	Main Researcher UJI subproject	Main Researcher AIJU subproject

Abstract

Cutting-edge technology in the field of Computer Graphics is currently found in 3D engines for computer games. This is due to the successful combination of research conducted by universities with the traditional techniques used in the digital games sector. While this technology was first developed for computer games, it is expanding its scope of application to include other sectors which are less developed as far as realistic rendering is concerned. Virtual Reality and 3D web content are the sectors which show the greatest interest in incorporating the new technologies. Therefore the purpose of this project is twofold: to continue the research on computer games technologies carried out in the project TIC2001-2416 and to apply the new technologies to the fields of Virtual Reality and 3D web content.
Keywords: computer games, virtual reality, web 3D

1 Project Objectives

Global objectives of the coordinated project:

1. Improvement on game technology in the areas of illumination and geometry

2. Application of game technology to virtual reality and web content

3. To transfer the scientific results to industrial sectors

The research that will be done by the teams at UdG and UJI is complementary, tackling both the two main aspects in virtual reality and web3D, the illumination quality and the possibility of inclusion of complex models. The team at AIJU will state the requirements, will validate the results and will disseminate them in the industrial environment, and will also lead the exploitation of

* Email: mateu@ima.udg.es
** Email: chover@lsi.uji.es
*** Email: fibanyez@aiju.info

results. We believe that the coordination of the three groups is crucial and gives a very important added value, especially in technology transfer.

Specific objectives of UdG subproject

Subproject MATER-UdG deals with improvements in the area of illumination. To this means we established the following objectives:

UdG1- Improvement of obscurances theory to obtain real-time in its recalculation and its application to complex environments including vegetation.

UdG2- Development of two-phase algorithms to solve the problem of global illumination in non-diffuse environments. The first phase computes and stores the lighting. The second one is the execution in real-time of a module that generates the illumination from the stored parameters.

UdG3- Incorporation of image based rendering techniques and combination of solutions with real images.

UdG4- Acceleration of ray-tracing, i.e. using path coherence, and its inclusion in a game engine for the computation of illumination.

UdG5- Use of new programmable graphics cards to obtain the acceleration of the computation of illumination, with new algorithms or by adapting existing ones.

UdG6- Study and introduction of systematic sampling in the different illumination computations.

UdG7- Application to virtual reality and Web3D systems of the developed technology.

Specific objectives of UJI subproject

The subproject MATER - UJI consists in the improvement of computer games technology in the geometry field. These improvements are applied in the context of virtual reality and Web 3D. In particular the subproject includes the following objectives:

UJI 1- Definition of a new framework for the comparison of multiresolution models.

UJI2- Development of new Level of Detail (LOD) models that exploit connectivity.

UJI3- Development of new LOD models for specific objects (trees, leaves ...).

UJI4- Use of Graphics Hardware to speed up geometry rendering.

UJI5- Application to virtual reality and Web3D systems of the developed technology.

Specific objectives of AIJU subproject

The main aims of subproject MATER-AIJU is to establish the requirements and to validate the research results for the ulterior transfer to the industry. To achieve this purpose we have defined the objectives listed below:

AIJU1- To establish the needs and requirements of the toys and games industries in order to conduct the research work towards applicable results in the industry.

AIJU2- To validate the research results from suprojects MATER-UdG and MATER-UJI in Virtual Reality and 3D Web.

AIJU3- To Transfer the obtained results to industry.

Coordinate Objectives

Research between UdG and UJI:

C1- Application of illumination techniques for our Level of Detail models.

C2- Simplification, compression and progressive transmission of geometry for Web 3D applications.

Research between UdG, UJI and AIJU:

C3- Development of an application for Web3D.

C4- Development of an application for Virtual Reality.

2 Level of success of the project

We have advanced in the three global objectives:

1. Improvement on game technology in the areas of illumination and geometry

2. Application of game technology to virtual reality and web content

3. To transfer the scientific results to industrial sectors

Next we describe the advancement in the specific objectives of the three subprojects:

Specific objectives of UdG subproject

We have advanced in all the objectives. Next we review them one by one:

UdG1- Improvement of obscurances theory to obtain real-time in its recalculation and its application to complex environments including vegetation.

We have applied obscurances to vegetation, for real time rendering in [Garcia05A] and to ray-tracing in [Mendez06B]. In [Mendez06D, Szecsi06] a new obscurances computation using depth peeling is presented, and real-time recalculation is considered, although it is still in a first stage. In [Feixas06A] a viewpoint based technique is used to obtain obscurances-like illumination that is also applied to vegetation.

UdG2- Development of two-phase algorithms to solve the problem of global illumination in non-diffuse environments. First phase computes and stores the lighting. Second one is the execution in real-time of a module that generates the illumination from the stored parameters.

In [Szirmay05A, Szecsi06, Szirmay06] two phase algorithms are described that allow real time rendering of global illuminated scenes and participating media.

UdG3- Incorporation of image based rendering techniques and combination of solutions with real images.

In [Garcia05A, Garcia05B], image based rendering techniques are used to render trees in real time, and also photographs of leaves are used via indirect texturing to enhance realism.

UdG4- Acceleration of ray-tracing, i.e. using path coherence, and its inclusion in a game engine for the computation of illumination.

In [Mendez06C] the unbiased reuse of camera paths is obtained and in [Mendez06A] a new paradigm, the frame array, is defined by reusing both paths from camera and light source. In [Szecsi06, Szirmay06] precomputed light paths that make extensive use of coherence are used to compute global illumination in real-time. In [Szirmay05A, Szirmay06] paths are obtained for participating media using Illumination Networks, which allow real time rendering. In [Szirmay05B] the "Go with Winners" strategy defined first in Physics is applied to Path-Tracing. Finally, in [Flo06B, Florez06A] interval theory is used to accelerate the raytracing of implicit surfaces.

UdG5- Use of new programmable graphics cards to obtain the acceleration of the computation of illumination, with new algorithms or by adapting existing ones.

In [Mendez06D, Szecsi06] a new obscurances computation using depth peeling with graphics card is presented. In [Szirmay05A, Szirmay06] Illumination Networks are stored in textures and this allows real time rendering using the GPU. In [Szecsi06, Szirmay06] the global illumination is computed in real-time using also the graphics programmable unit. New and efficient algorithms that make use of graphics hardware to compute the obscurances are presented.

UdG6- Study and introduction of systematic sampling in the different illumination computations.

In [Sbert06] the systematic sampling is applied to path-tracing and also demonstrated for obscurances.

UdG7- Application to virtual reality and Web3D systems of the developed technology.

The new techniques developed have been or are in process of being integrated into several game engines, Ogre, Shark, which are at the core of virtual reality engines. They have been also developed as either stand alone techniques or plugins for Maya or 3D studio, which allows their easy integration in the visualization pipeline. See below at section 2, technology transfer, for the industrial application of our results.

Specific objectives of UJI subproject

UJI 1- Definition of a new framework for the comparison of multiresolution models.

Related with the first goal, we have developed a tool to compare multiresolution models. This application loads any model defined in a dynamic library with a standard specification. The tool analyzes the extraction and visualization times of each model, to validate its performance in a real application and gives general information (number of triangles, strips, vertices send to the GPU, etc .....) [Castello05B].

UJI2- Development of new Level of Detail models that exploits connectivity.

The second objective is related with the creation of multiresolution models. We have worked in different techniques related with the construction of LOD models [Ripolles05A, Ripolles05B] including simplification algorithms [Castello05A, Castello06, Castello07, Gonzalez07]. We have developed four variants of a model that has the following characteristics [Belmonte04, Belmonte06, Ramos06A, Ramos06B]: it is a continuous representation, based on the use of triangle strips, which exploits the graphics hardware, which handles objects formed by sub meshes, which considers its textures and attributes and which can be animated by skeletons. Each one of these variants is an evolution of the previous one including new properties like: lower storage cost, higher rendering speed, progressive transmission or variable resolution. Some of these variants are integrated in OGRE, a game engine that can be used to develop Virtual Reality systems.

UJI3- Development of new Level of Detail models for specific objects (trees, leaves ...).

The third objective consists of the definition of new multiresolution models for objects with special features (as for example trees and plants). We have worked in the development of two models for the representation of leaves. The first of them [Rebollo06A, Rebollo07B] maintains a representation of greater quality for the different levels of detail, whereas the second one [Rebollo07A] is designed specially to exploit graphics hardware and allows the instantiation of different species for rendering forest scenes. The two variants are integrated in the OGRE and SANDRA engines respectively.

UJI4- Use of Graphic Hardware to speed up the rendering of geometry.

The fourth objective is based on the exploitation of the graphical hardware; we have proposed some improvements on actual architectures [Gumbau06A]. With respect to this objective all of the models developed in objectives UJI 2 and UJI 3 have been designed specially for the graphics hardware [Ramos06B, Rebollo06B, Rebollo07A, Rebollo07B]. As an extension over the initial objectives, we have developed a library for the management of scenes with a great number of multiresolution objects [Gumbau07]. This library is able to adapt the level of detail of each object, taking into account the capability of the graphics hardware and the application requirements. Finally we have developed a method for normal map generation on the GPU, to enhance our multiresolution models [Gonzalez05, Gumbau06B].

UJI5- Application to virtual reality and Web3D systems of the developed technology.

The last one of the objectives looks for the use of the developed technology in Virtual Reality systems. In order to facilitate the integration of the proposed models, we have created a tool that automates the generation of multiresolution models. From a simple interface it is possible to load any object and to construct its multiresolution representation, as much for general meshes as for trees and plants.

We have worked in the following areas, related with the objectives addressed to the collaboration between subprojects

- With UdG. We have worked to develop geometric simplification techniques that are used for the construction of the multiresolution models [Castello07]. During this year, we will apply the techniques of illumination developed by the UdG to our models (this task has been postponed 6 months in its execution, due to agenda problems).

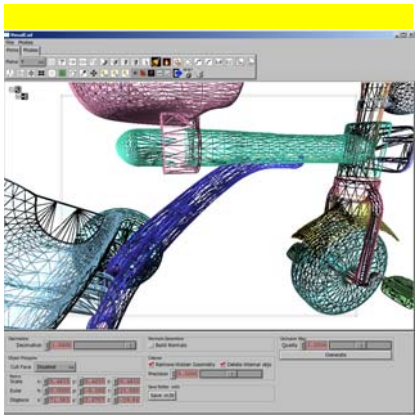
- With AIJU. We have participated in the system specification for the validation of the developed technology. At this moment, we work in the integration of our models in two applications: the first one for the exploration of 3D products in the Web and the other one for natural scene rendering with a Cave. These applications will be finished during this year.

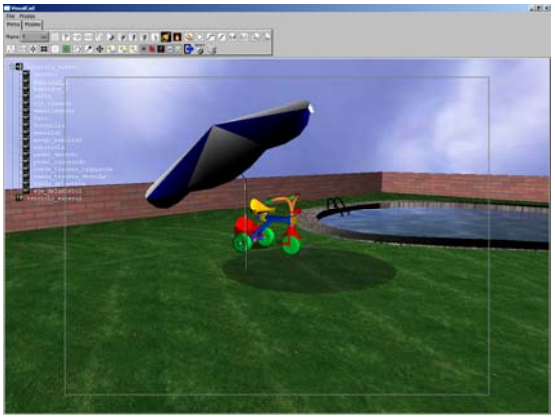
Specific objectives of AIJU subproject

Regarding the objective AIJU1, a questionnaire has been designed and distributed among a sample of European videogames companies. The results have been analysed and according to the requirements, the research has been conducted by the Universities – UdG and UJI.

The libraries developed by UdG and UJI have been integrated in a specific real-time graphic software (e-Studio from Brainstorm multimedia). After a very hard work of integration, toys designs have been used for the validation of the libraries integrated specifically geometry and illumination. So the objective AIJU2 (Validation of research results) and AIJU3 (Transfer of results to industry) have been fulfilled. The toy designs have been displayed in virtual reality environments

and in 3D web, checking the enhancements introduced by the integration of the geometry and illumination libraries developed in the project.





3 Results indicators

Subproject UdG:

Personnel in formation.

At the University of Girona, Jaume Rigau read his dissertation in December 2006 and Alex Méndez's dissertation is already approved and will be read in April 2007. Nico Sunyer, contracted by the project, has started a PhD on the use of new hardware facilities, like the cell computer or new graphics card, for global illumination. Francisco González and Ismael García are doing this year their Master studies and will start a PhD next year. Their Master thesis subject deals with tree rendering and viewpoint selection, respectively. All of them have participated in research and publications related with the project. Sergi Funtané is collaborating in technology transfer and has also participated in one publication.

Results can be found at the web page of the project http://ima.udg.es/iiia/GGG/TIN2004-07451-C03-01/galeria.html

Publications related with the project

In the context of the project, more than 30 publications have been obtained so far. The list is presented in section 4, references.

Technological transference

The three teams UdG, UJI and AIJU have participated in the technology transfer project "Visual Cad, Desarrollo de un optimizador de diseños modelados en sistemas CAD para su visualización en tiempo real", from the Spanish Ministery of Industry and led by Brainstorm company. This has allowed the transfer to industry of several of the results of the present project. Also, within the context of European project Gametools (see next section), relationships has been estabished with several Computer Games and virtual reality companies, such as Digital Legends, Gedas, and others that has allowed to estabish an stable framework for technology transfer. Also, a spinoff company that will allow further exploitation of the results of this project is currently under viability study at UdG technology transfer office.

Participation in international projects

We want also to outline that the three groups of this coordinated project, UdG, UJI, AIJU, participate together with Vienna Technical University, MSI laboratory of Limoges University, Technical University of Budapest and several companies of the area of computer games, simulation and virtual reality in the European project of the VIth Framework Gametools www.gametools.org coordinated by Mateu Sbert from UdG, that deals with the creation of highly realistic graphics software libraries for their use in computer games and virtual reality. A follow-up of this project is currently being prepared and will be submitted to the First Call of the VIIth. The three partners will participate and the project will be again coordinated by Mateu Sbert.
There has also been a Joint Action between the Technical University of Budapest and the UdG, led by Laszlo Szirmay and Mateu Sbert, during the period 2005-2006, "Synthesis of Photo-realistic Images in Computer Games and Virtual Reality Systems", that has allowed to deepen the collaboration between both groups.

Collaborations with other research groups.

About foreign groups working in similar subjects, and with whom there is a relationship, we find:

- Max Planck Institut of Saarbrucken, with Karol Myszowkski as contact person
- The Computer Graphics Institute of the Vienna Technical University, led by professor Werner Purgathofer
- Graphics Group from the Technical University of Budapest, led by professor Laszlo Szirmay
- MSI laboratory of the Université de Limoges led by professor Dimitri Plemenos, now on leave, and Djamchid Gharzanfapour.
- Multimedia Centre of Limburg University, with Philippe Bekaert as contact person
- Graphics Group from the University of Plzen, led by professor Vaclav Skala
- Computer Graphics Center of Carolus Prague University, with Vlastimil Havran as contact person
- University of Tianjin, in China, with professor Qing Xu as contact person. Common publications within this period are [Xu06A], [Xu06B].
- University of North Carolina at Chapel Hill, with professor John Halton as contact person
- In the USA we have also contacts with, Per Christensen, from Pixar, with Brian Barsky, professor at Berkeley University.

Within the Girona Graphics Group there is a collaboration with project TIN2004-08065-C02-02 "Interfaces avanzadas para la interaccion transparente con modelos altamente complejos en

entornos de realidad virtual", with publications [Bardera05], [Bardera06A], [Bardera06B], [Bardera06C], [Prados06]. A.Bardera and F.Prados are coadvised by Miquel Feixas.

Mateu Sbert coadvises a PhD student, Jorge Flórez, together with Josep Vehí, from Control Engineering and Intelligent Systems (EXIT) group within the Institut d'Informàtica i Aplicacions from the UdG. This has led so far to the publications [Flo06B] and [Florez06A]. There has been also a collaboration with the Broadband Communications and Distributed Systems (BCDS) from the same institute, with result published in the awarded paper [Urra05].

We are keeping in touch with most of the existing Spanish Computer graphics groups that work in areas related to simulation and Virtual Reality, which are:

- Computer Graphics section of the Technical University of Catalonia, Sección de Informática Gráfica de la Universidad Politécnica de Cataluña. Pere Pau Vázquez, formerly at UdG and whose PhD was advised by Mateu Sbert, is now a member from this group. A common publication within this period is [Vazquez06].
- Advanced Computer Graphics Group, Grupo de Informática Gráfica Avanzada, from University of Zaragoza.
- Graphics Group from the University of Granada, with Carlos Ureña. This group participates in a new project application together with UdG, UJI, and AIJU to the Spanish National Research Foundation.
- Graphics Group from the Technical University of Valencia, Grupo de Gráficos de la Universidad Politécnica de Valencia, with Roberto Vivó and Emilio Camahort. This group participates in European project Gametools, www.gametools.org

Subproject UJI:

Personnel in formation.

PhDs related with the subproject MATER-UJI.

- Inmaculada Remolar. "Real-time rendering of vegetable species". 2005. Universitat Jaume I. Advisor: Miguel Chover.
- Cristina Rebollo. "A multiresolution model for leaves rendering". 2006. Universitat Jaume I. Advisor: Miguel Chover.

In addition three PhD students are finalizing their works in subjects related to the project.

- Francisco Ramos. "LODStrips: A Level of Detail Model with Triangle Strips". Universitat Jaume I. Advisor: Miguel Chover.
- Pascual Castelló. "View-point Based Simplification for Polygonal Meshes". Universitat Jaume I. Advisors: Mateu Sbert and Miguel Chover.
- Alejandro Garcés. "Moderately Open Agent-Based Systems for Games". Universitat Jaume I. Advisors: Ricardo Quirós and Miguel Chover.

Other students starting their PhDs:

- Oscar Ripollés. "Geometric Modelling for the GPU".
- Jesús Gumbau. "Rendering Natural Environments"
- Carlos Gonzalez. "Geometric Simplification of CAD Models"
- Anna Puig. "Automatic Scene Creation"

Publications related with the project

In the context of the project, more than 22 publications have been obtained so far. Two of them are journal papers, five are publications in LNCS series, twelve in international conferences indexed by ISI and three are papers published in Spanish conferences. The list is presented in section 4, references.

Technological transference

The UJI team has participated with UdG and AIJU in the technology transfer project "Visual Cad, Desarrollo de un optimizador de diseños modelados en sistemas CAD para su visualización en tiempo real" (FIT-350101-2004-15), from the Spanish Ministry of Industry and led by Brainstorm company. This has allowed the transfer to industry of several of the results of the present project.

As a result of the developed technology and the knowledge acquired in the project, the Centre of Interactive Visualization of the University Jaume I has been created. This centre has born with the mission of transferring technological results in the context of Virtual reality and Web3D content using the new techniques developed in the field of the Computer Games. The first collaboration has been made with INFOGRAF S.L., a company that works in the real state field.

Participation in international projects

We have participated in the European project GameTools as a responsible partner for the geometry modules, in which it is tried to impel the development and transference of algorithms to the industry to facilitate the development of computer games: "GAMETOOLS - Advanced tools for developing highly realistic computer games" IST-2-004363 (www.gametools.org).

Collaborations with other research groups

International groups:

- The Computer Graphics Institute of the Vienna Technical University, led by Professor Werner Purgathofer
- Graphics Group from the Technical University of Budapest, led by professor Laszlo Szirmay
- MSI laboratory of the University of Limoges led by professor Dimitri Plemenos
- Graphics Group from the University of Plzen, led by professor Vaclav Skala

Spanish Groups:

- Computer Graphics Section of Polytechnic University of Valencia.
- Computer Graphics Group from University of Valencia

- Computer Graphics Group of University of Vigo.
- Computer Graphics Group of University of illes Balears
- Advanced Computer Graphics Group of University of Zaragoza.

Subproject AIJU:

Personnel in formation.

One specific person has been contracted by AIJU to support the integration of the resulting libraries in the eStudio application and to test and validate the enhancements obtained with the application of the research results. This specific person has been also in charge of create different applications to validate the results in different virtual reality environments and 3D web.

Publications related with the project

No scientific publications have been produced, but dissemination articles to companies in AIJU's Magazine have been published.

Technological transference

The resulting results have been transferred to Brainstorm Multimedia. This company has integrated some of the obtained libraries in the real time 3D graphics application e-Studio. Although the integration process has been very hard, the obtained results have been very positive.

Participation in international projects

We are participating in the European project GameTools in charge of create a good communication channel and a technology transfer path between Academia and Industry. "GAMETOOLS - Advanced tools for developing highly realistic computer games" IST-2-004363 (www.gametools.org).

Collaborations with other research groups.

In addition to the collaborations with the partners of the European project, we have established relations with important companies like Brainstorm Multimedia and Eon Reality.

We have set up collaboration agreements with Spanish Groups in different universities:

- Universitat Politècnica de València
- Universitat d'Elx

4 References

[Acebo05] E. Acebo, M. Sbert. Benfords Law for Natural and Synthetic Images, Eurographics Workshop on Computational Aesthetics in Graphics, Visualization and Imaging, Girona (Spain), May 2005

[Bardera05] A. Bardera, M. Feixas, I. Boada, M. Sbert. Medical Image Registration Based on Random Line Sampling, IEEE International Conference on Image Processing (ICIP 2005), Genova (Italy), 2005.

[Bardera06A] A. Bardera, M. Feixas, I. Boada, J.Rigau and M. Sbert. Medical Image Registration based on BSP and Quad-tree partitioning, Proceedings of the Third International Workshop on Biomedical Image Registration 2006, Utrecht, Netherlands, July 2006.

[Bardera06B] A. Bardera, M. Feixas, I. Boada and M. Sbert. High-dimensional normalized mutual informacion for image registration using random lines, Proceedings of the Third International Workshop on Biomedical Image Registration 2006, Utrecht, Netherlands, July 2006.

[Bardera06C] A. Bardera, M. Feixas, I. Boada and M. Sbert. Compression-based image registration, Proceedings of the IEEE International Symposium on Information Theory, 2006, Seattle, USA, July 2006.

[Belmonte04] O. Belmonte, I. Remolar, J. Ribelles, M. Chover, Marcos Fernández. Efficiently using connectivity information between triangles in a mesh for real-time rendering. Future Generation Computer Systems. Elsevier. ISBN: 0167-739X. 20. 8. pp. 1263-1273. Amsterdam, The Netherlands. 2004.

[Belmonte06] O. Belmonte, S. Aguado, S. Sancho. Triangle Strip Multiresolution Modelling using sorted edges. ICCS 2006 (LNCS 3992), University of Reading, UK, May, 2006.

[Castello0A] P. Castelló, M. Chover, M.Sbert, M.Feixas. Técnicas para Calcular la Entropía Dependiente de la Vista de una Escena 3D, XV Congreso Español de Informática Gráfica, ISBN/ISSN 84-9732-431-5, Granada, pp. 277-280, September 2005.

[Castello05B] Pascual Castelló, J. Francisco Ramos, M. Chover. A Comparative Study of Acceleration Techniques for Geometric Visualization. Lecture Notes in Computer Science 3515. International Conference on Computational Science (ICCS 2005). Springer-Verlag. ISBN: 3-540-26043-9. 2. pp. 240-247. Atlanta (USA). 2005.

[Castello06] P. Castelló, M. Sbert, M. Chover, M. Feixas. Techniques for Computing Viewpoint Entropy of a 3D Scene, Proceedings of ICCS 2006 (LNCS 3992), Reading, UK, May 2006.

[Castello07] P. Castelló, M. Sbert, M. Chover, M. Feixas. Viewpoint Entropy-Driven Simplification. WSCG2007. Plzen (Czech Republic), February. 2007.

[Feixas06A] Miquel Feixas, Mateu Sbert, Francisco González. A Unified Information-Theoretic Framework for Viewpoint Selection and Mesh Saliency, IIiA-06-06-RR, Institut d'Informàtica i Aplicacions, Universitat de Girona (Girona, Spain), 10-04-2006.

[Florez06A] Jorge Florez, Mateu Sbert, Josep Vehí and Miguel A. Sainz. Parallel implementation of a ray tracing of implicit surfaces using Interval Arithmetic, GAMM - IMACS International Symposion on Scientific Computing, Computer Arithmetic and Validated Numerics, Sep 2006.

[Florez06B] Jorge Florez, Mateu Sbert, Miguel A. Sainz, Josep Vehí. Improving the Interval Ray Tracing of Implicit Surfaces, Proceedings of Computer Graphics International 2006.

[Garces06] A. Garcés, R. Quirós, M. Chover, E. Camahort. Implementing Moderately Open Agent-Based Systems. IADIS International Conference, WWW/Internet. Murcia (Spain). 2006.

[Garces07] A. Garcés, R. Quirós, M. Chover, J. Huerta, E. Camahort. A Development Methodology for Moderately Open Multi-Agent Systems. 25th IASTED International Multi-Conference on Applied Informatics. Innsbruck (Austria). 2007.

[Garcia05A] Ismael García, Laszlo Szirmay-Kalos and Mateu Sbert. Tree rendering with billboard clouds, Third Hungarian Conference on Computer Graphics and Geometry, Budapest, 2005.

[Garcia05B] I. Garcia, M. Sbert, L. Szirmay-Kalos. Leaf Cluster Impostors for Tree Rendering with Parallax, Eurographics05, Dublin (Ireland), 2005.

[Gonzalez05] C. González, J. Gumbau, M. Chover. Generación por Hardware de Mapas de normales. Congreso Español de Informática (CEDI). (CEIG 2005). 2005.

[Gonzalez07] C. González, P. Castelló, M. Chover. A Texture-Based Metric Extension for Simplification Methods. GRAPP 2007. Barcelona (Spain). 2007.

[Gumbau06AJ. Gumbau, M. Chover. Conceptual design of a programmable geometry generator. Winter School of Computer Graphics (WSCG 2006) (Short paper). 2006.

[Gumbau06B] J. Gumbau, C. González, M. Chover. Fast GPU-based normal map generation for simplified models. Winter School of Computer Graphics (WSCG 2006) (Short paper). 2006.

[Gumbau07] J. Gumbau, O. Ripolles, M. Chover. LODManager: A Framework for Rendering Multiresolution Models in Real-time Applications. WSCG2007 (short paper). Plzen (Czech Republic), February. 2007.

[Mendez06A] Alex Méndez Feliu, Mateu Sbert. Efficient rendering of light and camera animation for navigating a frame array, Computer Animation and Social Agents CASA 2006 Geneve, Switzerland, July 2006.

[Mendez06B] Alex Méndez Feliu, Mateu Sbert. Obscurances in General Environments, Graphicon 2006, Novosibirsk, Russia, July 2006.

[Mendez06C] Alex Méndez Feliu, Mateu Sbert, Laszlo Szirmay-Kalos. Reusing Frames in Camera Animation, Journal of WSCG, ISSN 1231-6972, Vol. 14, 2006.

[Mendez06D] Alex Méndez Feliu, Mateu Sbert, J. Catàr, N. Sunyer, S. Funtané. Real-time Obscurances with Color Bleeding, SHADERX4 Sebastian St-Laurent and Wolfrang Engel (Eds.) Charles River Media, Hingham, MA, USA. January 2006.

[Prados06] F.Prados, A. Bardera, M. Feixas, I. Boada and M. Sbert. A Monte Carlo-based Fiber Tracking Algorithm using Diffusion Tensor MRI, Proceedings of the 19th IEEE International Symposium on Computer-Based Medical Systems 2006, Salt Lake City, USA, June 2006.

[Puig07] A. Puig-Centelles, O. Ripolles, M. Chover. Reviewing P2P Network Community Detection, EuroIMSA 2007 (IASTED). Chamonix, France. March, 2007.

[Ramos06A] F. Ramos, M. Chover, O. Ripolles and C. Granell. Efficient Implementation of LodStrips. Proceedings of the Sixth IASTED International Conference on. Visualization, Imaging and Image Processing, ISBN 0-88986-598-1, ISSN. 1482-7921, pp. 365-370. 2006.

[Ramos06B] F. Ramos, M. Chover, O. Ripolles and C. Granell. Continuous Level of Detail on Graphics Hardware. Lecture Notes in Computer Science Vol. 4245.Proceedings of the. 13th International Conference on Discrete Geometry for Computer Imagery. ISBN: 3-540-47651-2. 2006.

[Rebollo06A] C. Rebollo, I. Remolar, M. Chover, O. Ripollés. An Efficient Continuous Level of Detail Model for Foliage. Winter School of Computer Graphics (WSCG 2006). ISBN: 80-86943-03-8. 2006.

[Rebollo06B] C. Rebollo, I. Remolar, M. Chover, and J. Gumbau. Hardware-Oriented Visualisation of Trees. ISCIS 2006, LNCS 4263 proceedings.

[Rebollo07A] C. Rebollo, J. Gumbau, O. Ripolles, M. Chover, and I. Remolar. Fast Rendering of Leaves. The Ninth IASTED International Conference on Computer Graphics and Imaging CGIM 2007. February, 2007. Innsbruck (Austria). 2007.

[Rebollo07B] C. Rebollo, J. Gumbau, O. Ripolles, M. Chover, and I. Remolar. REVISTA

[Rigau05A] J. Rigau, M. Feixas, M. Sbert. Shape Complexity Based on Mutual Information, Proceedings of IEEE International Conference on Shape Modeling and Applications, MIT, Cambridge (USA), June 2005.

[Rigau05B] J. Rigau, M. Feixas, M. Sbert. An Information Theoretic Framework for Image Complexity, Eurographics Workshop on Computacional Aesthetics in Graphics, Visualization and Imaging, Girona (Spain), May 2005.

[Rigau06] J.Rigau, M.Feixas, M.Sbert. An Information-Theoretic Approach to Shape Complexity, Computer Graphics & Geometry Internet Journal (Vol. 8, No. 1), 2006.

[Ripolles05A] O. Ripollés, M. Chover. Búsqueda de tiras para modelos multirresolución estáticos. Congreso Español de Informática (CEDI). (CEIG 2005). 2005.

[Ripolles05B] O. Ripollés, M. Chover, F. Ramos. Quality Strips for Models with Level of Detail. Visualization, Imaging, and Image Processing (VIIP 2005). ACTA Press. ISBN: 0-88986-528-0. pp. 268-273. Benidorm (Spain). 2005.

[Rovira05] J. Rovira, P. Wonka, F. Castro, M. Sbert. Point Sampling with Uniformly Distributed Lines Symposium on Point-Based, Graphics (PBG05), June 2005, Stony Brook (USA).

[Sbert05] M. Sbert, D. Plemenos, M. Feixas, F. González. Viewpoint Quality: Measures and Applications, Eurographics Workshop on Computacional Aesthetics in Graphics, Visualization and Imaging, Girona (Spain), May 2005.

[Sbert06] M. Sbert, J. Rigau, M. Feixas and Laszlo Neumann. Systematic Sampling in Image-Synthesis, Proceedings of ICCSA 2006 (LNCS 3980), Glasgow, UK, May 2006.

[Szecsi06] Laszlo Szecsi, Laszlo Szirmay-Kalos, and Mateu Sbert. Light Animation with Precomputed Light Paths on the GPU, Proceedings of Graphics Interface, 2006.

[Szirmay05A] Laszlo Szirmay-Kalos, Mateu Sbert, Tamas Ummenhoffer. Real-Time Multiple Scattering in Participating Media with Illumination Networks, Rendering Techniques'05, ACM Siggraph Press, 2005.

[Szirmay05B] Laszlo Szirmay-Kalos, Gyorgy Antal, Mateu Sbert. Go with Winners Strategy in Path-Tracing, Journal of WSCG, Vol.13, No.2, pp.49-56, 2005.

[Szirmay06] Laszlo Szirmay-Kalos, Laszlo Szecsi and Mateu Sbert. GPUGI: Global Illumination Effects on the GPU, Eurographics 2006, Tutorial.

[Urra05] A. Urra, J.L. Marzo, M. Sbert, E. Calle. Estimation of the Probability of Congestion using Monte Carlo method in OPS Networks, IEEE Symposium on Computers and Communications (ISCC 2005), Cartagena (Spain), June 2005.

[Vazquez06] Pere-Pau Vazquez, Miquel Feixas, Mateu Sbert and Antoni Llobet. Real-time automatic selection of good molecular views, Computers & Graphics, Vol. 30, Issue 1, Pages 98-110, February 2006.

[Viola06] Ivan Viola, Miquel Feixas, Mateu Sbert and Meister Eduard Groeller. Importance-Driven Focus of Attention, IEEE Transactions on Visualization and Computer Graphics, Vol 12, No 5, September/October 2006.

[Xu06A]Qing Xu, Mateu Sbert. Adaptive Sampling Based on Fuzzy Inference, GRAPHITE 2006.

[Xu06B] Qing Xu, Mateu Sbert, Wei Wang and Lianping Xing. Fuzziness Driven Adaptive Sampling for Monte Carlo Global Illuminated Rendering, Proceedings of Computer Graphics International 2006.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

UPIC: towards an Unified framework for the Procurement and Implementation of information system Components TIN2004-07461-C02

Xavier Franch
Dept. Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya (LSI – UPC)
UPC-Campus Nord, mòdul Omega, room 120
E-08034 Barcelona (Spain)

Abstract

This work reports on the activities undertaken in the Project TIN2004-07461-C02.

Keywords: Information Systems, Off-The-Shelf Components.

1 Project objectives

Currently, enterprise information systems (IS) are mainly being developed as a combination of several components. These are usually purchased from specialized vendors who are furnished from a huge and continuously growing market that, requiring a significant implementation effort by the organization, including the development of in-house code. In the light of this new situation, organizations require new methodological patterns that take into account the diversity of implied processes and that focus specifically on these kinds of IS components. That is, those structural elements that are recognized in the first architectural decomposition of the system (namely, the one that directly comes from users requirements), which a target functionality of the system, and which become subject of selection, implementation and licensing processes.

The motivation of the UPIC project is to study, extend and apply methods, models, techniques and languages (hereafter, support utilities) to help the effective and efficient procurement and implementation processes of IS components in the context of the creation or renewal of an organization IS. The goal is, on one hand, to improve the quality of the product and the process, and, on the other hand, to improve the satisfaction of the people involved in the above mentioned processes, hence considering at the same time the evolutionary aspect of component-based IS. For this reason, these studies, extensions and applications are carried out within an unified perspective, focusing on those common support utilities that, although initially conceived to be applied in some particular provision or implementation activities, are also envisioned to be applicable in others. In this project, we consider mainly those IS components usually known as Commercial Off-The-Shelf

(COTS) components, although we are open to consider other types of IS components or even other distribution channels.

The UPIC project has four main goals: 1) to acquire, classify and characterise the knowledge that exists in the organizations related to the IS components market; 2) to define support utilities, and to develop tool support, for the effective and efficient IS component procurement in an evolutionary framework; 3) to define support utilities, and to develop tool support, for the efficient and reliable IS component implementation in an evolutionary framework; 4) to identify and solve the limitations observed in the support utilities when used in the context of the project.

This project is the continuation of the project TIC2001-2165.

2 Current State of the Project

We analyse each of the stated objectives of the project and mention references and other activities for each of them.

2.1 To acquire, classify and characterize the knowledge that exists in the enterprise setting concerning the offering of IS components

We have explored the notion of taxonomy as the way to organize the knowledge about IS components. We have formulated a method called GOThIC (formerly, GBRAM) composed of 7 stages. Most of our work has been the analysis of each of these stages. This is the subject of the PhD theses of C. Ayala, which is expected to be presented during 2007.

References: [ABF05a], [ABF05b], [AF05], [AF06a], [AF06b], [AF06c]

Activities:
- Collaboration with NTNU (see 3.6).
- Construction of tool support for building taxonomies for COTS components.

2.2 To define methods and techniques, and to build tool support, for the efficient and suitable procurement of IS components in an evolvable context

Our research has focused in the following lines:
- Quality models. We have performed some more experiences and focused our attention in the extension of usual quality models with non-technical factors.
- Agile methods into off-the-self-based development. We have started to analyse how agile methods can be adopted in COTS development and also in product line engineering.
- Formalization of the activities undertaken in off-the-shelf based development. We have defined usual activities, artefacts, roles and tools and established software processes upon them.

References: [Car06], [CF06], [CFQ05], [CFQ06a], [CFQ06b], [CFQ07a], [CFQ07b], [FT05], [NBF05a], [NBF05b], [NBF05c], [NBF06], [NBF07], [QFL05], [QFL06]

Activities:

- Collaboration with Politecnico de Torino, Centre Henry Tudor and University of Texas at Dallas (see 3.6).
- Construction of tool support for quality model construction.

2.3 To define methods and techniques, and to build tool support, for the efficient and suitable implementation of IS components in an evolvable context.

With regard to the implementation of integrated enterprise information systems, we continued the work on ERP implementation best practises research. After the successful defence of Jose Esteves' thesis on "Definition and Analysis of Critical Satisfaction Factors for ERP Implementation Projects", we started similar endeavours with the accepted thesis proposals of three new PhD. students: Luis Bibiano, on "Engineering implementation methods for CRM information systems"; Alberto Caldelas, on "Engineering implementation methods for SCM information systems"; and Rafael Silveira, on "Engineering IS integration methods through systematic EAI procurement and deployment". In relation to improving risk management in information systems development, we collaborated through an action-research project with a real project at UPC for the development of the new integrated academic management information system, where we extended and applied a classical risk management method from the Software Engineering Institute with additional risks factors drawn from our prior research on ERP implementation projects.

References: [ACEM05], [BDEP06], [BP06], [CP06], [CP06b], [CPG06], [EP05], [EP06], [EPC05], [EDP05b], [EPRR05], [RPS06], [SP06].

Collaborations:

- Collaboration with the Software Engineering Group from the Public Research Center Henry Tudor from the University of Luxemburg. They consider to use our results and methods for their support service in ERP procurement projects.
- Collaboration with the PRISMA academia management information system project at UPC, consisting in designing a personalized risk management method and supporting the development team in its application, plus designing a quality assurance procedure.
- Organization in Barcelona of the Internacional Congreso on PLS'05 (Partial Least Squares). Our group has used PLS as an analytical tool for studying several issues within ERP implementation critical success factors.
- Collaboration with SAP America and IE Business School for studying the impact of SAP products on Spanish SME companies.

2.4 To identify and solve the limitations presented by the methods, models, techniques and languages used in the context of the project

Concerning the conceptual tools used in the procurement part, we have focused our research in the formal underpinnings of quality models, goal-oriented engineering and UML. Several PhD students are performing their PhD thesis in these topics. For quality modelling, we have developed a framework that acts as ontology for quality which allows deriving concrete quality frameworks (this is the PhD topic of X. Burgués). For goal-oriented engineering, we are analysing both methods and languages. For methods, C. Cares is focusing on the requirements engineering stage, whilst G. Grau is defining a reengineering methodology based on goal-oriented concepts. For language, we have defined a reference framework for the *i** language.

With regard to the formalization of semantic interoperability and integration within an ontological framework, we continued prior work con concept analysis and semantic modelling. This work comes from an AI background, where it is also being applied, but is also applicable to COTS procurement and specially integration. A new PhD student, Manuel Atencia, is starting to work on the issue of ontological semantic alignment. For the case of dependability analysis, a real case study on a neonatal intensive care unit could be addressed, while its formalization has been tackled also with OCL within the context of UML. For what has to do with the trial of varied research methods in IS, we have applied grounded theory to the analysis of data resulting from in-depth case studies in ERP implementations, and action-research approaches in our collaboration with real IS development projects.

References: [AFCF05], [Aya+05], [AS06], [BFM07], [BFR05], [BKMT05], [CFMa06], [CFM06b], [CFM07], [CFMA06], [CFMQ07], [Clo+07], [CK05], [EP05], [EP05b], [FGQ05], [FMQ07], [Fra05], [Fra06], [Fra+07], [GCFN06], [GFA06], [GFM05a], [GFM05b], [GFM07], [Gra+05], [HFM05], [KA06], [KK05], [KK06], [KPRS05], [KS05], [KSU05], [SK05], [Sch05], [KSU05]

Collaborations:

- Collaboration with City University, Linz University, University College London and ITCirs (see 3.6).
- Development of different prototypes as tool support.
- Participation of Marco Schorlemmer and Yannis Kalfoglou in the organization of the Dagstuhl international research seminar on "Semantic Interoperability and Integration".
- Participation of Marco Schorlemmer in the preparation of the european ESPRIT Project OpenKnowledge.
- Participation of Juliana Küster in the creation and organization of the first international workshop FESCA, first held in Barcelona, as well as in the organization of its second edition, held in Edimburg.
- Participation in the accepted proposal of the excellence network MIFFISIS and its associated events, such as the "Workshop en Métodos de Investigación y Fundamentos Filosóficos en Ingeniería del Software y Sistemas de Información", and derived event PHISE'05 "Philosophical Foundations of Information Systems Engineering", an ancillary workshop to the CAiSE'05 conference.

3 Results of the Project

We include next information about the research team for the whole integrated project, summary of publications (see section 4 for a complete list), involvement in academic events, industrial experiences, collaboration with other research groups, participation in other projects and software developed. The information is restricted to the period of the project (from December 28th, 2004).

3.1 Research team

The staff in the project is composed of 15 people most of them with full dedication, up to 13.5 EDPs, originally 12 doctors (X. Franch, project responsible; P. Botella, J. Corcó, J. Küster, E. Mayol, J.A. Pastor, C. Quer, J.M. Ribó and M. Schorlemmer, full dedication; T. Aluja, J. Casanovas, J. Kafoglou, half dedication) and 3 PhD. students (C. Ayala, X. Burgués and J. Marco, full dedication). Besides, some other PhD students have joint the team and participate in the activities of the project: L. Bibiano, A. Caldelas, C. Cares, J.P. Carvallo, R. Clotet (who obtained the FPI

position assigned to the project), G. Grau, M. Haya, M. Lempert, L. López, O. Méndez, F.J. Navarrete, J. Pradel, J. Raya and R. Silveira. During the period of UPIC, two of the PhD students have presented the thesis, J. Marco and J.P. Carvallo. Others have presented their PhD project: L. Bibiano, A. Caldelas, C. Cares, G. Grau, F.J. Navarrete and R. Silveira.

3.2 List of publications

We have published 60 refereed publications since January 2005 whilst we have 9 publications accepted and pending of publication. The complete list appears in section 4. When the cell is of the form *x*+*y*, it means *x* already available publications and *y* accepted but not published yet. The table below includes some classification criteria:

	Total	Scope	Indicator	Review	Joint	Editorial	Others
Journals	6+4	I: 5+4 R: 1+0	I: 0+3	F: 2+1 S: 3+3	1	IEEE: 1+1 Elsevier: 0+1 ACM: 1+0 Springer: 1+1	
Conferences	34+3	I: 32+3 R: 2+0	N: 10+0 A: 2+1	F: 32+3 A: 2+0	4+1	IEEE: 9+3 LNCS: 8+0 LNAI: 3+0 Springer: 1+0	T: 4+0 S: 2+0
Workshops	16	I: 14 R: 2		F: 16	1		S: 1
Book chapters	4+2	I: 4+2		F: 1+0 I: 0+2		Idea Group: 4+0 MIT Press: 0+2	

Scope of the publication or conference: I: International; R: Regional
Indicator: X: indexed; N: classified as "Notable" by the UPC; A: award received
Review: F: full paper; A: abstract; S: selected from former work; I: invitation
Joint: with other research groups
Editorial: we highlight some reputed editorials and collections
Others: T: description of tool; S: published as short paper

Concerning editorial responsibilities:

- Xavier Franch was *Proceedings co-editor* of the 4th *International Conference on COTS-Based Software Systems* (ICCBSS, http://www.iccbss.org), held at Bilbao (Spain), February 2005.
- Xavier Franch is co-editor of the Proceedings of the *1st International Workshop on Agile Product Line Engineering* (APLE, http://www.lsi.upc.edu/events/aple/), published in electronic form with ISSN 1613-0073.
- Xavier Franch is currently *Co-editor* of a special issue of the SCI-indexed *Journal of Systems and Software* (Elsevier).

3.3 Participation in academic events

The members of the group have been and are being involved in the following academic events:

- Organizing committees:
  - Pere Botella was *Organizing Committee chair* of the XI *Jornadas de Ingeniería del Software y Bases de Datos* (JISBD, http://congress.cimne.upc.es/jisbd06), held at Sitges (Spain), October 2006.

- Xavier Franch was *Tutorial chair* of the XI *Jornadas de Ingeniería del Software y Bases de Datos* (JISBD, http://congress.cimne.upc.es/jisbd06), held at Sitges (Spain), October 2006.
- Xavier Franch has been workshop co-organizer of the following workshops:
  - *2nd International Workshop on Models and Processes for Evaluating off-the-shelf Components* (MPEC, http://www.lsi.upc.edu/events/mpec/), held during the 27th *International Conference on Software Engineering* (ICSE), Saint Louis (Missouri, USA), May 2005.
  - *1st International Workshop on Service-Oriented Computing: Consequences for Engineering Requirements* (SOCCER, http://www.lsi.upc.edu/events/soccer/), held during the 13th *International Conference on Requirements Engineering* (RE), Paris (France), August 2005.
  - *2nd International Workshop on Service-Oriented Computing:Consequences for Engineering Requirements* (SOCCER, http://www.lsi.upc.edu/events/soccer/), held during the 14th *International Conference on Requirements Engineering* (RE), Minneapolis (Minessota, USA), September 2006.
  - *1st International Workshop on Agile Product Line Engineering* (APLE, http://www.lsi.upc.edu/events/aple/), held during the 10th *International Software Product Lines Conference* (SPLC), Baltimore (Maryland, USA), August 2006.

- Program committees:
  - Xavier Franch was *Program Committee co-chair* of the 5th *International Conference on COTS-Based Software Systems* (ICCBSS, http://www.iccbss.org), held at Orlando (Florida, USA), February 2006.
  - Xavier Franch is currently *Program Committee chair* of the XII *Jornadas de Ingeniería del Software y Bases de Datos* (JISBD, http://congress.cimne.upc.es/jisbd06), to be held at Zaragoza (Spain), September 2007.
  - The members of the group have been, or currently are, members of more than 20 program committees belonging to the following conferences and workshops: CAiSE, RE, ICCBSS, SEKE, Euromicro, WER, JISBD, PHISE, IDEAS, SBES, ITNG, DSOA, APLE, MPEC, ICCSEA.
  - The members of the group have been, or currently are, reviewing papers for the following journals: IEEE Software, IEE Proceedings Software, JSS (Elsevier).
- Other activities:
  - Pere Botella was invited speaker at the *Philosophical Foundations of Information Systems Engineering Workshop* (PHISE), an ancillary workshop to the CAiSE'05 conference.

3.4 Industrial experiences

The members of the group have participated, or are currently participating in the following industrial collaborations:

- Analysis of the feasibility of the adoption of the UML and J2EE technologies by the UPCnet organization.
  Members of the group: X. Franch, P. Botella, J. Pradel, J. Raya
  Duration: February 2005 – May 2005
  Total amount: 6.750 euros
- Assessment activities in requirements management in the information systems development of the UPCnet organization.
  Members of the group: X. Franch
  Duration: July 2006 – September 2006
  Total amount: 1.540 euros

3.5 Other projects and innitiatives

The members of the group have participated, or are currently participating, in the following projects:

- Requirements Engineering for Multi-Stakeholder Distributed Systems (MSDS).
  Acciones Integradas program (HU2005-0021), MEC
  Partner: Johannes Kepler University (Linz, Austria)
  Members of the group: X. Franch (responsible), J. Marco, R. Clotet, L. López
  Duration: January 2006 – December 2007
  Total amount: 6.280 euros
- SODA: Services Oriented Devices & Delivery Architectures
  PROFIT program (FIT-340000-2006-312), Ministery of Industry
  Members of the group: E. Mayol (responsible), P. Botella, J. Marco, J.A. Pastor
  Duration: January 2006 – December 2008
  Total amount: 58.362 euros
- NESSI (Networked European Software&Services Initiative, http://www.nessi-europe.com)
  European Technology Platform on software and services.
  The group is member of NESSI (representative: Pere Botella), and also member of the NESSI WG on Software Engineering (Pere Botella, Xavier Franch).
- INES (Iniciativa Española de Software y Servicios, http://www.ines.org.es/)
  Spanish Technology Platform on software and services.
  The grous is member of INES, elected member of the INES Board and member of the NESSI WG on Software Engineering (representative: Pere Botella).
- PROMETEO (http://www.prometeo-office.org/)
  Spanish Technology Platform for Embedded Systems
  The group (P. Botella) has participated in some meetings as member of PROMETEO.

3.6 Collaboration with other research groups

We are collaborating, or about to collaborate, with the following research groups represented by the mentioned researchers:

- City University, HCI group, responsible Neil Maiden. We are using the modelling language *i** for representing software systems architecture, in which this group has lots of expertise. The collaboration started at 2003 and is currently active. In the period of the period of UPIC, we have written 2 papers together and have another accepted for publication. Also, we are extending the REDEPEND tool developed at the City University with several functionalities. We have co-organized the two editions of the SOCCER workshop.
- Linz University, responsible Paul Grünbacher. We are putting together the *i** language, the win-win negotiation approach and variability modelling for representing and monitoring multistakeholder distributed systems. We have got an Acciones Integradas billateral project for the period 2006-07. We have a first joint paper accepted and others are in process.
- Politècnico de Torino, responsible Marco Torchiano. We are collaborating in the field of COTS based development. In the period of UPIC, we have produced a paper about the roles and activities on COTS selection processes. We have applied twice for a joint project in the Acciones Integradas program but with negative results.

- University College London, responsible Anthony Finkelstein. In the field of requirements negotiation during COTS selection, one former PhD student of Anthony Finkelstein, Carina Alves (recently presented the thesis), is using our quality models for describing the features of the components. Our collaboration resulted in a joint paper.
- Centre Henry Tudor, responsible Eric Dubois. This group is building tool support for COTS selection expected to be tailored with different methods and processes, among them they are interested in ours. We are currently collaborating in the identification of a catalogue of requirements patterns from their requirements books and have planned a paper on that issue.
- Center ITC-irst, responsible Anna Perini. This group has a prominent position in *i** related tools and methodologies. One of our PhD students, Carlos Cares, has made a visit of 3 months, to validate some of the ongoing results of his thesis. As a result, we have scheduled the publication of 3 joint papers and also are developing a XML format for *i** to be adopted by the community.
- University of Texas at Dallas, responsible Lawrence Chung and Kendra Cooper. This group has great expertise in goal-oriented requirements engineering and off-the-shelf component-based systems. We have collaborated in the past organizing the MPEC workshop, and now we are collaborating in the analysis of agile methodologies applied to the product line context, which has generated the APLE workshop and a special issue on the JSS journal.
- NTNU Center (Trondheim), responsible Reidar Conradi. This group has great expertise in the application of empirical methods for validating research hypothesis, especially in the context of off-the-shelf component-based systems. One of our PhD students, Claudia Ayala, has made a visit of 3 months, to validate the findings of her thesis. As a result, she has built a first prototype of wiki-collaborative system for managing component information, and we have submitted a first joint paper to an OSS conference. We are currently planning to apply for a European project.

3.7 Software developed

- DesCOTS, a tool for the Description of COTS components [GCFQ04]. The tool is composed of four subsystems [QFL05, QFL06] and is currently available at http://www.lsi.upc.es/~gessi/QMTool/QMTool.html.
- REDEPEND-REACT, a plugin in the REDEPEND tool for evaluating Requirements Architectures [GFM05]. The tool is currently available at http://www.lsi.upc.es/~ggrau/REDEPEND-REACT/. We are currently coordinating further efforts with the owner of the REDEPEND tool, the City University.
- J-PR*i*M, a tool for reengineering information systems [GFA06]. The tool is currently available at http://www.lsi.upc.es/~ggrau/JPRIM/.
- AR3L, a tool for assigning responsibilities to layers in a 3-layered architecture based on AndroMDA. The tool is currently available at http://www.lsi.upc.edu/~gessi/AR3L/.
- ASIRCA, an in-house tool for assigning responsibilities to layers in a 3-layered architecture. The tool is currently available at http://www.lsi.upc.edu/~gessi/ASIRCA/.
- OCLassistant, a tool for generating OCL expressions by the application of several patterns. The tool is currently available at http://www.lsi.upc.edu/~gessi/OCLassistant.

4 References

[ABF05a] C. Ayala, P. Botella, X. Franch. "On Goal-Oriented COTS Taxonomies Construction". Procs. 4th ICCBSS Conference, LNCS 3412, Springer-Verlag, Feb. 2005.

[ABF05b] C. Ayala, P. Botella, X. Franch. "Construction of a Taxonomy for Requirements Engineering COTS Components". *Journal of Computer Science and Technology*, 5(2), Institute of Computing Technology, July 2005.

[AF05] C. Ayala, X. Franch. "Transforming Software Package Classification Hierarchies into Goal-Based Taxonomies". *Procs. 16th DEXA Conference*, LNCS 3588, Springer-Verlag, Sept 2005.

[AF06a] C. Ayala, X. Franch. "A Goal-Oriented Strategy for Supporting Commercial Off-The-Shelf Components Selection". *9th International Conference on Software Reuse*, LNCS 4039. 2006.

[AF06b] C. Ayala, X. Franch. "Overcoming COTS Marketplace Evolvability and Interoperability". *CAiSE Forum,* 2006.

[AF06c] C. Ayala, X. Franch. "Domain Analysis for Supporting Commercial Off-The-Shelf Components Selection", *Procs. 24th International Conference on Conceptual Modeling* (ER'06), LNCS 4215, Springer Verlag, Tucson (USA), Nov. 2006.

[AFCF05] C. Alves, X. Franch, J.P. Carvallo, A. Finkelstein. "Using Goals and Quality Models to Support the Matching Analysis during COTS Selection". Procs. 4th ICCBSS Conference, LNCS 3412, Springer-Verlag, Feb. 2005.

[AS06] M. Atencia, M. Schorlemmer. "Situated Semantic Alignment". *Proceedings of the 4th European Workshop on Multi-Agent Systems* (EUMAS'06), Dunin-Keplicz, Barbara; Omicini, Andrea; Padget, Julian eds., CEUR Workshop Proceedings, Vol. 223, CEUR-WS, 2006. [ISSN 1613-0073], Lisbon, Portugal, December 2006.

[Aya+05] C. Ayala *et al.* "A Comparative Analysis of i*-Based Agent-Oriented Modeling Languages". *Procs. 17th SEKE Conference*, KSI, July 2005.

[BDEP06] V. W. Bohorquez, E. Dans, J. Esteves, J. A. Pastor. "Analyzing the Impact of Enterprise Systems in the Personnel Size of SMEs". *International ICIS Workshop on Enterprise Systems Research in Management Information Systems*, International Conference on Information Systems (ICIS), Association for Information Systems (AIS), Milwakee, USA, December 2006.

[BFR05] X. Burgués, X. Franch, J.M. Ribó. "A MOF-Compliant Approach to Software Quality Modeling". Procs. 24th International Conference on Conceptual Modeling (ER'05), LNCS 3716, Springer Verlag, Klagenfurt (Austria), Oct. 2005.

[BKMT05] G. Baxter, J. Küster-Filipe, A. Miguel and K. Tan. "The Effects of Timing and Collaboration on Dependability in the Neonatal Intensive Care Unit". *Constituents of Modern System-safety Thinking: Proceedings of the Thirteenth Safety-critical Systems Symposium.* Springer. London, UK, 2005.

[BP06] L. Bibiano, J. A. Pastor. "Towards a Definition of a CRM Systems Life-Cycle". *International European and Mediterranean Conference on Information Systems* (EMCIS'06), Brunel University Ed., Alicante, Spain, July 2006.

[Car06] J.P. Carvallo. "Supporting Organizational Induction and Goals Alignment for COTS Components Selection by Means of i*". *International Conference on COTS-Based Software Systems* (ICCBSS'06). IEEE Computer Society Press. 2006.

[CF06] J.P. Carvallo, X. Franch. "Extending the ISO/IEC 9126-1 Quality Model with Non-Technical Factors for COTS Components Selection". *Workshop on Software Quality* (WOSQ'06). 2006.

[CFMa06] C. Cares, X Franch, E. Mayol. "Perspectives about paradigms in software engineering". Workshop on Philosophical Foundations of Information Systems Engineering. PHISE. Luxembourg, _, Luxembourg. June, 2006.

[CFM06b] C. Cares, X. Franch, E. Mayol. "Extending Tropos for a Prolog Implementation: A Case Study Using the Food Collecting Agent Problem". Lecture Notes in Artificial Intelligence. 3900. 0. ISSN:0302-9743. January, 2006

[CFM07] C. Cares, X. Franch, E. Mayol. "Using Antimodels to Define Agents' Strategy". VII Clima conference, Lecture Notes in Artificial Intelligence (LNAI). 0. ISSN:0302-9743. 2007.

[CFMA06] C. Cares, X. Franch, E. Mayol, E. Alvarez. "Goal-Driven Agent-Oriented Software Processes". *Proceedings IEEE 32th Euromicro*, 2006.

[CFQ05] J.P. Carvallo, X. Franch, C. Quer. "A Quality Model for Requirements Management Tools". Book chapter in *Requirements Engineering for Sociotechnical Systems*, Idea Group Publishing, Feb. 2005.

[CFQ06a] J.P. Carvallo, X. Franch, C. Quer. "Managing Non-Technical Requirements in COTS Components Selection". *Proceedings of the IEEE International Conference on Requirements Engineering* (RE), 2006

[CFQ06b] J.P. Carvallo, X. Franch, C. Quer. "Un Catálogo de Factores de Calidad para la Definición de Requisitos No-Técnicos en la Selección de Componentes COTS". *Actas 9th. Workshop on Requirements Engineering*, WER 2006.

[CK05] A. Cavarra, J. Küster-Filipe. "Combining Sequence Diagrams and OCL for Liveness". *Proceedings of ETAPS Satellite Workshop on Semantic Foundations of Engineering Design Languages*. Electronic Notes in Theoretical Computer Science (Elsvier), Barcelona, Spain, January 2005.

[CP06a] A. Caldelas, J. A. Pastor. "A Definition of SCM Information Systems through SCOR". *International European and Mediterranean Conference on Information Systems* (EMCIS'06), Brunel University Ed., Alicante, Spain, July 2006.

[CP06b] A. Caldelas, J. A. Pastor. "Towards a Research Agenda for SCM Information Systems". *International Americas Conference on Information Systems* (AMCIS'06), Association for Information Systems (AIS), Acapulco, México, August 2006.

[CPG06] A. Caldelas, J. A. Pastor, J. N.D. Gupta. "An Initial Research Agenda for SCM Information Systems". *International Journal on Networking and Virtual Organizations* (ISSN: 1470-9503), Inderscience Publishers.

[EDP05b] J. Esteves, E. Dans, J. A. Pastor. "Achieving Business Benefits through Enterprise Systems Use in SME". *Proceedings of the IsOneWorld Information Systems Conference,* Las Vegas, USA, March 2005.

[EP05] J. Esteves, J. A. Pastor. "An ERP Life-Cycle Costs Model". *Encyclopedia of Information Science and Technology* (Idea Group Publishing), USA, June 2005.

[EP05b] J. Esteves, J. A. Pastor. "Understanding an ERP System Implementation in a Higher Education Institution: A Grounded Theory Approach". *Qualitative Case Studies on Implementation of Enterprise Wide Systems*, Idea Group Publishing, Inc., USA, January 2005.

[EP06] J. Esteves, J. A. Pastor. "A Critical Success Factor's Relevance Model for SAP Implementation Projects". *Managing Business with SAP: Planning, Implementation and Evaluation*, Idea Group Publishing, Inc., pag. 240-261, USA, February 2006.

[FGQ05] X. Franch, G. Grau, C. Quer. "Un Marco para la Definición de Métricas sobre Modelos de Dependencias de Actores". *Anais do WER05 - Workshop em Engenharia de Requisitos*, Porto, Portugal, Junho 13-14, 2005. Joao Araújo, Amador Durán Toro, Joao Falcao e Cunha. ISBN:972-752-079-0. 2005.

[FPR06] X. Franch, J. Pradel, J. Raya. "Asignación sistemática de responsabilidades en una arquitectura en tres capas". *Actas XI Jornadas en Ingeniería del Software y Bases de Datos*, 2006.

[Fra05] X. Franch. "On the Lightweight Use of Goal-Oriented Models for Software Package Selection". *Procs. 17th CAiSE Conference*, LNCS 3520, Springer-Verlag, June 2005.

[Fra06] X. Franch. "On the Quantitative Analysis of Agent-Oriented Models". *Proceedings CAiSE*, LNCS 4001, 2006.

[FT05] X. Franch, M. Torchiano. "Towards a Reference Framework for COTS-Based Development: A Proposal". *Procs. 2nd MPEC Workshop*, ACM Digital Library, May 2005.

[GCFN06] G. Grau, C. Cares, X. Franch, F.J. Navarrete. "A Comparative Analysis of i*Agent-Oriented Modelling Techniques". *Proceedings of The Eighteenth International Conference on Software Engineering and Knowledge Engineering.* 2006.

[GFA06] G. Grau, X. Franch, S. Ávila. "J-PRiM: A Java Tool for a Process Reengineering i* Methodology". *Proceedings of the 14th IEEE International Requirements Engineering Conference.* 2006.

[GFM05a] G. Grau, X. Franch, N.A.M. Maiden. "REDEPEND-REACT: an Architecture Analysis Tool". *13th IEEE Requirements Engineering International Conference* (RE'05). IEEE Computer Society. ISBN:0-7695-2445-1. 2005.

[GFM05b] G. Grau, X. Franch, N. Maiden. "A Goal-Based Round-Trip Method for System Development". *Procs. 10th REFSQ Workshop*, Essener Informatik Beiträge, June 2005.

[Gra+05] G. Grau et al. "RiSD: A Methodology for Building i* Strategic Dependency Models". *Procs. 17th SEKE Conference*, KSI, July 2005.

[HFM05] M. Haya, X. Franch, E. Mayol. "The Use of UML Activity Diagrams and the i* Language in the modeling of the Balanced Scorecard Implantation Process". *Journal of Computer Science and Technology*, 5(2), Institute of Computing Technology, July 2005.

[KA06] J. Küster-Filipe, S. Anderson. "On a time enriched OCL liveness template". *International Journal on Software Tools for Technology Transfer (STTT)* 8(2):156-166, Springer. June 2006.

[KK05] L. Kloul, J. Küster-Filipe. "From Interaction Overview Diagrams to PEPA nets". *Proceedings of the 4th Workshop on Process Algebras and Timed Activities (PASTA'05).* Electronic Notes in Theoretical Computer Science (Elsvier), Edimburgh, Scotland, UK, September 2005.

[KK06] L. Kloul, J. Küster-Filipe. "Modelling Mobility with UML2.0 and PEPA Nets". *Sixth International Conference on Application of Concurrency to System Design (ACSD 2006*. IEEE Computer Society Press. Turku, Findland, June 2006.

[KPRS05] J. Küster-Filipe, I. Poernomo, R. Reussner and S. Shukla (Eds.), *Proceedings of the First International Workshop on Formal Foundations of Embedded Software and Component-based Software Architectures (FESCA 2004).* Electronic Notes in Theoretical Computer Science (Elsvier), Barcelona, Spain, December 2004.

[KPRS05] J. Küster-Filipe, I. Poernomo, R. Reussner and S. Shukla (Eds.). *Proceedings of the Second International Workshop on Formal Foundations of Embedded Software and Component-based Software Architectures (FESCA 2005).* Electronic Notes in Theoretical Computer Science (Elsvier), Edimburgh, Scotland, UK, December 2005.

[KS05] Y. Kalfoglou, M. Schorlemmer. "Using Formal Concept Analysis and Information Flow for Modelling and Sharing Common Semantics: Lessons Learnt and Emergent Issues". *Proceedings of the 13th International Conference on Conceptual Structures: Common Semantics for Sharing Knowledge.* Lecture Notes in Artificial Intelligence (ISSN 0302-9743, ISBN 3-540-27783-8), Kassel, Germany, July 2005.

[KSU05] Y. Kalfoglou, M. Schorlemmer, M. Uschold, A. Sheth, S. Staab. "Semantic Interoperability and Integration: Executive Summary". *Semantic Interoperability and Integration. Dagstuhl Seminar Proceedings 04391*, IBFI Schloss Dagstuhl, Dagstuhl, Germany, 2005.

[MFA05] J. Marco, X. Franch, J. Alvarez. "Implementando acceso directo y secuencial a colecciones de datos mediante aspectos". *IEEE América Latina*, IEEE-CS, March 2005.

[NBF05a] F.J. Navarrete, P. Botella, X. Franch. "Maturity levels for Commercial Off-The-Shelf (COTS) selection processes". *Procs. 2nd International Conference on Software Processes and Software Metrics* (SPSM'05), Madrid (Spain), Oct. 2005.

[NBF05b] F.J. Navarrete, P. Botella, X. Franch. "How Agile COTS selections methods are (and can be)?". *Procs. 30th IEEE EUROMICRO Conference*, IEEE-CS, Sept. 2005.

[NBF05c] F.J. Navarrete, P. Botella, X. Franch. "Análisis de los Métodos de Selección de Componentes COTS". *Actas IX Jornadas de Ingeniería del Software y Bases de Datos* (JISBD). Granada, Granada, España. September, 2005.

[NBF06] F.J. Navarrete, P. Botella, X. Franch. "An Approach to Reconcile the Agile and CMMI Contexts in Product Line Development". International Workshop on Agile Product Line Engineering, agosto 2006.

[QFL05] C. Quer, X. Franch, X. Lopez-Pelegrín. "DesCOTS-EV: A Tool for the Evaluation of COTS Components". *International Conference on Requirements Engineering*, 2005. RE'05. Sorbonne, Paris, France. IEEE Computer Society. September, 2005.

[QFL06] C. Quer, X. Franch, X. López-Pelegrín. "DesCOTS-SL: A Tool for the Evaluation of COTS Components". *Proceedings of the 14th IEEE International Requirements Engineering Conference* (RE'06). ISBN 0-7695-2555-5, 2006.

[RPS06] J. Rodon, J. A. Pastor, F. Sesé. "Promoting the Integration and Use of an IOIS: an Action-Oriented Perspective". *International ICIS Workshop on Difussion and Integration of Information systems and Technology* (DIGIT), International Conference on Information Systems (ICIS), Association for Information Systems (AIS), Milwakee, USA, December 2006.

[Sch05] M. Schorlemmer. "On the Mathematical Foundations of Semantic Interoperability and Integration". *Semantic Interoperability and Integration. Dagstuhl Seminar Proceedings 04391*, IBFI Schloss Dagstuhl, Dagstuhl, Germany, 2005.

[SK05] M. Schorlemmer, Y. Kalfoglou. "Progressive Ontology Alignment for Meaning Coordination: An Information-Theoretic Foundation". *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multi-Agent Systems*. (ISBN 1-59593-094-9), Utrecht, The Netherlands, July 2005.

[SP06] R. Silveira, J. A. Pastor. "A Model for EAI Tools Evaluation". *International European and Mediterranean Conference on Information Systems* (EMCIS'06), Brunel University Ed., Alicante, Spain, July 2006.

Pending of publication

[BFM07] X. Burgués, X. Franch, J.M. Ribó. "Improving the Accuracy of UML Metamodel Extensions by Introducing Induced Associations". *Software and Systems Modeling Journal* (SoSyM), Springer, 2007.

[CFMQ07] C. Cares, X. Franch, E. Mayol, C. Quer. "A Reference Model for *i**". Book chapter in Giorgini P., Maiden N., Mylopoulos J., Eric Yu (eds.) *Social Modeling for Requirements Engineering*, the MIT Press, to be published in 2007.

[CFQ07a] J.P. Carvallo, X. Franch, C. Quer. "Determining Criteria for Selecting Software Components: Lessons Learned". *IEEE Software*, March/April 2007.

[CFQ07b] J.P. Carvallo, X. Franch, C. Quer. "Towards a Unified Catalogue of Non-Technical Quality Attributes to Support COTS-Based Systems Lifecycle Activities". *Proceedings of the 6th IEEE International Conference on COTS-Based Software Systems* (ICCBSS'07), 2007.

[Clo+07] R. Clotet, X. Franch, P. Grünbacher, L. López, J. Marco, M. Quintus, N. Seyff, "Requirements Modelling for Multi-Stakeholder Distributed Systems: Challenges and Techniques". 1st IEEE International Conference on Research and Challenges in Information Systems, 2007.

[FMQ07] X. Franch, E. Mayol, C. Quer. "On the Use of *i** for COTS Components Selection: Principles and Consequence". Book chapter in Giorgini P., Maiden N., Mylopoulos J., Eric Yu (eds.) *Social Modeling for Requirements Engineering*, the MIT Press, to be published in 2007.

[Fra+07] Franch, X.; Grau, G.; Mayol, E.; Quer, C.; Ayala, C.; Cares, C.; Navarrete, F.; Haya, M.; Botella, P. "Systematic Construction of i* Strategic Dependency Models for Socio-Technical Systems". *International Journal of Software Engineering and Knowledge Engineering* (IJSEKE), 2007.

[GFM07] G. Grau, X. Franch, N. Maiden. "PRiM: A Methodology for Reverse Engineering". *Information Systems and Technology*, Elsevier 2007.

[NBF07] F.J. Navarrete, P. Botella, X. Franch. "Reconciling Agility and Discipline in COTS Selection Processes". *Proceedings of the 6th IEEE International Conference on COTS-Based Software Systems* (ICCBSS'07), 2007.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Infraestructura software para el desarrollo de servicios en red confiables y adaptables TIN2004-07474-C02

Ricardo Jiménez Peris *
Universidad Politécnica de Madrid

Francisco J. Ballesteros Cámara †
Universidad Rey Juan Carlos

Abstract

The goal of this project is the development of an integral infrastructure for the support of dependable networked services. The main purpose is to achieve a single software architecture that supports the development of applications and services deployed in different settings such as ubiquitous environments and wide and local area networks. The resulting architecture will focus on service dependability due to the increasing dependence that people and organizations have on information systems. Service dependability is measured in terms of attributes such as availability, adaptability, scalability and data consistency. Some of these attributes have not been successfully combined till now, what means that the study of how to provide them in an integrated fashion is an important scientific challenge. Two applications will be built in order to validate experimentally the resulting platform: an adaptable domotic environment for disabled people and a mobile e-commerce application. This project will embrace the new oriented service computing paradigm in which the architecture is structured as a pyramid of interoperable services at different abstraction levels. Following the spirit of service oriented computing, services will deployed in very different environments such as local area networks, ubiquitous environments, extended area networks and heterogeneous networks (combination of the previous ones). Dealing with this variety of environments raises interesting problems to be solved. Ubiquitous environments differ from traditional local and extended area networks: they exhibit an inherent lack of resources (memory, CPU, batteries, etc.), show a high level of dynamicity (there are mobile devices which connect and disconnect frequently) and require specific services (localization, physical environment models, etc.). Nowadays, there is no platform in which networked applications across heterogeneous networks can be built. In this project, a service oriented software architecture will be developed that will provide the necessary support for ubiquitous and non ubiquitous environments (LAN, WAN, Mirynet). Thanks to the coordination of the two teams in the project, it will be possible to offer an integral infrastructure for dependable services in which applications could be deployed in heterogeneous networks what will enable to combine the advantages of the different kinds of environments.

Keywords: Database Replication, Pervasive Computing, Middleware, Application Server Replication, Transactions, Service-Oriented Architectures (SOA),

*Email: rjimenez@fi.upm.es
†Email: nemo@lsub.org

1 Project Goals

The goals of the project:

- Support for dependable and adaptive services.
- Support for transactional services.
- Support for service composition.
- Ubiquitous service architecture.
- Identification and development of services for ubiquitous applications.
- Modeling of user interface.
- Development of a prototype for a smart space.

The Pert chart can be found at the end.

2 Progress Report

The LSD has achieved most of the objectives of its subproject. Regarding database replication it has overcome two out of the three bottlenecks of the scalability of current approaches. It has overcome the bottleneck introduced by serializability by defining a new correctness criterion, 1-copy snapshot isolation (1CSI). It has proposed and implemented protocols that satisfy this criterion (this has been formally proven). It has overcome the bottleneck introduced by full replication resorting to a novel approach to partial replication based on 1CSI. Currently, LSD is working in overcoming the last bottleneck that lies in the communication overhead. For this purpose Myrinet technology is being used to offload part of this inherent overhead to the network interface card. Regarding adaptation LSD has obtained protocols for self-optimization and self-healing with excellent performance results.

LSD has also obtained very good results for application server replication. In this area, LSD has proposed a primary-backup protocol providing highly available transactions what is especially important for long-running transactions. It has also proposed new protocols based on 1CSI for vertical replication of application servers and databases. They have been implemented and evaluated demonstrating very good scalability. LSD is currently working in online recovery and partial replication protocols for replication of the application server.

Regarding the transactional support, LSD has completed engines for acid and long-running transactions. The latter is an open source reference implementation of the OASIS standard WS-CAF in which technical committee LSD has participated. It is available as an open source project led by LSD at ObjectWeb (Jass project).

Regarding service composition, LSD has built a web service orchestration engine based on the OASIS BPEL4WS standard. LSD is participating in the technical committee of the standard. LSD has licensed its pre-industrial prototype to Bull. It is now commercialized as BSOA Orchestra. The BPEL engine built by LSD is reflective and therefore extensible without having to modify it. This reflection has been used to distributed the orchestration load in a cluster, and its being used to add other non-functional aspects to the web service orchestration.

Most of the objectives of the subproject of the LSU have been already fulfilled. The researchers achieved even more than it was expected, as it is explained below. Only the last objective is still unfinished, although it is expected to be finished it by the end of this project.

As a result of the research of this project, a new software architecture to develop ubiquitous services has been designed. The architecture principles have been applied to the design of a new operating system, Plan B, which has qualified by several experts as excellent research when they have peer reviewed the communications and papers already published about both, the architecture and the operating system the LSU has built.

LSU researchers are proud of having been able to test the results of our own work, a research practice that is strongly recommended by the systems research community, because it enables to develop better, more robust and functional systems. In these two years LSU researchers have built and then tested on the field several versions of the Plan B operating system. This way, they have been able to use it for daily work at LSU offices in Universidad Rey Juan Carlos. Current version is fully functional and supports a smart space built to provide the members of the laboratory with software and hardware infrastructure for their research, as well as for carrying out their daily tasks.

The resulting pervasive computing environment that Plan B supports is highly adaptable, and dynamically responds to changes in the resource availability in the environment, thus fulfilling the goals of the project. Specific services, including context awareness and location management, pervasive user interface support, voice and multimodal interfaces, among others, have been designed and implemented. All these are completely functional in the installed infrastructure in our lab. LSU researchers have also been able to use the system to provide support for lectures and student advising.

Regarding the last objective initially considered for the project, employing the infrastructure for people with disabilities, it is still ongoing. Severe delays in the partners responsible for setting up a domestic environment with domotic elements (partners who are not part of the research team for the project) made us reconsider the testbed of our research. As a result of this, intelligent spaces for systems researchers and programmers was chosen as testbed for our research, instead of intelligent spaces in assistencial environments. In this respect, the system resulting from the project provides an excellent pervasive computing system where all devices and resources are made available to the researchers and programmers within the LSU lab. There are several impressive demonstrations at http://lsub.org.

In any case, there is still time within the project time frame and there is hope the LSU will be able to apply all the research results to the domotic home-like environment finally built by our partners (it was made available one month ago).

Among the main results from the project, these should be mentioned (all of them implemented and being used):

- An operating system supporting pervasive computing environments
- User Interface Management System capable of combining multiple devices without application intervention.
- User Interface Technology capable of splitting, replicating, and migrating user interfaces (even individual components or parts of them) among different devices, without application support.

- Simple context management, yet capable of doing anything done by any other context management toolkit known to the authors.
- Adaptation technology to integrate resources from the environment into the user space.
- Human-Centered Authentication Architecture that enables a real Single Sign-On and provides secure resource sharing between users at smart spaces.
- Example applications.

3 Excellence Indicators

3.1 Publications

- J. Salas and R. Jiménez-Peris and M. Patiño-Martínez. Exploiting Reflection to Enable Scalable and Performant Database Replication at the Middleware Level. Chapter of the book. Software Engineering and Fault Tolerance. P. Pelliccione, H. Muccini, N. Guelfi, A. Romanovsky (eds). Kluwer. 2007. In Press.
- R. Jiménez-Peris, M. Patiño-Martínez, B. Kemme. Enterprise Grids: Challenges Ahead. Special issue on "Data management in Grid environments" Journal of Grid Computing. Springer. 2007. In Press.
- F. Perez, J. Vuckovic, M. Patiño-Martínez, R. Jiménez-Peris. Highly Available Long Running Transactions and Activities for J2EE Applications. 26th IEEE Int. Conf. on Distributed Computing Systems (ICDCS), Lisboa, Portugal, July 2006.
- J. Salas, F. Perez, M. Patiño-Martínez, R. Jiménez-Peris. WS-Replication: A Framework for Highly Available Web Services. ACM WWW Conf., Edinburgh, Scotland, May 2006.
- E. Jiménez, S. Arévalo, A. Fernández. Implementing unreliable failure detectors with unknown membership. Information Processing Letters. Vol. 100. n. 2. Pp. 60-63. Oct. 2006 Elsevier.
- A. Fernández, E. Jiménez, M. Raynal. Eventual Leader Election with Weak Assumptions on Initial Knowledge, Communication Reliability, and Synchrony. IEEE Int. Symp. on Dependable Systems and Networks (DSN). pp. 166-178. 2006.
- J. Salas, R. Jiménez-Peris, M. Patiño-Martínez, B. Kemme. Lightweight Reflection for Middleware-based Database Replication. IEEE Int. Symp. on Reliable Distributed Systems (SRDS), Leeds, England, Oct. 2006.
- A. Fernández, E. Jiménez, S. Arévalo. Minimal System Conditions to Implement Unreliable Failure Detectors. 12th IEEE Int. Symp. Pacific Rim Dependable Computing (PRDC), California, EEUU. Dec. 2006.
- R. Jiménez-Peris, M. Patiño-Martínez, B. Kemme. Enterprise Grids: Challenges Ahead. Invited paper corresponding to the keynote speech at Int. Workshop on High-Performance Data Management in Grid Environments co-located with VECPAR 7th Int. Conf. on High Performance Computing for Computational Science, July 2006.

- Y. Lin, B. Kemme, M. Patiño-Martínez, R. Jiménez-Peris. Applying database replication to multi-player online games. ACM SIGCOM Workshop on Network and System Support for Games (Netgames). Singapore, Oct. 2006.
- Tonghong Li. A Mobile Ad Hoc Routing Protocol for Metropolitan Bus Networks. The 6th Int. Conf. on ITS Telecommunications (ITST06).
- Tonghong Li, W. Seah. Multi-Path Routing Protocol for Underwater Sensor Network, Poster at Seventh ACM International Symposium on Mobile Ad Hoc Networking and Computing (mobihoc 2006).
- V. Martins, E. Pacitti, R. Jiménez-Peris, P. Valduriez: Scalable and Available Reconciliation on P2P Networks. BDA 2006
- Y. Lin, B. Kemme, M. Patiño-Martínez, R. Jiménez-Peris. Middleware based Data Replication providing Snapshot Isolation. ACM SIGMOD Int. Conf. on Management of Data. Baltimore, Maryland, USA. Jun. 2005.
- M. Patino-Martinez, R. Jimenez-Peris, B. Kemme, G. Alonso. Consistent Database Replication at the Middleware Level. ACM Transactions on Computer Systems (TOCS). n. 4, v. 23, pp. 375-423, Nov. 2005.
- R. Baldoni, R. Jiménez-Peris, M. Patiño-Martínez, L. Querzoni and A. Virgillito. Dynamic Quorums for DHT-based P2P Networks. 4th IEEE Int. Symp. on Network Computing and Applications (NCA). Cambridge, MA, USA. July 2005.
- A. Bartoli, R. Jimenez-Peris, B. Kemme, et al. The Adapt Framework for Adaptable and Composable Web Services. IEEE Distributed Systems On Line, September 2005.
- Y. Lin, B. Kemme, M. Patiño-Martínez, R. Jiménez-Peris. Consistent Data Replication: Is it feasible in WANs? Europar Conf. (EUROPAR), Lisbon, Portugal, 2005.
- A. Fernández, E. Jiménez, S. Arévalo. Minimal system conditions to implement unreliable failure detectors. Brief announcement. pp. 207. ACM Int. Conf. on Principles of Distributed Computing (PODC). 2005.
- B. Kemme, R. Jimenez, M. Patiño, J. Salas. Exactly once interaction in a multi-tier architecture. VLDB Conf. Trondheim, Norway. Aug. 2005. Workshop on design, implementation, and deployment of database replication. .
- J. Salas, F. Perez, M. Patiño-Martínez, R. Jiménez-Peris. Towards Replication of Web Services in WANs (Fast abstract). IEEE Int. Conf. on Dependable Systems and Networks (DSN), Yokohama, Japan, Jun. 2005.
- A. Bartoli, R. Jiménez-Peris, B. Kemme et al. Adapt: Towards Autonomic Web Services. European Dependable Computing Conference (EDCC), Budapest, Hungary, April 2005.
- A. Martinez, F. Pérez, M. Patiño-Martínez, R. Jiménez-Peris. ZenFlow: A Visual Tool for Web Service Composition. IEEE Symp. on Visual Languages and Human-Centric Computing. Dallas, Texas. Sept. 2005.

LSD is the first Spanish research group to publish in ACM TOCS in its 30 years of history, the journal with highest impact according citeseer (http://citeseer.ist.psu.edu/impact.html). The LSD is the second Spanish research group to publish in ACM SIGMOD (acceptance rate 15% in 2005) the premier database conference (also the 1st database conf. according citeseer). LSD has also published in ACM WWW (acceptance rate 11% in 2006), IEEE ICDCS (accept. rate 13% in 2006), IEEE DSN (accept. rate 18% in 2006), ACM/IFIP/USENIX Middleware (accept. rate 13% in 2004), the top conferences in web technologies, distributed systems, dependability, and middleware, respectively.

- F. J Ballesteros, G. Guardiola, K. Leal, E. Soriano. Omero: Ubiquitous User Interfaces in the Plan B Oper. Syst. IEEE PerCom 2006.
- F. J. Ballesteros, E. Soriano, K. Leal, G. Guardiola. Plan B: An Operating System for Ubiquitous Comp. Environ. IEEE PerCom 2006.
- FJ Ballesteros, E Soriano, G Guardiola, K Leal. Plan B: Using Files instead of Middleware Abstractions for Pervasive Comp. Environm. IEEE Pervasive Computing. 2007.
- F. J. Ballesteros, G. Guardiola, E. Soriano, K. Leal. Traditional Systems can Work Well for Pervasive Applications. IEEE Percom 2005.
- FJ Ballesteros, E Soriano, G Guardiola, K Leal. The Plan B OS for Ubiquitous Comput. Voice, Security, and Terminals as Case Studies. Pervasive and Mobile Computing Journal. 2006. Vol 2. N. 4.
- K. Leal, F. J. Ballesteros, E. Soriano, G. Guardiola. UbiTerm: A hand-held control-center for user's activity mobility. IEEE Int. Conf. on Pervasive Services 2005.**?**
- F. J. Ballesteros, E. Soriano. Give Me Back My Personal Mainframe. Lessons Learned from Plan B and Position Statement. Int. Workshop on Plan 9. Madrid, 2006.
- E. Soriano, F. J. Ballesteros, G. Guardiola. SHAD: A Human Centered Security Architecture for the Plan B Operating System. IEEE PerCom 2007
- E. Soriano, F. J. Ballesteros, G. Guardiola. A P2P Single Sign-On which remains Single on Smart Spaces. Int. Workshop on Ubiquitous Computing and Ambient Intelligence.
- G. Guardiola. CUROCO: a distributed architecture for the dynamic generation, composition and use of context in highly dynamic and heterogeneous environments. Doctoral Symp., ACM/USENIX Middleware 2004 Workshops, Toronto (Canada), October 2004.
- E. Soriano. SHAD: A Human Centered Security Architecture for Partitionable, Dynamic and Heterogeneous Distributed Systems. (PhD Thesis proposal) Doctoral Symposium, ACM/USENIX Middleware 2004 Workshops, Toronto (Canada), October 2004.
- G. Guardiola, K. Leal. A Mobile and ubiquitous system for the new millennium. Doing new things with ancient technology from the 70s!. Demonstration at WMCSA 2004.
- F. J. Ballesteros, E. M. Castro, G. Guardiola, K. Leal, and P. de las Heras. /net. A Network Abstraction for Mobile and Ubiquitous Computing Environments in the Plan B Operating System. WMCSA, 2004.

- FJ Ballesteros, G. Guardiola, K. Leal, E. Soriano, P. de las Heras, E. M. Castro, A. Leonardo, S. Arévalo. Plan B: Boxes for network resources. JBCS special issue on Adaptable computing systems., 2004.

- F.J. Ballesteros, E. Soriano, G. Guardiola. The LS smart space for programmers. Int. Workshop on Ubiquitous Computing and Ambient Intelligence 2006.

- G. Guardiola, R. Cox, E. Van Hensbergen. Persistent 9P Sessions for Plan 9. Workshop on Plan 9. 2006.

- E. Castro, T. de Miguel, S. Pavón. Transition of Applications to IPv6, Upgrade Magazine, vol VI, n.2, 2005.

- E. Castro, T. de Miguel, S. Pavón. Novática, n. 174, 2005.

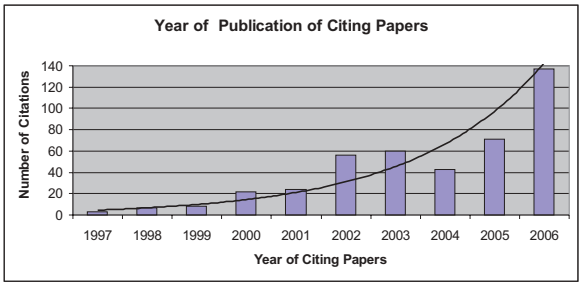
In 2005 previous work to Plan B was published in the IEEE Int. Conf. on Pervasive Computing and Communications (PerCom 2005), whose acceptance rate was 16%. The same year a component of the system was introduced in the ICPS'05.**?** The Plan B operating system was introduced through two communications in PerCom'06. Its acceptance rate was 13%. The introduction of Plan B in this conference provoked great expectation and generated ample positive discussions while and after the sessions. The LSU got another paper accepted for Percom 2007, this time, acceptance rate was 12%.

Until now, the work done in the development of the Plan B operating system has generated several publications in international journals of the first ranks. The IEEE Pervasive Computing (ISSN 1536-1268) journal will publish a paper describing the whole Plan B operating system. This journal has an impact factor of 2.036 in the JCR of 2005, and achieves the 5th position in the Journal Citation Report, in the category of telecommunications. The journal Pervasive and Mobile Computing will publish another paper describing the system through several case studies. This journal was created on 2005, so there is no information about its impact.

As part of the dissemination of results, the LSU organized IW9P, a peer reviewed workshop where an impressive group of more than 40 researchers met in Univ. Rey Juan Carlos, placing in a single room much of the heritage of UNIX and Plan 9 operating systems. Researchers from top research institutions such as IBM Research at Austin (USA), Bell Labs at Murray Hill (USA), MIT (USA), Carnegie Mellon (USA), York University (UK), and Twente University (The Netherlands), among others, visited the LSU group. The results of this project were presented at length in the workshop. Apart of the communications, two panels were dedicated to redesigning important protocols used in Plan B. All the materials presented by the LSU group and the rest of participants were published (http://lsub.org/iwp9).

3.2 Citations

The publications from LSD has over 450 citations (excluding self-citations from all co-authors) from which over 230 has been received in 2005-2006 and over 40% are citations to publications of the present project. Find below the graph of the citations of LSD publications (a complete list of citations of the lab co-directors can be found at http://lsd.ls.fi.upm.es/lsd/citations.html).



The LSU publications have 430 citations, and we expect more to come soon due to recently published results.

3.3 Technology Transfer

There have been some important technology transfer activities with the companies that expressed interest in the research of the project.

LSD developed a pre-industrial prototype for a web service orchestration environment that has been licensed to Bull in a technology transfer contract in 2005. Now it is available as a product commercialized by Bull named "BSOA Orchestra".

LSD has also signed a cooperation agreement with Telefonica Soluciones to help them to develop a highly scalable middleware for a service-oriented data warehousing. The cooperation agreement was signed in 2005 and the first component of the product has been sold to Telefonica. A new contract is being signed to sell them a second component of the product.

Sun microsystems researchers have been visiting the LSU to transfer technology from Plan B in the area of User Interface technology. They are now applying Plan B technology to their Java user interface applications.

LSU is currently initiating a collaboration with Synergex, a multinational company based on Canada which is interested in integrating the pervasive technologies implemented in Plan B in their RFID-related business.

3.4 Participation in International Projects

- Adapt: Middleware Technologies for Adaptive and Composable Distributed Components. European Programme Framework (IST-2001-37126). Primary investigator: Ricardo Jiménez-Peris. **Global coordinator of the project**. Sept. 2002-Oct. 2005.

- S4ALL. Services for All. European EUREKA/ITEA project (ITEA Label 04025). MITyC. Period: 2005-2007. Primary Investigator: Prof. Marta Patiño-Martínez.

- SEMEASY: Semantics makes middleware easy. EUREKA/ITEA project (Label 05016). 2006-2007. Primary Investigator: Prof. Ricardo Jiménez-Peris.

- AutoMan: Autonomic Management of Grid-Based Enterprise Services. ARC - Action de Recherche Coopérative. Funding Agency: INRIA. Period: 2006-2007. Primary Investigator: Prof. Marta Patiño-Martínez.

- High Performance Database Replication exploiting Storage Area Networks. Microsoft European PhD 2005. Funding Programme: EuroPhD programme 2005. Funding Agency: Microsoft Research Cambridge. 2005-2008. Primary Investigator: Prof. Marta Patiño-Martínez.

- Support for application consistency in P2P networks. Funding Programme: Binational Italian-Spanish Integrated Actions. Funding Agency: Spanish Ministry of Science and Education MEC (HI2003-0036). 2004-2005. Primary Investigators: Prof. Ricardo Jiménez-Peris (Spanish team) and prof. Roberto Baldoni (Italian team).

- Large scale data streaming. Funding Programme: EuroPhD programme 2006. Funding Agency: Microsoft Research Cambridge. 2006-2009. Primary Investigator: Ricardo Jiménez-Peris.

- Cabernet: Network of Excellence on Dependable and Distributed Systems. Funding Programme: European Framework Programme. Funding Agency: European Commission (IST-2000-25088). 2001-2004.

3.5 Participation in National Projects

- AUTONOMIC: Autonomic, Dependable and High Performance Distributed Systems. Funding Programme: Regional Networks of Excellence (area of information and communication technologies TIC, only group funded in Distributed and High Performance Computing). Funding Agency: Madrid Research Council CAM (S-0505/TIC/000285). 2006-2009. Primary Investigator (UPM) and Project Coordinator: Prof. Ricardo Jiménez-Peris. Primary Investigator (URJC): Francisco J. Ballesteros.

- Towards an open management architecture for services (TOMAS). Funding Programme: PROFIT (tractores call). Funding Agency: MyTIC (FIT-340001-2005-8). 2005-2006. Primary Investigator: Prof. Marta Patiño-Martínez.

- OpenCities: eGovernment Platform oriented to processes, documents and services. Funding Programme: PROFIT (open source call). Funding Agency: MyTIC (FIT-350110-2005-107). 2005-2006. Primary Investigator: Prof. Marta Patiño-Martínez.

3.6 Collaboration with Other Research Groups

- Prof. Bettina Kemme. McGill Univ, Montreal, Canada. Database replication.

- Prof. Gustavo Alonso. ETHZ, Zurich. Switzerland. Database replication.

- Prof. Roberto Baldoni. Univ. La Sapienza di Roma, Roma, Italy. P2P systems.

- Prof. Peter Popov. City Univ, London, UK. Autonomic Replication.

- Wolfgang Gerteis. SAP Research, Belfast, UK. High Performance Application servers.

- Prof. Patrick Valduriez. INRIA. LINA. Nantes, France. Data streaming.

- IBM Research at Austin. Dr. Eric Van Hensvergen. High Performance Computing

- Spyros Lalis. University of Thessaly. Pervasive Computing
- Xavier Alaman. Universidad Autónoma de Madrid. Pervasive Computing
- Jose Bravo. Universidad de Castilla La Mancha. Pervasive Computing.

3.7 Patents

LSD is currently patenting results on application server replication.

3.8 Training: Completed PhDs

- Ernesto Jimenez. PhD in 2004.
- Enrique Soriano. PhD in 2006
- Leonardo Querzoni. PhD in cotutela. 2007.
- Jesús Milán. PhD in 2007.

3.9 Training: PhDs under Supervision

At LSD: Francisco Pérez, Damián Serrano, Jorge Salas, José Antonio Sánchez, Andrea Ricci.
At LSU: Gorka Guardiola Muzquiz, Juan Cespedes

3.10 PhD and Master Internships

- Leonardo Querzoni, PhD student (Università La Sapienza di Roma). 2005-2006.
- Valentina Cristina del Franco, Master Student (Università di Napoli). 2005-2006.
- Jaksa Vuckovic, PhD student (Università di Bologna). 2005.
- Nicolas Lefebvre, Master student (ETH Zurich). 2005.
- Cayse Llorens, Master student (Univ. of Illinois at Urbana-Champaign). 2004.
- Gorka Guardiola Muzquiz. IBM Research at Austin. 2005 and 2006.

3.11 Awards

- Microsoft Research Cambridge. European PhD Award. 2005. Prof. Marta Patiño. This awards funds the PhD of Mr. Damián Serrano.
- Accenture. Best master thesis award. 2005. Business process coordination in web service environments. Implementation of the Business Activity standard. Iván García Muñoz (as student author of the master thesis) Prof. Marta Patiño (as supervisor of the thesis).
- Microsoft Research Cambridge. European PhD Award. 2006. Prof. Ricardo Jiménez. This award funds the PhD of Mr. Jorge Salas.

Pert Chart

Activities	Center	People involved	Year 1(*)	Year 2(*)	Year 3(*)	
Activity A1. DB and AS Replication	UPM					
A1.1 Partial Replication	UPM	RJP, BK, JMF	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ● ● ● ● ● ●	● ● ● ● ● ● ● ● ● ● ● ●
A1.2 Adaptability	UPM	MPM, AB,JMF	● ● ● ● ● ● ● ● ● ● ● ●	● ● ● ● ● ● ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
A1.3 WAN Replication.	UPM	BK,MPM,LHA, EJM,TEC	● ● ● ● ● ● ● ● ● ● ● ●	● ● ● ● ● ● ● ● ● ● ● ●	● ● ● ● ● ● ● ● ● ● ● ●
A1.4 Myrinet.	UPM	AB,RJP,EJM, LHA,TEC	● ● ● ● ● ● ● ● ● ● ● ●	● ● ● ● ● ● ● ● ● ● ● ●	● ● ● ● ● ● ● ● ● ● ● ●
Activity A2. Transactional support	UPM					
A2.1 Transactional ACID engine	UPM	RJP, FPS, DJP	● ● ● ● ● ● ● ● ● ● ● ●	● ● ● ● ● ● ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
A2.2 Advanced Transactions Engine	UPM	MPM, FPS, DJP	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ● ● ● ● ● ●	● ● ● ● ● ● ● ● ● ● ● ●
Activity A3. Composition support	UPM					
A3.1 Workflow engine	UPM	MPM, FPS, EJM	● ● ● ● ● ● ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
A3.2 Dynamic adaptation	UPM	RJP, EJM, DCL	○ ○ ○ ○ ○ ○ ● ● ● ● ● ●	● ● ● ● ● ● ● ● ● ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
A3.3 Dependability	UPM	MPM, EJM, DCL	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ● ● ●	● ● ● ● ● ● ● ● ● ● ● ●
Activity A4. Case study	UPM, URJC					
A4.1 System Requirements	UPM, URJC	TL, FJBC	● ● ● ● ● ● ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
A4.2 Design & Implementation.	UPM, URJC	TL, EJM, FJBC	○ ○ ○ ○ ○ ○ ● ● ● ● ● ●	● ● ● ● ● ● ● ● ● ● ● ●	● ● ● ● ● ● ○ ○ ○ ○ ○ ○
A4.3 Performance Evaluation	UPM, URJC	TL, EJM, FJBC	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ● ● ● ● ● ●

TIN2004-07474-C02

Activities	Center	Person	Year 1(*)	Year 2(*)	Year 3 (*)
Activity B1. Pervasive serv. definition	URJC				
B1.1. Low level services	URJC	SAV, FJBC, MLA ALM, KLA, GGM	● ● ● ● ● ● ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
B1.2. System services	URJC	FJBC, PHQ, ECB	● ● ● ● ● ● ● ● ● ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
Activity B2. Service access prot.	URJC				
B2.1 Common resource manag. prot.	URJC	FJBC, ALM, SAV	○ ○ ○ ○ ○ ○ ● ● ● ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
B2.2 Resource discovery prot.	URJC	FJBC, KLA	○ ○ ○ ○ ○ ○ ● ● ● ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
B2.3 Notification services	URJC	MLA, ALM	○ ○ ○ ○ ○ ○ ○ ○ ○ ● ● ●	● ● ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
B2.4 Component integration model	URJC	FJBC, KLA	○ ○ ○ ○ ○ ○ ○ ○ ○ ● ● ●	● ● ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
Activity B3. Pervasive serv. devel.	URJC				
B3.1 Low level services devel.	URJC	MLA, ALM, TL	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ● ● ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
B3.2 Construcción de servicios de sistema	URJC	SAV, PHQ, ESS GGM, ECB	○ ○ ○ ○ ○ ○ ○ ○ ○ ● ● ●	● ● ● ● ● ● ● ● ● ● ● ●	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
Activity B4. Middleware	URJC				
B4.1 Naming	URJC	GGM	○ ○ ○ ○ ○ ○ ● ● ● ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
B4.2 Service access interface	URJC	FJBC, KLA	○ ○ ○ ○ ○ ○ ○ ○ ○ ● ● ●	● ● ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
B4.3 Selection engine	URJC	FJBC, KLA	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ● ● ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
B4.4 Adaptors	URJC	FJBC, MLA, ALM KLA	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ● ● ● ● ● ● ● ●	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
Activity B5. Case study	URJC, UPM				
B5.1 Design and implementation	UPM, URJC	FJBC, TODOS URJC	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	● ● ● ● ● ● ● ○ ○ ○ ○ ○
B5.2 Adaptation of the environment	UPM, URJC KLA, GGM, ESS TL, MPH, MBR, RMP	FJBC, MLA, ALM	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ● ● ● ● ● ● ●
B5.3 Terminal development	URJC	SAV, PHQ, ECB	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ● ● ● ● ● ● ●

TIN2004-07474-C02

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

El problema de la escalabilidad en visualizaciones complejas de la información y del *software*. TIN2004-07568

J. Ángel Velázquez Iturbide*
Departamento de Lenguajes y Sistemas Informáticos
Universidad Rey Juan Carlos

Abstract

A visualization is an abstract representation of an object that improves our understanding of it. The term information visualization is used for the emerging discipline that studies the design and implementation of computer-generated visualizations. One of the main problems of visualization, named the scalability problem, is given when the number of elements of compound information is so large that the visualization is unable to represent it adequately. The main aim of the project consists in developing techniques to effectively visualize information of large size. The techniques will be developed for and applied to three domains: functional programs, object-oriented programs and students' activity in e-learning platforms. Each domain deserves a different treatment, since the scalability problem has received different degrees of attention and the effort necessary to construct visualizations also greatly varies.
Keywords: information visualization, software visualization, scaling problem, effortlessness, functional programming, object-oriented programming, e-learning platforms.

1 Aims of the Project

A visualization is an abstract representation of an object that improves our understanding of it. The term information visualization is used for the emerging discipline that studies the design and implementation of computer-generated visualizations. One of the main problems of visualization, named the scalability problem, is given when the number of elements of compound information is so large that the visualization is unable to represent it adequately.

The main aim of the project consists in developing techniques to effectively visualize large-scale information. We seek domain-independent techniques and techniques for three specific domains: functional programming, object-oriented (OO) programming, and e-learning platforms.

The proposal is grounded on experience of the applicant team, mainly in program visualization and educational applications. Previous developments that were substantially used for the project are the functional programming environment WinHIPE and a web-based e-learning platform.

The project proposes to enhance visualization techniques mainly by means of three approaches. Firstly, the combined use of different techniques will have a synergic effect. We propose to use, at least, the multiple views, navigation, and simplification techniques. Secondly,

* Email: angel.velazquez@urjc.es

porting a technique from a given domain to another one will require enhancing it. Thirdly, new visualization techniques will probably be necessary. A key element in the three approaches is effectiveness evaluation by means of controlled experiments.

Each domain has specific features that must be taken into account. For instance, software is often visualized dynamically (as an animation) in educational contexts. Another important difference lies in the role of the user. Information visualizations are completely generated by computer programs, so the user only interacts with them. However, software animation systems typically provide the infrastructure to the teacher, who must build the animation with great effort.

The initial scheduling of the project is shown in Table 1. For the sake of brevity, we omit tasks concerning mundane activities (task A) or bibliographic study (subtasks *.1). Tasks correspond to the domains of study: functional programming (task A), OO programming (task B) and e-learning platforms (task C). Roughly, all of the tasks consist in a sequence of design, implementation and tuning of a visualization model. Task B contains an intermediate subtask (B.3) to implement a language processor. Task D contains twice such a sequence: first for multiple, partial visualizations of students' activity, and then for a comprehensive visualization model. There is no specific task to develop general visualization techniques, but they will be a side effect of research in the 3 domains.

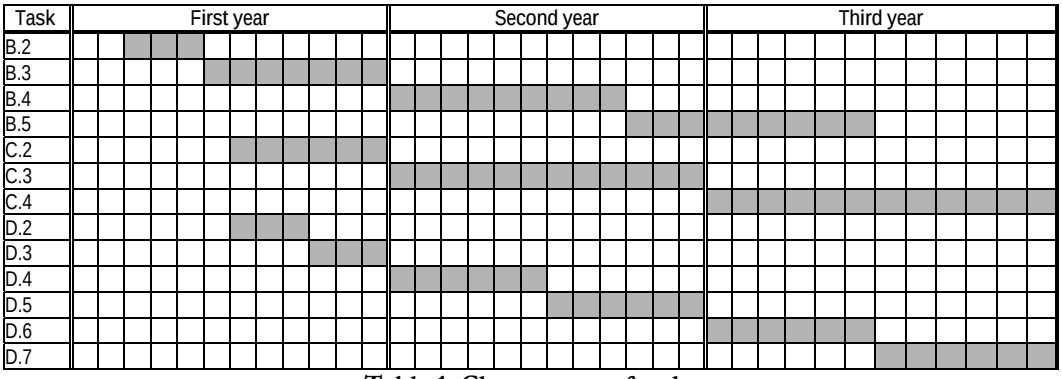


Table 1. Chronogram of tasks

2 Technical Results

2.1 Key Decisions

Task B was altered so that a new functional programming processor was not developed: we considered more productive to make use of a processor already available by the group. Task C also was modified so that a single approach was not addressed, but several exploratory works were performed. Task D is being performed pretty similar to the initial schedule. Despite these deviations from the initial schedule, contributions obtained so far justify these decisions.

With respect to visualization of programs, we focused in educational contexts. As explained above, a major source of difficulty in education is the effort required to generate animations. This workload is posed on educators. Consequently, our research efforts on software visualization are performed within effortless frameworks, typically based on program visualization.

2.2 Visualization of Functional Programs

Objectives Fulfilled to Date

In summary, we extended the WinHIPE functional programming environment to address in depth the effort and the scaling issues.

We refined the effortless approach of WinHIPE to build discrete animations of the evaluation of functional expressions [35]. Further support was given to customize and maintain web animations [26][27]. Extensions to integrate tests also were considered [24]. The approach was assessed with students of a programming languages course. Statistical evidence was obtained about the fact that students who construct an animation perform better than those just watching it [28].

The scale problem was addressed in three aspects: size of an expression, number of expressions, and number of animations.

An intermediate expression can be too large to fit a window, therefore being difficult to understand it. We designed a technique to filter the parts of an expression that are less interesting. It is a logical fisheye view [32] based on five design requirements.

The selection of the visualizations that form an animation is difficult when their number is large, even if they are displayed as miniatures, because they do not fit a window. We designed a focus+context technique to deal with miniatures, called r-zoom [25]. The technique satisfies specific requirements of program visualizations and provides enhanced user interaction. It was empirically evaluated and compared to an overview+detail interface [30]. Experienced users of r-zoom completed tasks faster and with fewer errors than users of the overview+detail interface. Furthermore, users felt more comfortable with r-zoom.

When an educator develops animations, she will probably want to reuse them for different courses. WinHIPE allows constructing and maintaining hierarchical collections of web animations [26][27]. The user may handle the hierarchy, the animations and their look. A collection is visualized as a tree of folders with expansion/contraction facilities.

Current and Future Work

We are working on four issues. Firstly, we made a long-term evaluation of the educational effectiveness of WinHIPE during two months. We still have to statistically analyze collected data. Secondly, we are developing a pedagogical guide for educators to design animations. Thirdly, a general API for r-zoom, amenable to be used by any Java container and visualizer, is under development. Finally, we plan to evaluate the facilities for collections of web animations.

2.3 Visualization of Object-Oriented and Imperative Programs

Objectives Fulfilled to Date

We conducted different exploratory works on visualization of object-oriented and imperative programs. The goal was to explore several domains, to identify the role of visualization in them, and to study the importance of the scaling problem. The domains are: OO debugging, tutors and assistants on OO programming, symbol tables, and algorithm design techniques.

We identified the limitations of line-oriented debuggers for the practitioner programmer and developed a structural debugger [5]. It solved those problems with structural actions interactively invoked with a graphical interface. It is based on JavaMod, an API developed by us.

We addressed the design of a series of tutors for Java OO concepts. They are designed within the framework provided by Bloom's taxonomy [8]. Three tutors, aimed at the 3 lowest levels of the taxonomy, were developed. Firstly, we developed a tutor on inheritance [6][7]. It contained components for theory, Flash animations and adaptive tests. The tutor was evaluated, yielding

enhanced learning at the comprehension level and students' positive attitude. Other tutors [9] were: GeCom, that automatically generates comments in programs, and CreOO, that randomly generates and corrects problems on object creation.

We also developed TextOO, an assistant aimed at the application and upper levels of Bloom's taxonomy [13][14]. It assists in developing class, object and interaction UML diagrams and linking them to parts of a natural language specification. The assistant also corrects students' diagrams.

Visualization of other kinds of imperative programs was also addressed. SOTA is a tool aimed to graphically demonstrate the behaviour of symbol tables [2][3]. Its educational effectiveness was empirically evaluated [4][29]. Students using SOTA performed the same as students not using it, but in less time. They also had a positive attitude towards using SOTA for self-study.

A final line of research is visualization of algorithms based on design techniques. We performed a comprehensive study on visualizations for four algorithm design techniques [1]. In order to implement their visualization, we developed a generic framework.

Current and Future Work

The work described so far continues in several lines. Two works are in progress to explore the extension of our tutors to other programming domains: programming syntax and greedy algorithms. The TextOO assistant also is being enhanced to better assess UML diagrams. Two works on SOTA are in progress: a new evaluation of its educational effectiveness, and a new version to deal with any source language; most importantly, this new version will include scaling facilities. With respect to the visualization of algorithm design techniques, we are currently extending the generic framework to visualize recursion using multiple views.

2.4 Visualization of Students' Activity in e-Learning Platforms

Objectives Fulfilled to Date

Our first contribution to visualization in e-learning platforms was focused on timetable and exam schedule [33]. Tabular representations were designed, where collisions were detected graphically.

Next efforts were devoted to visualize students' activity. We followed two approaches. For the first one, we used an e-learning platform developed by our team (http://trancos.escet.urjc.es/urjc). We identified issues that educators are typically interested in about their students' activity, we studied 10 tools for analyzing web sites and we concluded that they fail in tracking individual behaviour [10]. Consequently, we developed two extensions for our platform, respectively based on a database of students' actions and on cookies. A comparison of functionality provided by commercial analysis tools, our prototype and surveys was conducted [11], as well as a comparison with other e-learning platforms [12].

A different approach was concurrently used to analyze students' activity of an on-line course offered within the ADA-Madrid project (http://www.upm.es/adamadrid), where a different e-learning platform was used. Data collected by the platform were used for exploratory data analysis and data mining [21][22]. Surprising conclusions were obtained about university registration procedures, students' skills in computer usage, interaction patterns, descents in performance, students' evaluations, or self-confidence.

Current and Future Work

The two platform extensions are collecting access data from 6 class groups. In February, we will be able to analyze and compare data obtained by both tools. A questionnaire will also be submitted to instructors. Based on our experiences, we will design a more comprehensive visualization model of students' interaction. However, evaluation will be done after the end of the project.

2.5 Other Contributions

We delivered several states of the art [16][18][19] and WG reports [17][20] on visualization. Other contributions were produced as a side effect of research on visualization. Animations of functional programs were used in a freshmen course on programming languages [31]. We restated algorithm design by transformations [15]. An on-line course on visualization was offered [21][23].

3 Indicators of Success

3.1 Personnel in Training

Three PhD Thesis are being conducted. The first one is finished and will be presented in 2007:

- Jaime Urquiza. "Generación semiautomática de animaciones de programas funcionales con fines educativos".
- Isidoro Hernán. "Taxonomía de Bloom como marco de diseño de tutores programación".
- Raquel Hijón. "Análisis visual de actividad de los alumnos en plataformas de e-learning".

The project has also been a framework for research grants and works:

- Micael Gallego, URJC doctoral grant.
- Antonio Pérez, project grant.
- Research projects of graduate students (Luis Fernández, Raquel Hijón, David Redondo).
- Capstone projects of undergraduate students (11 projects).

3.2 Publications

The project has produced so far 35 publications. They are here summarized by categories. The complete list of these publications can be found in the references section.

Edited books	**1**
Book chapters	**4**
Articles in international journals or magazines	**4**
Papers in international conferences	**17**
Papers in international workshops	**1**
Posters and sw demos in international conferences	**3**
Papers in national conferences	**2**
Papers in national workshops	**2**
Technical reports	**1**

3.3 System Implementations

We have delivered several applications, already mentioned above:

- Extensions of the WinHIPE IDE to generate and manage animations.
- The JavaMod API to model Java programs and a Java structural debugger.
- Three OOP tutors (anonymous tutor on inheritance, GeCom and CreOO).

- The TextOO assistant to develop UML designs based on natural language specifications.
- The SOTA system to visualize symbol tables.
- A framework to visualize algorithm design techniques and a prototype for recursion.
- A web application to schedule timetables and exams.
- Two extensions of our e-learning platform (based on databases and cookies).

3.4 Collaboration with National and Foreign Groups

Our research team has collaborated with several groups. Firstly, we participated in working groups of the *ACM Annual Conference on Innovation and Technology in Computer Science Education* (*ITiCSE*). In particular, Jaime Urquiza participated in a group at *ITiCSE 2005*, Lisbon, and Jaime Urquiza and Ángel Velázquez participated at *ITiCSE 2006*, Bologna. Jaime Urquiza participated at the 2006 GI-Dagstuhl-Seminar "Human-Centered Visualization Environments" held in Wadern, Germany.

Secondly, visits were made to other universities. Jaime Urquiza visited the EdTech group at Joensuu University, Finland, from July to August 2006 (2 months). Maximiliano Paredes visited the CHICO group at the Universidad de Castilla – La Mancha from June to September 2006 (3 months), as well as the Universidad Privada Antenor Orrego, Peru, from November to December 2006 (one month). Manuel Rubio visited the COSMAL group at University of California, San Diego, on September 2006, and has been accepted to perform a 6-month research visit in 2007.

3.5 Other Activities

We organized a seminar in 2006 [34], co-supported by the URJC. A second edition is being organized for 2007. Manuel Rubio and Ángel Velázquez received the Best Poster Award at the *I Jornada de Innovación Pedagógica del Proyecto ADA-Madrid* for their analysis of students' activity [22][23].

4 References

[1] L. Fernández, Á. Velázquez. "Estudio sobre la visualización de las técnicas de diseño de algoritmos". In *VII Congreso Internacional de Interacción Persona-Ordenador*, M.Á. Redondo, C. Bravo, M. Ortega (eds.), 2006 (ISBN 84-690-1613-X), 315-324.

[2] M. Gallego, F. Gortázar, J. Urquiza, Á. Velázquez. "SOTA, una herramienta educativa para la enseñanza de la tabla de símbolos". In *XI Jornadas de Enseñanza Universitaria de la Informática*, Thomson, 2005 (ISBN 84-9732-421-8), 339-345.

[3] M. Gallego, F. Gortázar, J. Urquiza, Á. Velázquez. "SOTA: a visualization tool for symbol tables". In *10th Annual Conference on Innovation and Technology in Computer Science Education*, ACM Press, 2005 (ISBN 1-59593-024-8), 385. Reprinted in *ACM SIGCSE Bulletin*, 37(3):385, 2005 (ISSN 0097-8418).

[4] M. Gallego, F. Gortázar, J. Urquiza, Á. Velázquez. "Experimental report – Evaluation of SOTA, a symbol table visualization tool". Universidad Rey Juan Carlos, Working Paper 2006/11, 2006 (ISBN 84/689-8795-6), 30 pp.

[5] M. Gallego, F. Gortázar, Á. Velázquez. "Depuración estructural: acercando la práctica a la teoría de la programación". In *VI Simposio Internacional de Informática Educativa*, J.M. Sánchez, J.A. Gómez, M.Á. Vega, B. Fernández, J. Bravo (eds.), Universidad de Extremadura, 2004 (ISBN 84-7723-654-4), 8 pp.

[6] I. Hernán, C. Lázaro, Á. Velázquez. "Una aplicación educativa basada en la jerarquía de Bloom para el aprendizaje de la herencia de POO". In *VII Simpósio Internacional de Informática Educativa*, A. Mendes, I. Pereira, R. Costa (eds.), E.S.E. de Leiria, 2005 (ISBN 972-95207-4-7), 107-112.

[7] I. Hernán, C. Lázaro, Á. Velázquez. "An educative application based on Bloom's taxonomy for the learning of inheritance in oriented-object programming". In *Computers and Education: Towards Educational Change and Innovation*, A. Mendes, I. Pereira, R. Costa (eds.), Springer-Science, in press.

[8] I. Hernán, Á. Velázquez, C. Lázaro. "Programming learning tools based on Bloom's taxonomy: Proposal and accomplishments". In *8th International Symposium on Computers in Education (Vol. 1)*, L. Panizo, L. Sánchez, B. Fernández, M. Llamas (eds.), Universidad de León, 2006 (ISBN 84-9773-301-0), 343-351.

[9] I. Hernán, Á. Velázquez, C. Lázaro. "Dos herramientas educativas para el aprendizaje de programación: generación de comentarios y creación de objetos". In *VII Congreso Internacional de Interacción Persona-Ordenador*, M.Á. Redondo, C. Bravo, M. Ortega (eds.), 2006 (ISBN 84-690-1613-X), 325-334.

[10] R. Hijón, Á. Velázquez. "Steps towards the analysis of students' activity in e-learning platforms". In *E-Learn 2005 – World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education*, G. Richards (ed.), AACE, 2005 (ISBN 1-880094-57-6), 738-743.

[11] R. Hijón, Á. Velázquez. "Web, log analysis and surveys for tracking university students". In *IADIS International Conference Applied Computing 2006*, N. Guimarães, P. Isaías, A. Goicoetxea (eds.), IADIS Press, 2006 (ISBN 972-8924-09-7), 561-564.

[12] R. Hijón, Á. Velázquez. "E-learning platforms analysis and development of students tracking". In *ED-MEDIA 2006 – World Conference on Educational Multimedia, Hypermedia & Telecommunications*, E. Pearson, P. Bohman (eds.), AACE, 2006 (ISBN 1-880094-59-2), 2,823-2,828.

[13] C. Lázaro, Á. Velázquez, I. Hernán, F. Gortázar, M. Gallego. "TextOO: Learning object oriented modeling using enunciates". In *Fifth Koli Calling Conference on Computer Science Education*, T. Slakoski, T. Mäntylä, M. Laakso (eds.), Turku Centre for Computer Science, TUCS General Publications Series nº 41:181-182, 2005 (ISBN 951-29-3006-4, ISSN 1239-1905).

[14] C. Lázaro, Á. Velázquez, R. Hijón, I. Hernán. "TextOO: An object-oriented learning tool based on enunciates". In *8th International Symposium on Computers in Education (Vol. 1)*, L. Panizo, L. Sánchez, B. Fernández, M. Llamas (eds.), Universidad de León, 2006 (ISBN 84-9773-301-0), 182-189.

[15] L. Llana, C. Pareja, Á. Velázquez. "Desarrollo de software con el modelo transformacional". *CNCIIC 2005: Avances en Tecnologías de la Información*, ANIEI (ISBN 970-31-0528-9), ch. 2, 6 pp.

[16] O. Kulyk, R. Kosara, J. Urquiza, I.H.C. Wassink. "Human-centered aspects". In *Human-Centered Visualization Environments*, A. Kerren, A. Ebert, J. Meyer (eds.), Springer-Verlag, LNCS Tutorials, in press.

[17] T. Naps, G. Roessling, P. Brusilovsky, J. English, D. Jarc, V. Karavirta, C. Leska, M. McNally, A. Moreno, R.J. Ross, J. Urquiza, "Development of XML-based tools to support user interaction with algorithm visualization", *ACM SIGCSE Bulletin*, 37(4):123-138, 2005 (ISSN 0097-8418).

[18] C. Pareja, Á. Velázquez. "Program execution and visualization on the Web". *Encyclopedia of Information Science and Technology*, M. Khosrow-Pour (ed.), Idea Group, 2005 (ISBN 1-59140-553-X, eISBN 1-59140-794-X), 2,330-2,334.

[19] C. Pareja, J. Urquiza, Á. Velázquez. "Web-based algorithm and program visualization for education". *Encyclopedia of Information Science and Technology*, M. Khosrow-Pour (ed.), 2nd ed., Idea Group, in press.

[20] G. Roessling, T. Naps, M.S. Hall, V. Karavirta, A. Kerren, C. Leska, A. Moreno, R. Oechsle, S. H. Rodger, J. Urquiza, Á. Velázquez, "Merging interactive visualizations with hypertextbooks and course management", *ACM SIGCSE Bulletin*, 38(4):166-181, 2006 (ISSN 0097-8418).

[21] M. Rubio, S. Alcalde, M. López, Á. Velázquez. "«Cómo mostrar visualmente datos y explicaciones»: análisis de interacción". In *II Jornada de Innovación Pedagógica del Proyecto ADA-Madrid*, J. Garrido, J.V.Conde, A.García, P. Shih Ma (eds.), 2006 (ISBN 84-690-2366-7), 67-72.

[22] M. Rubio, R. Hijón, F. Domínguez, Á. Velázquez. "How course management systems can benefit from exploratory analysis of student on-line activity". In *8th International Symposium on Computers in Education (Vol. 1)*, L. Panizo, L. Sánchez, B. Fernández, M. Llamas (eds.), Universidad de León, 2006 (ISBN 84-9773-301-0), 182-189.

[23] M. Rubio, Á. Velázquez. "Diseño y evaluación de la asignatura «Cómo mostrar visualmente datos y explicaciones»". In *I Jornada Pedagógica del Proyecto ADA-Madrid*, R. Criado, J.V. Conde (eds.), Universidad Politécnica de Madrid, 2005 (ISBN 689-4330-4), 119-125.

[24] J. Urquiza, R. Calvo, R. Hijón, Á. Velázquez. "Generation, management and publication of XML-based e-questionaries". In *IADIS International Conference Applied Computing 2006*, N. Guimarães, P. Isaías, A.Goicoetxea (eds.), IADIS Press, 2006 (ISBN 972-8924-09-7), 153-160.

[25] J. Urquiza, Á. Velázquez. "R-zoom: A visualization technique for algorithm animation construction". *IADIS Internacional Conference Applied Computing 2005*, N. Guimarães, P. Isaías (eds.), IADIS Press, 2005 (ISBN 972-99353-6-X), 145-152.

[26] J. Urquiza, Á. Velázquez. "Effortless construction and management of program animations on the Web". *Advances in Web-Based Learning – ICWL 2005*, R.W.H. Lau, Q. Li, R. Cheung, W. Liu (eds.), Springer-Verlag, LNCS 3583, 2005 (ISBN 3-540-27895-8, ISSN 0302-9743), 163-173.

[27] J. Urquiza, Á. Velázquez. "Reusable collections of Web-based program animations". *10th Annual Conference on Innovation and Technology in Computer Science Education*, ACM Press, 2005 (ISBN 1-59593-024-8), 386. Reprinted in *ACM SIGCSE Bulletin*, 37(3):386 (ISSN 0097-8418).

[28] J. Urquiza, Á. Velázquez. "An evaluation of the effortless approach to build algorithm animations with WinHIPE". In *Fourth Program Visualization Workshop*, G. Roessling (ed.), University of Florence, 2006, 29-33.

[29] J. Urquiza, Á. Velázquez, M. Gallego, F. Gortázar. "An evaluation of a symbol table visualization tool". In *8th International Symposium on Computers in Education (Vol. 1)*, L. Panizo, L. Sánchez, B. Fernández, M. Llamas (eds.), Universidad de León, 2006 (ISBN 84-9773-301-0), 198-205.

[30] J. Urquiza, Á. Velázquez, C. Lázaro. "Design and evaluation of r-zoom, a new focus+context visualization technique". In *VII Congreso Internacional de Interacción Persona-Ordenador*, M.Á. Redondo, C. Bravo, M. Ortega (eds.), 2006 (ISBN 84-690-1613-X), 79-88.

[31] Á. Velázquez. "A programming languages course for freshmen". In *10th Annual Conference on Innovation and Technology in Computer Science Education*, ACM Press, 2005 (ISBN 1-59593-024-8), 271-275. Reprinted in *ACM SIGCSE Bulletin*, 37(3):271-275, 2005 (ISSN 0097-8418).

[32] Á. Velázquez. "Principled design of logical fisheye views of functional expressions". *ACM SIGPLAN Notices*, 41(8):34-43, 2006 (ISSN 0362-1340).

[33] Á. Velázquez, A.Mª. Aguado, C. Rubio. "Aplicaciones web para gestión de la docencia universitaria: horarios, exámenes y proyectos de fin de carrera". In *VI Simposio Internacional de Informática Educativa*, J.M. Sánchez, J.A. Gómez, M.Á. Vega, B. Fernández, J. Bravo (eds.), Universidad de Extremadura, 2004 (ISBN 84-7723-654-4), 8 pp.

[34] Á. Velázquez, M. Paredes (eds.). *Tecnologías del Software*. Dyckinson, in press.

[35] Á. Velázquez, C. Pareja, J. Urquiza. "An approach to effortless construction of program animations". *Computers & Education* (ISSN 0360-1315), in press.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

RIMMEL: Recuperación de Información Multilingüe y Multimedia y su Evaluación TIN2004-07588-C03

José Miguel Goñi Menoyo
Project's Chief Investigator
Universidad Politécnica de Madrid
josemiguel.goni@upm.es

Antonio Moreno Sandoval
Principal Investigator for subproject 2
Universidad Autónoma de Madrid
antonio.msandoval@uam.es

Karmele Intxausti Goikoetxea
Principal Investigator for subproject 3
IKERLAN
kintxausti@ikerlan.es

Abstract

The object of the project is the development, integration, evaluation and application of a multilingual and multimedia Information Retrieval system, in a production environment. The system will be technologically competitive in international fora. The research groups that participate in this project proposal are working collaboratively for the last years in this research field. As a result of this collaboration, they have integrated some resources and tools developed previously in a separate way. The current challenges are to develop and to integrate resources and own technology specifically for multilingual and multimedia Information Retrieval, targeting to enhance performance, to cover a significant number of languages, to test different techniques for all these aspects, to evaluate them in an international context and to apply them in an industrial scenario developing a specific prototype. For this evaluation, we consider the participation of the team in the international evaluation fora, mainly CLEF (Cross Language Evaluation Forum), although it is not discarded to participate in other similar to it, such as TREC (Text REtrieval Conference), NTCIR (NII Test Collection for Information Retrieval), or those other that could substitute them.

This project is located inside the high-priority lines of the National Program of Computer Technologies, high-priority objectives 4.1, 4.2, 4.3, 3.5, 3.6 and 6.2, and of the VI Frame Programme of the European Union (priority 2, Information Society, key action 3, multimedia contents and tools).

Keywords: Information retrieval, trie data structure, pseudo relevance feedback, evaluation of IR systems, multilingualism, linguistic resources for IR,

1 Project Objectives

This project's main objective is the analysis of the current available methods and techniques in the field of Multilingual Information Retrieval and their application, especially to the case in which one of the involved languages is Spanish. In addition to that, contributing to the application of new techniques, the improvement of existing ones, and the hybridation or combination of all them are also objectives. These techniques will be evaluated in international evaluation fora. In addition to that, the project will be not only limited to documental objects, but also it will try to extend these techniques to the retrieval of multimedia objects, which are conveniently "described" by particular documental representations.

The availability of linguistic resources is a prerequisite for the development of linguistic engineering systems, whether they implement symbolic or statistical models. Their compilation is a costly task, and special abilities are needed, as well as a deep knowledge of languages. Ideally, linguistic resources should be developed by linguists with expertise. Multilingual IR has new challenges: to extend the number of supported languages, to include written and spoken language (that is, include some multimodality), and including different variants of Spanish, regarding the increasing importance of the Spanish in America in the Information Society. One of our objectives is the extension of existing systems to languages such as Arabic, Hindi and Japanese, at least to a basic level.

Finally, we plan to explore the usefulness of these systems in a real exploitation industrial scenario, that one of the Technical Support Services. In these environments, a huge amount of information is available, both textual –whether structured or not- and graphical –whether annoted or not-, that must be accessed in a simple, easy and quick way for maintenance technicians. In addition to that, the application domain is a very limited one, and this eases that semantic techniques and domain modelling (ontologies) can be also used to complement standard ones, and this can improve the retrieval results. So, metadata and Semantic Web modelling techniques are also planned to be included in our prototypes.

These general objectives will be implemented via these specific objectives:

- To develop a platform that integrates linguistic components and tools and resources. These elements can be existing ones or new-developed.
- To apply this platform to Multilingual and Multimedia Information Retrieval.
- To apply this platform for Information Extraction and the semi-automatic generation of ontologies, using techniques from Knowledge Engineering, for restricted domains. The selected domain will be an industrial real domain, that of maintenance services. A structured model will be developed to integrate heterogeneous available information.
- To evaluate the platform in usual international evaluation fora, mainly CLEF (Cross-Language Evaluation Forum) [3].
- To develop specific linguistic resources for languages such Arabic, Hindi and Japanese.
- To incorporate linguistic resources that allows us a better treatment of spontaneous spoken language.

In order to reach these objectives, the project team counts on several means: computational and communications infrastructure (partly acquired with funding coming from the project budget),

linguistic resources and software tools previously developed in the research groups (some of them as outcomes of other research projects):

- **Lexical resources:** The Lexical Database ARIES was developed in former research projects by part of the research team. This resource was inscribed in the Spanish Intellectual Property Registry in 1996, and it is under continuous enhancement, partly during the development of this project. The EuroWordNet [6] lexical resource was also previously available.
- **TRIELIB library:** As a joint outcome of the work of some researchers of the UPM team and from the company DAEDALUS-Data, Decisions and Language, S.A. [5], a university spin-off and EPO of this project, the TRIELIB library has been developed [7]. It is a software library for trie [2] management, oriented to lexical databases for lexical storage or index building. Its inscription in the Spanish Intellectual Property Registry has been requested on December 2006.
- **Lexical resources management tools:** A library of tools for lexical resources management based on the TRIELIB library is also available as a result of previous research. These tools permit access times independent of the size of the lexical database.
- **Morphological processing module:** The morphological model used in the ARIES lexical database is well adapted to the trie data structure used in the TRIELIB library. It allows us a very efficient processing of the morphemes grouped in continuation classes.
- **Specific tools:** Initial versions of specific tools for the different linguistic processes are available, such as stemmers, text segmenters, stopwords lists, and others. Some of them are obtained from other research groups or companies, being open-source software.
- S**pontaneous speech:** The results and the methodology of the European Project C-ORAL-ROM, a multilingual and multimedia corpus of spontaneous speech in Italian, French, Portuguese and Spanish, developed by the UAM team, is an outstanding experience for further exploitation and development of both corpora and tools for annotation and retrieval.

For the execution of the project the timing diagram shown in Appendix I was initially scheduled.

2 Success level carried out in the project

Semantic (restricted) domain modelling and use of ontologies for IR

The application of this technology in a technical domain (Technical Support Services) has allowed us the development of a domain ontology. This ontology uses sub models or views in order to structure the domain: (i) the different functional views or functional models of a product, (ii) the physical structure of the product, and (iii) process-models associated to the product, with a description of the operations and resources (assembly process, maintenance process, etc.).

It is important to bring out the parallelism between these models and the ones used in the case of the product configuration problem (functional model or commercial view of the product, product physical structure or hierarchy of components/parts, and process-models or manufacturing process with routes and operations).

This approach for a technical domain used in RIMMEL seems to be general enough to be applicable in any other technical domain/product, raising a reference model for this kind of problems. In the same way, RIMSAT works with a Dublin Core–based documental ontology, to assign documental meta-data to each of the resources of the corpus (documents, plans, maps…). In a technical domain, these meta–data can frequently be created in an automatic way because the

general documentation of a variant of a product is also generated automatically from the different product-process models.

A first study and analysis was carried out for the case submitted from IKERLAN. It is included in the Annex to the Technical Report ISYS 06-07588/1, as well as the analysis of the structure of the documents coming from the company ORONA (diagrams, manuals, technical reports regarding elevators). As a result a bilingual (English Spanish, aligned) domain ontologies were developed, consisting 60 nodes in 8 levels.

Prototypes development

The first prototype was introduced in the Project Meeting held on October, 5th, 2005. It is fully described in the Technical Report ISYS 06-07588/1. This prototype system consists of three modules: (i) module for modifications of elevators ontologies made by an expert user, as well as accessing the documents searcher; (ii) Local information retrieval hybrid module, combining vector space model and semantic searches (using ontologies), and (iii) module for describing and referencing the diagrams and images that appear in the texts for proper retrieval.
A second version of the prototype includes different linguistic resources, for translating, segment identification, proper nouns detection, morphological analysis and web retrieval (Google/yahoo) for integrating local and web results, if the user so decides. A fuzzy integration scheme for both searches is now under development.
The prototypes have been evaluated the TIME [12] reference collection, which consists of 450 texts and 20 queries. The results have been qualitatively good. During 2006, Ikerlan has also evaluated the prototype developed by UPM from the point of view of the needs raised by the documental corpus given by Orona. This corpus is composed by complex and highly structured documents, using the most used graphic formats in the technical industry (*.dxf, *.tiff). As a result of this analysis, RIMSAT has raised new functional needs for the next version of the prototype.

Resources development

The version of the TRIELIB library available when the project started has been adapted for its use for document indexing. On top of the library, a specific tool has been built for indexing huge document collections, as well as for querying these indexes. The retrieval modules implements both the vector and probabilistic (BM25 [10]) models. This tool can be used for indexing and retrieval text documents written in any language and any encoding (ISO 8859, UTF, and others). This was not possible using other freely available indexing and retrieval tools. The trie structure used has allowed us a very quick and efficient indexing for the huge collections used in the evaluation campaigns, much better than using the Xapian [14] tool, publicly available and previously used.

Navigation tools for the EuroWordNet resource have also been developed, based on the storage of this resource in a trie-based structure. This tool can be used for translating words between the different languages available in EuroWordNet or for lexical expansion of terms based on semantic relationships.

Specific tools for IR have been developed. One of the research lines explored was the combination of results lists obtained using different approaches for querying the same collection using the same query, or querying a multilingual collection using the same query translated to each language. Different algorithms were tested, so different combination tools were scripted.

An entities detection and tagging tool is under development (a first prototype is available). We think that IR processes can be greatly enhanced with a proper detection, tagging and normalization of entities (proper nouns, toponyms, company's names, etc.). This field is one of the ones where research in IR and QA is strongly directed to.

The UAM team has improved their morphological analyzer, based on GRAMPAL and ARIES lexicons (jointly developed with the UPM team in previous projects). This morphological analyzer/tagger had been specifically adapted to spontaneous speech corpora, in particular the C-ORAL-ROM corpus. That is, it has adapted to the tokenization, disambiguation rules and lexicon for spoken texts.

A parallel corpus in Arabic, Spanish and English of 100.000 words each approximately has been compiled. This is a unique resource, never previously done for those three languages. This is a main contribution of the Doctoral Dissertation by Doaa Samy, member of the UAM team. Another contribution has been a first prototype for an Arabic Named Entities recognizer.

With a similar methodology, a Hindi-Spanish parallel corpus is been compiled by the UAM team. Up to date, a small corpus of Spanish and Hindi news and a several chapters from "La rebelión de las masas" by Ortega y Gasset have been collected. For Japanese, over 5 hours of spontaneous speech recorded in Japan and Madrid from native speakers is the main result. The next step is to start the transcription of the recorded materials.

Evaluation in international fora

The research team of subprojects 1 and 2, as well as other researchers coming from Universidad Carlos III in Madrid, and from the private company DAEDALUS has participated in the CLEF international forum (Cross-Language Evaluation Forum [3]). The team, so-called MIRACLE, is participating in CLEF since 2003, that is, we have been in four evaluation campaigns. The 2005 and 2006 campaigns were celebrated during the development of this project. A lot of research groups, whose research targets several fields related to multilingual IR, participate in CLEF. As an example, in 2006, the groups were 90 (74 in 2005): 60 from Europe, 14 from USA, 10 from Asia, 4 from South-America and 2 from Australia.

The contribution to CLEF has been developed in 6 tracks in 2005 and 4 in 2006:

- **AdHoc:** General experiments for Monolingual, Bilingual and Multilingual IR from a multilingual document collection. This is the CLEF original track, which attracts most participation.
- **ImageCLEF:** Information Retrieval from image collections, using IR techniques over textual annotations to images as well as content-based retrieval.
- **GeoCLEF:** Multilingual IR on the multilingual document collection, using geographical information to support the queries.
- **WebCLEF:** Multilingual IR over a collection of pages extracted from official sites from several European countries (not in 2006).
- **iCLEF:** Interactive experiments for IR, targeted to evaluate the efficiency and facility of use of the systems (not in 2006).
- **QA-CLEF:** Question Answering experiments over multilingual document collections.

The MIRACLE team is one of them that participate in a larger number of tracks in 2005 and 2006 campaigns. These campaigns do not permit to sort the different participating systems and approaches in an absolute manner, since the evaluation results compare only a particular type of

experiments but not the others, the texts are of a very specific category (news) and these ones have not changed significatively for years. In spite of that, we can say that they have been very satisfactory for us, and with these cautions we can mention the more important positions obtained:

- **CLEF 2005:**
  - AdHoc: Monolingual Hungarian (3rd), Bulgarian (4th). Bilingual Bulgarian (1st), French (5th), Portuguese (3rd), Multilingual (3rd).
  - ImageCLEF: Portuguese, Italian, Turkish, Norwegian, Philippine, Polish, Romanian, Bulgarian, Czech, Croatian, Finnish and Hungarian (1st).
  - WebCLEF: Mixed Monolingual (5th), Multilingual (2nd).
  - GeoCLEF: Monolingual German (3rd), English (5th).
- **CLEF 2006:**
  - AdHoc: Monolingual French (5th), Bulgarian (5th). Bilingual Bulgarian (1st), French (4th), Hungarian (1st), Portuguese (5th). Robust Monolingual Dutch (2nd), English (4th), German (3rd), Italian (4th), Spanish (4th). Bilingual Dutch (1st), German (1st), Spanish (3rd). Multilingual (2nd).
  - GeoCLEF: Monolingual German (4th), Spanish (3rd).
  - QA-CLEF: WiQA Monolingual Spanish (3rd, 4th). Real time QA (1st).

Our participation has submitted experiment runs in a significative number of languages. In fact, almost all possible languages in CLEF.

In 2005, we started an initial participation in an IR evaluation forum, oriented towards Asiatic languages (Chinese, Japanese and Korean), NTCIR (*NII-NACSIS Test Collection for IR Systems [9]).* The results obtained are limited to have been able to submit results for such different languages. Unfortunately, we have been unable to participate in the 2006 edition.

For 2007 we are planning a new participation in CLEF, maybe limited to a lesser number of tracks. We are also in a decision process for participating in NTCIR 2007 and other international fora, such as TC-STAR (TC-Star Evaluation Workshop on Speech-to-Speech Translation [11]), TREC (Text Retrieval Evaluation Conferences [13]), ACE (Automatic Content Extraction [1]), INEX (Initiative for the Evaluation of XML Retrieval [8]), or others.

3 Results Indicators

Research Personnel

As it was foreseen in the project proposal, two persons have been contracted for carrying out some tasks. The first, Rubén Granados Muñoz, has some responsibilities, such as the maintenance, enlargement and management of linguistic and ontological resources, and the implementation and integration of processing modules in the demonstrator prototypes. The second, Raúl de la Torre, have assisted the implementation of prototypes and some tools for tagging resources and for linguistic processing.

Some of the researches of this project have obtained their PhD. Thesis in the project:

- Manuel Alcántara Pla. **Anotación y recuperación de información semántica eventiva en corpus.** Departamento de Lingüística, Lógica y Filosofía de la Ciencia. Universidad Autónoma de Madrid. Director: Antonio Moreno Sandoval.

- Doaa Samy. **Recursos bilingües de Ingeniería Lingüística para el procesamiento de español y árabe.** Departamento de Lingüística, Lógica y Filosofía de la Ciencia. Universidad Autónoma de Madrid. Director: Antonio Moreno Sandoval.

Others are working towards the doctoral degree:

- Sara Lana Serrano. Evaluación de Técnicas para la Recuperación de Información: GEOCLEF 2005. Report for DEA degree. Departamento de Matemática Aplicada a las Tecn. de la Inf. ETSI de Telecomunicación. Universidad Politécnica de Madrid. Diciembre de 2006. Director: José Miguel Goñi Menoyo.
- Prem Prakash. Recursos Bilingües de Ingeniería Lingüística: Español-Hindi. Report for DEA degree. Departamento de Lingüística, Lógica y Filosofía de la Ciencia. Universidad Autónoma de Madrid. Director: Antonio Moreno Sandoval.
- Marta Garrote. Compilación de un corpus de habla infantil. Report for DEA degree. Departamento de Lingüística, Lógica y Filosofía de la Ciencia. Universidad Autónoma de Madrid. Director: Antonio Moreno Sandoval.

Some Final Career Projects have been developed and finished in the field of the project, and directed by its researches:

- Juan José Fernández Tébar. **Recuperación de información multilingüe y multimedia y su evaluación.** Departamento de Inteligencia Artificial. Facultad de Informática. Universidad Politécnica de Madrid. Diciembre de 2005. Director: Ana María García Serrano.
- Paloma de Juan González. **Desarrollo de una plataforma para la mejora de técnicas de recuperación de información multilingüe en Internet.** Departamento de Matemática Aplicada a las Tecn. de la Inf. ETSI de Telecomunicación. Universidad Politécnica de Madrid. Diciembre de 2006. Director: José Miguel Goñi Menoyo.

Publications CLEF 2006

- Goñi-Menoyo, José M.; González-Cristóbal, José C.; Lana-Serrano, Sara; and Martínez-González, Ángel. MIRACLE's Ad-Hoc and Geographical IR approaches for CLEF 2006. 7th Workshop of the Cross Language Evaluation Forum, 2006, CLEF 2006, Alicante, Spain, Revised Selected Papers (Carol Peters, Paul Clough, Fredric C. Gey, Julio Gonzalo, Douglas W. Oard, Maximilian Stempfhuber, Bernardo Magnini, Maarten de Rijke, Eds.). Lecture Notes in Computer Science. Springer, to be published.
- José Luis Martínez-Fernández, Ana García Serrano, Julio Villena-Román, and Paloma Martínez. Expanding Queries through Word Sense Disambiguation. 7th Workshop of the Cross Language Evaluation Forum, 2006, CLEF 2006, Alicante, Spain, Revised Selected Papers (Carol Peters, Paul Clough, Fredric C. Gey, Julio Gonzalo, Douglas W. Oard, Maximilian Stempfhuber, Bernardo Magnini, Maarten de Rijke, Eds.). Lecture Notes in Computer Science. Springer, to be published.
- César de Pablo-Sánchez, Ana González-Ledesma, Antonio Moreno-Sandoval, and Maria Teresa Vicente-Díaz. MIRACLE experiments in QA@CLEF 2006 in Spanish: main task, real-time QA and exploratory QA using Wikipedia (WiQA). 7th Workshop of the Cross Language Evaluation Forum, 2006, CLEF 2006, Alicante, Spain, Revised Selected Papers (Carol Peters, Paul Clough, Fredric C. Gey, Julio Gonzalo, Douglas W. Oard, Maximilian Stempfhuber, Bernardo Magnini, Maarten de Rijke, Eds.). Lecture Notes in Computer Science. Springer, to be published.

Publications CLEF 2005

- Goñi-Menoyo, José M.; González-Cristóbal, José C.; y Villena-Román, Julio. MIRACLE at Ad-Hoc CLEF 2005: Merging and Combining without Using a Single Approach. Accessing Multilingual Information Repositories: 6th Workshop of the Cross Language Evaluation Forum 2005, CLEF 2005, Vienna, Austria, Revised Selected Papers (Carol Peters, Fredric C. Gey, Julio Gonzalo, Gareth J.F.Jones, Michael Kluck, Bernardo Magnini, Henning Müller, Maarten de Rijke, Eds.). Lecture Notes in Computer Science, vol. 4022, págs. 44-53. Springer, 2006.
- Lana-Serrano, Sara; Goñi-Menoyo, José M.; y González-Cristóbal, José C. MIRACLE at GeoCLEF 2005: First Experiments in Geographical IR. Accessing Multilingual Information Repositories: 6th Workshop of the Cross Language Evaluation Forum 2005, CLEF 2005, Vienna, Austria, Revised Selected Papers (Carol Peters, Fredric C. Gey, Julio Gonzalo, Gareth J.F.Jones, Michael Kluck, Bernardo Magnini, Henning Müller, Maarten de Rijke, Eds.). Lecture Notes in Computer Science, vol. 4022, págs. 920-923. Springer, 2006.
- Villena-Román, Julio; Crespo-García, Raquel M; y González-Cristóbal, José C. Effect of Connective Functions in Interactive Image Retrieval. Accessing Multilingual Information Repositories: 6th Workshop of the Cross Language Evaluation Forum 2005, CLEF 2005, Vienna, Austria, Revised Selected Papers (Carol Peters, Fredric C. Gey, Julio Gonzalo, Gareth J.F.Jones, Michael Kluck, Bernardo Magnini, Henning Müller, Maarten de Rijke, Eds.). Lecture Notes in Computer Science, vol. 4022, págs. 293-296. Springer, 2006.
- César de Pablo-Sánchez, Ana González-Ledesma, José Luis Martínez-Fernández, José María Guirao, Paloma Martínez, and Antonio Moreno. MIRACLE's Cross-Lingual Question Answering Experiments with Spanish as a Target Language. Accessing Multilingual Information Repositories: 6th Workshop of the Cross Language Evaluation Forum 2005, CLEF 2005, Vienna, Austria, Revised Selected Papers (Carol Peters, Fredric C. Gey, Julio Gonzalo, Gareth J.F.Jones, Michael Kluck, Bernardo Magnini, Henning Müller, Maarten de Rijke, Eds.). Lecture Notes in Computer Science, vol. 4022, págs. 488-491. Springer, 2006.
- José Luis Martínez-Fernández, Julio Villena-Román, Ana García Serrano, and José Carlos González-Cristóbal. Combining Textual and Visual Features for Image Retrieval. Accessing Multilingual Information Repositories: 6th Workshop of the Cross Language Evaluation Forum 2005, CLEF 2005, Vienna, Austria, Revised Selected Papers (Carol Peters, Fredric C. Gey, Julio Gonzalo, Gareth J.F.Jones, Michael Kluck, Bernardo Magnini, Henning Müller, Maarten de Rijke, Eds.). Lecture Notes in Computer Science, vol. 4022, págs. 680-691. Springer, 2006.
- Ángel Martínez-González, José Luis Martínez-Fernández, César de Pablo-Sánchez, Julio Villena-Román. MIRACLE at WebCLEF 2005: Combining Web Specific and Linguistic Information. Accessing Multilingual Information Repositories: 6th Workshop of the Cross Language Evaluation Forum 2005, CLEF 2005, Vienna, Austria, Revised Selected Papers (Carol Peters, Fredric C. Gey, Julio Gonzalo, Gareth J.F.Jones, Michael Kluck, Bernardo Magnini, Henning Müller, Maarten de Rijke, Eds.). Lecture Notes in Computer Science, vol. 4022, págs. 869-872. Springer, 2006.

Other publications

- W.S. Wiguna, JJ Fernández-Tébar, A García-Serrano. Using a Fuzzy Model for Combining Search Results from Different Information Sources to Build a Metasearch Engine. Computational Intelligence, Theory and Applications. International Conference 9th Fuzzy Days in Dortmund, Germany, Sept. 18–20, 2006 Proceedings, págs. 325-334. Springer 2006.
- ALCÁNTARA, Manuel y BERTOMEU, Nuria (2005): "Ellipsis in spontaneous spoken language". Workshop Cross-modular Approaches to Ellipsis, ESSLLI, Edimburgo.

- DE CUADRA, Mª Teresa; SÁNCHEZ, Asunción; GONZÁLEZ, Ana; MORENO, Antonio (2005): "Adaptaciones necesarias de las normas de transcripción para el Corpus Oral sobre la Cata de Vinos". AESLA, Mallorca.
- GONZÁLEZ, Ana y MORENO, Antonio (2005): "Pragmatic analysis of man-machine interactions in a spontaneous speech corpus", Corpus Linguistics 2005, Birmingham.
- MORENO, Antonio, DE LA MADRID, Guillermo, ALCÁNTARA, Manuel, GONZÁLEZ, Ana, GUIRAO, José M. y DE LA TORRE, Raúl (2005): "The Spanish corpus", Cresti y Moneglia (eds.) C-ORAL-ROM Integrated Reference Corpora for Spoken Romance Languages, págs 135-161, Amsterdam, John Benjamins.
- SAMY, Doaa (2005) : "Named Entities: Structure and Translation. A study based on a Parallel Corpus (Arabic-English-Spanish)", RANLP-2005, Borovets, Bulgaria.
- SAMY, Doaa; MORENO, Antonio y GUIRAO, José María (2005): "A Proposal For An Arabic Named Entity Tagger Leveraging a Parallel Corpus", *Corpus Linguistics*, Birmingham.
- TORRALBO, Rafael; ALFONSECA, Enrique; GUIRAO, Jose María y MORENO, Antonio (2005) : "Description of the UAM system at DUC-2005". Proceedings of the Document Understanding Conference 2005, Vancouver, Canadá.
- TORRALBO, Rafael; ALFONSECA, Enrique; MORENO, Antonio y GUIRAO, Jose María (2005): "Automatic generation of term definitions using multidocument summarisation from the web". RANLP-2005, Borovets, Bulgaria.
- TORRE, Doroteo; MORENO, Antonio; COLÁS, José y GARRIDO, Javier (2005): "Acoustic-phonetic decoding of different types of spontaneous speech in Spanish" Disfluencies in Spontaneous Speech 2005, Aix-en-Provence, Francia.
- E. ALFONSECA, M. OKUMURA, A. MORENO-SANDOVAL y J. M. GUIRAO. "Googling answer's models in question-focused summarization". Proceedings of Document Understanding Conference 2006.
- J.M. GUIRAO, A. MORENO SANDOVAL, A. GONZÁLEZ LEDESMA, G. DE LA MADRID, M. ALCÁNTARA (2006): "Relating linguistic units to socio-contextual information in a spontaneous speech corpus of Spanish". En Wilson, Archer, y Rayson (eds.)(2006) Corpus linguistics around the world. Amsterdam: Rodopi, pp. 101-114.
- MORENO, A. y URRESTI, J. (2006): "El proyecto C-ORAL-ROM y su aplicación a la enseñanza del español", En Oralia, 8.
- MORENO, A. y GUIRAO, J.M. (2006):"Integrated tools for explotation of a spontaneous speech corpus of Spanish". En Proceedings of Computers, Literatura and Philology (CLIP 2006).
- D. SAMY, A. MORENO-SANDOVAL, J. M. GUIRAO y E. ALFONSECA (2006): "Building a Parallel Multilingual Corpus (Arabic-Spanish-English)". Proceedings of the Language Resources and Evaluation Conference, LREC-2006.

CLEF 2006 Workshop

- Goñi-Menoyo, José Miguel; González-Cristóbal, José Carlos y Villena-Román, Julio. Report of MIRACLE Team for the Ad-Hoc Track in CLEF 2006. Working Notes for the CLEF 2006 Workshop, ISBN 2-912335-23-x, ISSN 1818-8044, Alicante, España, sept. 2006. On line http://www.clef-campaign.org/2006/working_notes/working_notes2006/goniCLEF2006.pdf, last visited 10-1-2007.
- Lana-Serrano, Sara; Goñi-Menoyo, José M. y González-Cristóbal, José C. Report of MIRACLE Team for Geographical IR in CLEF 2006. Working Notes for the CLEF 2006 Workshop, ISBN

2-912335-23-x, ISSN 1818-8044, Alicante, España, sept. 2006. On line http://www.clef-campaign.org/2006/working_notes/workingnotes2006/lanaCLEF2006.pdf, last visited 10-1-2007.

- César de Pablo-Sánchez, Ana González-Ledesma, Antonio Moreno, José Luis Martínez-Fernández, and Paloma Martínez. MIRACLE at the Spanish QA@CLEF 2006. Working Notes for the CLEF 2006 Workshop, ISBN 2-912335-23-x, ISSN 1818-8044, Alicante, España, sept. 2006. On line http://www.clef-campaign.org/2006/working_notes/workingnotes2006/depabloCLEF2006.pdf, last visited 10-1-2007.
- César de Pablo-Sánchez, José Luis Martínez-Fernández, and Paloma Martínez. MIRACLE at the Spanish WiQA Pilot: Using Named Entities and Cosine Similarity to extend Wikipedia articles. Working Notes for the CLEF 2006 Workshop, ISBN 2-912335-23-x, ISSN 1818-8044, Alicante, España, sept. 2006. On line http://www.clef-campaign.org/2006/working_notes/workingnotes2006/depabloWiQACLEF2006.pdf, last visited 10-1-2007.
- José Luis Martínez-Fern*ández, Julio Villena, Ana García-Serrano, Paloma Martínez.* MIRACLE Team Report for ImageCLEF IR in CLEF 2006. Working Notes for the CLEF 2006 Workshop, ISBN 2-912335-23-x, ISSN 1818-8044, Alicante, España, sept. 2006. On line http://www.clef-campaign.org/2006/working_notes/workingnotes2006/villenaCLEF2006.pdf, last visited 10-1-2007.

NTCIR 2005 Workshop

- Villena-Román, Julio; Goñi-Menoyo, José Miguel; González-Cristóbal, José C.; y Martínez-Fernández, José Luis. MIRACLE Retrieval Experiments with East Asian Languages. Proceedings of the Fifth NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access. Tokio, Japón, dec. 2005. On line http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings5/data/CLIR/NTCIR5-CLIR-VillenaJ.pdf, last visited 13-3-2006.

CLEF 2005 Workshop

- González, José C.; Goñi-Menoyo, José Miguel y Villena-Román, Julio. Miracle's 2005 Approach to Cross-lingual Information Retrieval. Working Notes for the CLEF 2005 Workshop. Viena, Austria, sept. 2005. On line http://www.clef-ampaign.org/2005/working_notes/workingnotes2005/gonzalez05.pdf, last visited 13-3-2006.
- Goñi-Menoyo, José Miguel; González, José C. y Villena-Román, Julio. Miracle's 2005 Approach to Monolingual Information Retrieval. Working Notes for the CLEF 2005 Workshop. Viena, Austria, sept. 2005. On line http://www.clef-ampaign.org/2005/working_notes/workingnotes2005/menoyo05.pdf, last visited 13-3-2006.
- Julio Villena-Román, Raquel M. Crespo-García, and José Carlos González-Cristóbal. Boolean Operators in Interactive Search. Working Notes for the CLEF 2005 Workshop. Viena, Austria, sept. 2005. On line http://www.clef-campaign.org/2005/working_notes/workingnotes2005/villena-roman05.pdf, last visited 10-1-2007.
- César de Pablo-Sánchez, Ana González-Ledesma, José Luis Martínez-Fernández, José Maria Guirao, Paloma Martínez y Antonio Moreno. MIRACLE's 2005 Approach to Cross-Lingual Question Answering. Working Notes for the CLEF 2005 Workshop. Viena, Austria, sept. 2005. On line http://www.clef-ampaign.org/2005/working_notes/workingnotes2005/sanchez05.pdf, last visited 13-3-2006.

- J. L. Martínez-Fernández, J. Villena, Ana García-Serrano , S. González-Tortosa, F. Carbone y M.Castagnone. Exploiting Semantic Features for Image Retrieval at CLEF 2005. Working Notes for the CLEF 2005 Workshop. Viena, Austria, sept. 2005. On line http://www.clef-campaign.org/2005/working_notes/workingnotes2005/carbone05.pdf, last visited 13-3-2006.
- Villena-Román, Julio; González-Cristóbal, José Carlos; Goñi-Menoyo, José Miguel; Martínez Fernández, José Luis y Fernández, Juan José. MIRACLE's Combination of Visual and Textual Queries for Medical Images Retrieval. Working Notes for the CLEF 2005 Workshop. Viena, Austria, sept. 2005. On line http://www.clef-campaign.org/2005/working_notes/ workingnotes2005/cristobal05.pdf, last visited 13-3-2006.
- Villena-Román, Julio; González-Cristóbal, José Carlos; Goñi-Menoyo, José Miguel y Martínez Fernández, José Luis. MIRACLE's Naive Approach to Medical Images Annotation. Working Notes for the CLEF 2005 Workshop. Viena, Austria, sept. 2005. On line http://www.clef-campaign.org/2005/working_notes/workingnotes2005/villena-roman205.pdf, last visited 13-3-2006.
- Ángel Martínez-González, José Luis Martínez-Fernández, César de Pablo-Sánchez, Julio Villena-Román, Luis Jiménez-Cuadrado, Paloma Martínez, José Carlos González-Cristóbal. MIRACLE's Approach to Multilingual Web Retrieval. Working Notes for the CLEF 2005 Workshop. Viena, Austria, sept. 2005. On line http://www.clef-ampaign.org/2005/working_notes/ workingnotes2005/fernandez05.pdf, last visited 13-3-2006.
- Lana-Serrano, Sara; Goñi-Menoyo, José M. y González-Cristóbal, José C. MIRACLE's 2005 Approach to Geographical Information Retrieval. Working Notes for the CLEF 2005 Workshop. Viena, Austria, sept. 2005. On line http://www.clef-ampaign.org/2005/ working_notes/workingnotes2005/serrano05.pdf, last visited 13-3-2006.

Technical reports

- ISIS-GSI, Facultad de Infomática, Universidad Politécnica de Madrid. (2005, Octubre) Informe Técnico ISYS 06-07588/1.

Technological transfer

- TRIELIB: A Software Library for Trie Management. Inscription requested in the Spanish Intellectual Property Registry on December 2006. It was developed ointly by Universidad Politécnica de Madrid and the company DAEDALUS-Data, Decisions, and Language, S.A. Authors: José Miguel Goñi Menoyo, José Carlos González Cristóbal (UPM), and Jorge Fombella Mourelle, Julio Villena Román (DAEDALUS).
- The exploitation rights of TRIELIB were granted to DAEDALUS in September 2006 bu agreement between UPM and DAEDALUS.
- A Licence for using TRIELIB has been granted to Mundinteractivos, S.A. (propietary company of elmundo.es), June 2006.
- Licence agreement for "BASE LEXICA DE LA PLATAFORMA ARIES" granted to "DAEDALUS – DATA, DECISIONS AND LANGUAGE, S.A.". May, 2006. (This resource was inscribed in the Spanish Intellectual Property Registry in 1996, and it is under continuous enhancement, partly done in this project).

Collaborations with other Research Groups

- The project "MAVIR-CM: Mejorando el acceso y la visibilidad de la información multilingüe en red para la Comunidad de Madrid" (S-0505/TIC/000267) was granted and funded by

Comunidad de Madrid. This project is a network of research groups in Comunidad de Madrid that collaborate in research projects in Information Retrieval, Extraction, Question Answering and related fields. Participant groups are adscribed to Universidad Carlos III de Madrid, Universidad Autónoma de Madrid, Universidad Nacional de Educación a Distancia (coordinator), Universidad Europea de Madrid and Centro de Información y Documentación Científica del Consejo Superior de Investigaciones Científicas. January 2006 to December 2009.

- The project research team has joined a CONSOLIDER proposal (11 research groups, more than 70 doctors) entitled KNOWEB: Web, conocimiento, comunicación multilingüe y multimedia (CSD00C-07-21520). Its proposed duration is 5 years.
- The participating researches of UPM were adscribed formally to different research groups when the project started, althoug with a strong liasons of collaboration in academic and research activities. During 2005, the UPM University approved norms for formal Research Groups and a call for their formation. Now, all researchers in UPM belong formally to the same Research Group in UPM, so-called "Grupo de Sistemas Inteligentes" (ISYS-GSI).
- Two researchers of the UAM team, Francisco Marcos-Marín and Antonio Moreno-Sandoval are integrated in the workgroup for Linguistics and Language of the site CampusRed of Fundación Telefónica, being the first of them the Head of the group.

Dissemination

- Project and First prototype presentation to doctorate students in the Open Doors Week in Facultad de Informática in Universidad de Deusto, Bilbao (invited talk). April 2005.
- Project prototype presentation in the round table "AI applications" in Conferencia española para la inteligencia artificial CAEPIA, Santiago de Compostela, Galicia, November 2005.
- Open Doors Day in ISYS-GSI. Facultad de Informática. Semana de la Ciencia y la Tecnología de la Comunidad de Madrid (invited talk), November 2005.

4 References

[1] ACE page. http://www.nist.gov/speech/tests/ace/, last visited 18-1-2007.

[2] Aoe, Jun-Ichi; Morimoto, Katsushi; Sato, Takashi. An Efficient Implementation of Trie Structures. Software Practice and Experience 22(9): 695-721, 1992.

[3] CampusRed Languages and Linguistics page. On line http://www.campusred.net/area_lenguas/, last visited 18-1-2007.

[4] CLEF page. On line http://www.clef-campaign.org , last visited 18-1-2007.

[5] DAEDALUS home page. On line http://www.daedalus.es, last visited 18-1-2007.

[6] EuroWordNet page. On line http://www.illc.uva.nl/EuroWordNet/, last visited 18-1-2007.

[7] Goñi-Menoyo, José Miguel; González-Cristóbal, José Carlos and Fombella-Mourelle, Jorge. An optimised trie index for natural language processing lexicons. MIRACLE Technical Report. Universidad Politécnica de Madrid, 2004.

[8] INEX page. http://inex.is.informatik.uni-duisburg.de/, last visited 18-1-2007.

[9] NTICR page. On line http://research.nii.ac.jp/ntcir/, last visited 18-1-2007.

[10] Robertson, S.E. et al. Okapi at TREC-3. In Overview of the Third Text REtrieval Conference (TREC-3). D.K. Harman (Ed.). Gaithersburg, MD: NIST, April 1995.

[11] TC-STAR Workshop page. On line http://www.elda.org/tcstar-workshop/, last visited 18-1-2007.

[12] TIME reference collection. On line ftp://ftp.cs.cornell. edu/pub/smart/time/, last visited 18-1-2007.

[13] TREC page. On line http://trec.nist.gov/, last visited 18-1-2007.

[14] Xapian page. On line http://www.xapian.org/, last visited 18-1-2007.

Appendix I

Initial schedule of the project.

Subproject 1:

Activities/tasks	Centre	Responsible and people	First year	Second year	Third year
1.1.1 Management and coordination of the whole project	All	JMG,AMS,KI	x|x|x|x|x|x|x|x|x|x|x|x	x|x|x|x|x|x|x|x|x|x|x|x	x|x|x|x|x|x|x|x|x|x|x|x
1.1.2 Management and coordination of subproject 1	UPM	JMG, all	x|x|x|x|x|x|x|x|x|x|x|x	x|x|x|x|x|x|x|x|x|x|x|x	x|x|x|x|x|x|x|x|x|x|x|x
1.2.1 Improvement of existing inguistic resources	UPM	JVR,JHD,P1A	x|x|x|x|x|x|x|x|x|x|x|x	x|x|x|x|x|x|x|x|x|x|x|x	||||||||||||
1.2.2 Improvement of processing modules	UPM	JLM,JMG,P1B	|||||x|x|x|x|x|x	x|x|x|x|x|x|x|x|x|x|x|x	x|x|x|x|x|x||||||
1.3.1 Domain modeling	UPM	AGS,JHD,ANS	x|x|x|x|x|x|x|x|x|x|x|x	x|x|x|x|x|x||||||	||||||||||||
1.4.1 Analysis and design	UPM	JMG,JVR,JLM,JCG	x|x|x|x|x|x|||||||	x|x|x|x|x|x|||||||	x|x|x|x|x|x|||||||
1.4.2 Integration of modules and resources	UPM	ANS,JVR,JLM,P1A	|||x|x|x|x|x||||	||||x|x|x|x|x||||	||||x|x|x|x|x||||
1.5.1 Participation in evaluation fora	UPM	JCG,JVR,JLM,JMG	|||||x|x|x|x||||	||||||x|x|x|x||||	||||||x|x|x|x||||
1.5.2 Evaluation in an industrial production environment	UPM	JHD, AGS, P1B	|||||||||||||	||||||x|x|x|x|x|x||	x|x|x|x|x|x|x|x|x|x|x|x

Subproject 2: RILARIM

Activities/tasks	Centre	Responsible and people	First year	Second year	Third year
2.1.1 Managemento of subproject 2	UAM	AMS, all			
2.2.1 Design of an unified XML tagging language	UAM	AMS,JMGM,MAP,DOS, FMM			
2.2.2 Conversion to the unified format	UAM	JMGM,P2b,MAP,AMS			
2.2.3 Web-access querying system	UAM	JMGM,P2a			
2.3.1 Enlargement of Hindi-Spanish corpus	UAM	PRPS			
2.3.2 Enlargement of Japanese corpus	UAM	CHK			
2.3.3 Development of resources for Arabic	UAM	DOS, MEM, FMM			
2.3. 4 Development of resources for Spanish	UAM	MAP,AMS,P2b			
2.3.5 Development of resources for Hindi and Japanese	UAM	AMS, JMGM, MAP,DOS, CHK, PRPS			
2.4.1 Diseño de los corpus de prueba	UAM	AMS, MAP,DOS,CHK, PRPS,MEM,FMM			
2.4.2 Preparación para evaluación de manera automática	UAM	AMS, JMGM, FMM			
2.4.3 Análisis y verificación de resultados	UAM	AMS,all			
2.5.1 Sistema de consulta de corpus a través de internet	UAM	JMGM, MAP, DOS			
2.5.2 Publicación de resultados	UAM	AMS, all			

Subproject 3 RIMSAT

Activities/tasks	Centre	Responsible and people	First year	Second year	Third year
3.1.1 Management of subproject 3	IKERLAN	all			
3.2.1 Data analysis	IKERLAN	all			
3.2.2 Domain ontological modeling	IKERLAN	all			
3.3.1 Indexing with ontology	IKERLAN	all			
3.4.1 Definition of an ontollogy-based IR system	IKERLAN	all			
3.5.1 Evaluation	IKERLAN	all			

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

AVISA: A semi-automatic situation diagnosis system in surveillance tasks based in selective and Active VISual Attention with learning capacity techniques TIN2004-07661-C02

José Mira*
Univ. Nacional de Educación a Distancia

Antonio Fernández-Caballero**
Universidad de Castilla-La Mancha

Abstract

The project described develops a set of generic components to help humans (semi-automatic system) in surveillance and security tasks in several scenarios. These components are based in the computational model of a set of selective and Active VISual Attention mechanisms with learning capacity (AVISA) and in the superposition of an "intelligence" layer that incorporates the knowledge of human experts in security tasks. The project integrates the responses of these alert mechanisms in the synthesis of the three basic subtasks present in any surveillance and security activity: (1) Real-time monitoring. (2) Situation diagnosing. (3) Action planning and control. In order to augment the diversity of environments and situations where AVISA system may be used, as well as its efficiency as support to surveillance tasks, knowledge components derived from situating cameras on mobile platforms and using complementary sensors of a non-visual character are being developed.

Keywords: Dynamic selective visual attention; Visual surveillance; Monitoring; Situation diagnosis; Robotics.

1 Objectives of the Project

Surveillance is a multidisciplinary task affecting an increasing number of scenarios, services and customers. It aims to detect threats by continually observing large and vulnerable areas of a scenario considered to be of economic, social or strategic value because it can suffer theft, fire, vandalism or attacks. The range of scenarios is very wide and of very different complexity, going from the mere detection of movement that sets off an alarm to an integral control system that monitors the scene with different sensors, diagnoses the situation and plans a series of consistent actions. In any case, it always implies the observation of mobile objects (people, vehicles, etc.) in a predetermined environment to provide a description of their actions and interactions. Hence, this implies the detection of moving objects and their tracking, the recognition of objects, the analysis of the specific movement for each type of object and the interpretation of the activity. In recent

* Email: jmira@dia.uned.es
** Email: caballer@dsi.uclm.es

years there has been a significant number of works on the analysis of the activity in different application domains [12, 13, 22, 65, 68]. Most approximations for activity analysis activity are posed as a classification problem, i.e., they are based on defining models for each type of activity in a specific application domain, and they are highly dependent on the results obtained during tracking.

The *surveillance task* follows a common structure with general control tasks: 1) *monitoring* of a number of critical variables, whose deviation from normality is a sign of some dysfunction, 2) *diagnosis* of the problem, and 3) *planning* of relevant actions for solving the problem. In our surveillance approach a semi-automatic approximation, where at the end it is the operator of an alarms power station who takes the decisions, is considered.

One of the fundamental problems in this project is the enormous semantic gap between the physical signal level and the knowledge level. One of the ways of overcoming this gap is by segmenting it. There is consensus in the area, although with some variety and dispersion in the nomenclature, in accepting different description levels with an increasing degree of semantics [57, 58], which facilitates the injection of domain knowledge. In our work, four description levels are considered: *image*, *blobs*, *objects* and *activities/behaviours*. Each of these levels is modular and independent, and the information handled comes from the ontology of the proper level and from the lower level.

The *aims* of the project correspond to modelling and formalization of the tasks associated to each level and their integration. The UCLM subproject, namely "Design and implementation of selective visual attention system for the dynamic and distributed monitoring of scenarios with different kinds of moving non-rigid objects" focuses on the processes associated to pixels, blobs and static and dynamic objects considered for monitoring the scene. The *monitoring* output is a warning in the sense that AVISA system has detected some anomaly. Thus it is the alarm that activates a *diagnosis* process. The UNED subproject, "Design and implementation of a set of diagnosis, planning and control agents in surveillance tasks with capacity of learning and cooperation with humans", which shares the object level but focuses on the situation diagnosis task, starts from the results of the monitoring task. The differentiation between *monitoring* and *diagnosis* is consistent, for instance with Howarth and Buxton's work [23], where *monitoring* is said to be "passively moving from images to conceptual descriptions", while *watching* is "more active, task-focused", with feedback between a pre-intentional level from which behaviours emerge (from the images) and an intentional control level that focuses on the surveillance task process.

The chronogram of figure 1 describes the temporal sequencing of the four objectives/tasks associated to subprojects UNED (T_1^1 to T_4^1) and UCLM (T_1^2 to T_4^2).

T_1^1: Knowledge modelling of the human expert in surveillance tasks (situation diagnosis and coordinated action planning).

T_2^1: Formalization of the diagnosis and planning models.

T_3^1: Modelling, formalization and implementation of the reinforcement learning task for selective visual attention (SVA) and robot control.

T_4^1: Contribution to the development, implementation and evaluation of AVISA-SECISA prototypes in surveillance tasks in real scenarios (shared with subproject UCLM).

T_1^2: Knowledge modelling of the visual monitoring and scene pre-diagnosis tasks by means of SVA mechanisms.

T_2^2: Formalization of the SVA bottom-up and top-down models by means of accumulative computation (AC) and algorithmic lateral inhibition (ALI) methods.

T_3^2: Adaptation of reinforcement learning to the SVA, and action planning and control, which close the interaction loop of the moving cameras with the environment.

T_4^2: Contribution to the development, implementation and evaluation of AVISA-SECISA prototypes in surveillance tasks in real scenarios (shared with subproject UNED).

Tasks	Subproject	First year												Second year												Third year											
T_1^1	UNED	X	X	X	X	X	X	X	X	X																											
T_2^1	UNED							X	X	X	X	X	X																								
T_3^1	UNED	X	X	X	X	X	X	X									X	X	X	X	X	X	X	X	X	X	X	X									
T_4^1	UNED																			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
T_1^2	UCLM	X	X	X	X	X	X	X	X	X	X	X	X																								
T_2^2	UCLM							X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X														
T_3^2	UCLM	X	X	X	X	X	X	X									X	X	X	X	X	X	X	X	X	X	X	X									
T_4^2	UCLM																					X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Fig. 1. Project chronogram

2 Level of success reached in the project

According to the chronogram of figure 1, at the end of the second year tasks (T_1^1, T_2^1, T_2^2, T_2^2) should be accomplished, and tasks (T_3^1, T_4^1, T_3^2, T_4^2) should be initiated.

Tasks T_1^1 and T_2^1

The aim of task T_1^1 is to identify the sources and to conceptually model the knowledge, trying to reuse and to adapt the components of the libraries of methods and ontologies for the domain knowledge, whilst task T_2^1 has to formalize the previous concepts. The main problem faced in tasks T_1^1 and T_2^1 is that, when interpreting a scene the external knowledge necessary to fill the enormous semantic gap between the physical signal level and the knowledge level has to be injected, i.e., the need to adapt the sensory information to the abstraction level. The problem of analyzing data and obtaining a set of descriptions concurrent with different degrees of semantics is implicit in T_1^1 and T_2^1 tasks. Thus, when defining *blobs* – gotten from background-foreground separation - we really are speaking of segmentation, and when we introduce spatial-temporal configurations of events we are speaking of *objects* and their relations.

The entity handled at the first level is the *image* of the environment, which is captured by sensor "camera". An image does not contain semantic information on the application domain, but rather only raw environment data. After a pre-processing, where the entity type does not change, we move to the *blob* level, where very generic semantics are introduced. The regions of the image corresponding to the objects of interest or parts of the objects (foreground) are identified. At this level, the *object* level is superimposed, where all the information associated with the objects of interest is organised. Finally, the *activity* level is introduced, where the specific activity describing the scene with the appropriate level of abstraction for the target task is represented by means of states and events.

To exemplify the methodological proposal developed in tasks $\mathbf{T_1^1}$ and $\mathbf{T_2^1}$, initially a very simple scenario was defined, which was called "sofa scenario". It is an indoor space that is a pass-through area for humans. Humans can move freely, come in and go out of the observation area, sit down on the sofa, carry a briefcase, leave it, pick it up, etc. Figure 2 shows the state transition diagram for the automaton describing the scene at activity level. Each event is characterized by a series of generic attributes (identifier, type, context, value) and others dependent on the type of event (transition height, change in speed, duration ...). There is a distinction between primitive events, associated with transitions (in time) in the visual and spatial variables of the objects, and composed events, which will be defined by combination of primitive events by means of spatial-temporal relations (total or partial coincidence, sequencing) or by pattern recognition methods. In our approach primitive events are considered to be those directly produced by specific changes in the state of visual attributes [7] and no difference has been made between states and events [6]. Other more complex scenarios close to the proposal of the project have also been studied, such as the test case scenarios used by European projects CAVIAR[1] (Context Aware Vision using Image-based Active Recognition) and ISCAPS[2] (Integrated Surveillance of Crowded Areas for Public Security).

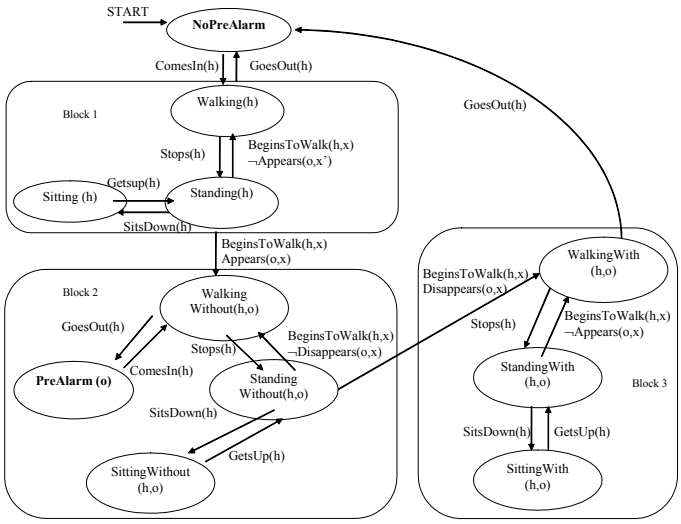


Fig. 2. State transition diagram used to model the activity level of the "sofa scenario"

The *difficulty* found in tasks T_1^1 and T_2^1 arises when connecting blobs to activities in different real scenarios. To solve this difficulty, at least partially, the following measures have been adopted: (1) Modelling and labelling of all families of sequences of interest in the proposed real scenarios and at the levels of objects, events and behaviours (construction of the ontologies of each level for each scenario). (2) Establishment of a clear frontier between automatic and human interpretation for each context (scenario, surveillance intention, information sources, robot accessibility, and so on). (3) Emphasizing in alarm pre-diagnosis as a complement to monitoring, leaving the decision on the action at each situation to the human operator. (4) Dedicating an additional effort to the construction of a case base that enables reasoning at each scenario by analogy to previous

[1] http://homepages.inf.ed.ac.uk/rbf/CAVIAR/
[2] http://www.iscaps.reading.ac.uk/

situations, in cooperation with tasks $\mathbf{T_3^1}$ and $\mathbf{T_3^2}$ (learning). (5) Exploring new ways of representing activities that enable limiting the combinatory explosion of the state transition diagrams of the deterministic finite automata. (6) Making explicit the difficulty of the situation diagnosis task and trying to integrate the efforts of this project with some complimentary ones (face recognition, etc.).

The success degree reached in T_1^1 and T_2^1 appears in the related publications. In summary it is: (1) Modelling and implementation of the motion detection task [52, 64]. (2) Modelling of the monitoring task in general and of the pre-diagnosis task in given scenarios [37, 67, 69]. (3) Establishment of a hierarchy of description levels with the ontologies corresponding to blobs, objects and simple events [11, 38, 73]. (4) Proposal of generic automata to describe the behaviours associated to a concrete scenario [5, 37]. (5) Generation of a tool for annotating images [66] and its later use in learning. (6) Some other precise developments on the modelling of humans and the recognition of parts of the body.

Tasks $\mathbf{T_3^1}$ and $\mathbf{T_3^2}$

The objectives of task T_3^1 (learning) are (1) to guide the search process in the SVA mechanisms by reinforcing the descriptions of the objects that look more like the predefined patterns, (2) to guide the robots in their pathways towards the selected coordinates, and, (3) to accumulate the AVISA experience in its collaboration with the human operator.

Of these three functions the first one is finished [28, 30, 59], with the complementation of reinforcement learning with other evolutionary models [14, 42] (see figure 3). Also, some incursions have been made in the treatment of uncertainty inherent to motion detection, tracking and analysis [54-56]. The second function is in development phase; to date an autonomous navigation system based on the calculation of the centre of area in open space is available [1-4]. The third function, associated to the interaction of the prototype with the human operator, will be developed during the third year in parallel with task T_4^1.

Fig. 3. A significant example of segmentation enhancement by evolutionary algorithms

The main problems of T_3^1 are given by the difficulty of constructing valid training sets for supervised learning in different scenarios. The solutions explored so far are the construction of a case base for each concrete scenario and the use of evolutionary procedures to enhance the segmentation task. The idea is to compare the angle and the module of the RGB vector associated to each image pixel corresponding to the background image, eliminating those objects that have been classified as noise in the ontology of the level, and facilitating image segmentation in terms of the moving objects labelled as interesting [9, 10, 40, 41].

Activities of task T_3^2 are coordinated with T_3^1 activities. Here, those parameters that contribute to a better discrimination (diagnosis) of different situations (video sequences) in accordance with the intention of the observer have to be reinforced [28, 30]. We have also explored the solution to some of the problems found in the calculation of the depth parameter by developing a stereovision prototype [15, 31, 32, 35, 36].

Tasks T_1^2 and T_2^2

The aim of task T_1^2 is to collect visual information from diverse coordinates of the scene and to pre-process them according to the SVA mechanisms. And, the function of T_2^3 is to focus attention of the cameras (static and mobile) on those elements of the scene that better fulfil the criteria specified by the human expert as "*objects of interest*". The processes corresponding to these tasks are related to the image, blob and object levels. Two basic mechanisms have been used: the accumulative computation (AC) and the algorithmic lateral inhibition (ALI) among agents with different degrees of semantic. The decision is taken in terms of the value of a set of parameters of motion, size and shape, firstly obtaining the blobs, and then injecting the semantic needed to get the objects of interest.

It has been possible to work on a real outdoor scenario [20, 21, 34, 48] monitored by the cooperating company SECISA (see left image at figure 4) and on the PETS2001 benchmark data (see right image of figure 4). This process has been developed in a sequential and incremental manner, starting with AC [15, 31, 32, 35] and ALI [26], formulated firstly at physical level and then in terms of multi-agent systems, where a totalizing process on the individual opinions of a set of working memories as agent coordination mechanism has been used. Next both mechanism have been combined [17, 19, 43, 72] to face the dynamic SVA problem [16, 24, 29], that is to say, "where to look to" and "what to look to". Our proposal is an integration of the bottom-up (connectionist) and top-down (symbolic) organizations usually accepted in Neuroscience.

Fig. 4. Some test case scenarios. Left: SECISA video sequence. Right: PETS2001 dataset

The main difficulties found are related to the interaction between moving objects (e.g., a human) and static ones (for instance, a suitcase) and to the complexity of working in real-time in real scenarios (camera movement, noise, 2D vs. 3D, calibration, stereovision, and so on). A partial solution has been provided by acquiring real-time vision processors [60, 71]. Nevertheless we again highlight the tremendous dependency on the context.

Tasks T_4^1 and T_4^2

Tasks T_4^1 and T_4^2 consider the construction of valid prototypes for real scenarios, which should be delivered during the third year. To date partial implementations that will be integrated during the whole year have been set up. Let us emphasize on the following finished implementations: (1) Navigation of moving robots by means of dynamic stereoscopic SVA [33, 70]. (2) SVA-based visual surveillance in real environments [27], modelled as multi-agent [61, 62] or holonic multi-sensorial [75] systems. (3) Visualization (see figure 5) and storage of video sequences [45], pre-diagnosis with simple pre-codified messages [47] and action planning in surveillance power station. (4) Real-time programs of all our SVA algorithms [25, 44, 60, 71].

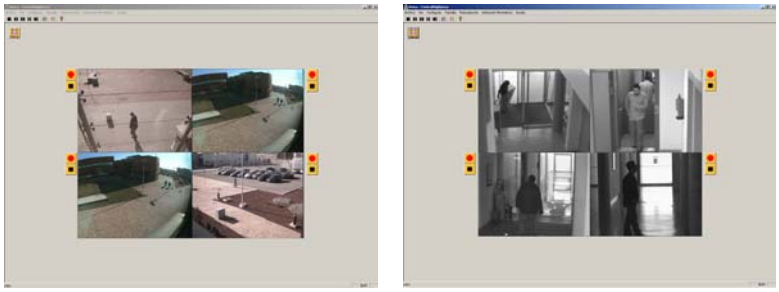


Fig. 5. Visualization task in prototype. Left: Indoor scenario. Right: Outdoor scenario

3 Result indicators

Student formation

- **PhD Dissertations:**
  - M.T. López Bonal. "Modelado computacional de los mecanismos de atención selectiva mediante redes de interacción lateral". ETSII. UNED, 2004.
  - J.M. López Valles. "Modelo de disparidad de carga: Un enfoque con inspiración biológica". Escuela Politécnica Superior de Albacete. UCLM, 2004.
- **Ongoing PhD Thesis:**
  - J. Martínez Cantos: "Lenguajes de descripción de escenas en la tarea de video-vigilancia". ETSII, UNED.
  - J.M. Cuadra Troncoso. "Creación de sensores virtuales multimodales para la monitorización y el diagnóstico en un entorno de vigilancia visual avanzada". ETSII, UNED.
- **DEAs:**
  - V. López Jaquero. UCLM, Year 2004.
  - Enrique Lazcorreta. UCLM, Year 2005.
  - J.M. Gascueña Noheda. UCLM, Year 2006.
  - JJ. Valencia Jiménez. UCLM, Year 2006.
  - E. Folgado Zúñiga UNED, Year 2006.
- **Scholarship holders and Contracted:**
  - J. Martínez Cantos. UNED, Scholarship holder of the Ministerio de Educación y Ciencia (MEC), Year 2005.
  - J.M. Gascueña Noheda. UCLM, Scholarship holder of the Junta de Comunidades de Castilla-La Mancha (JCCM), Years 2006-2009.
  - J.M. González Gómez. UCLM, Contracted of the JCCM, Years 2006-2008.
  - J.L. Ortiz Soria. UNED, Contracted of the MEC, Year 2006.
  - J.L. Herranz Frutos. UNED, Scholarship holder of the MEC, Year 2006.

Patents

- "Procedimiento para describir el comportamiento geométrico en humanos en una escena captada por un sistema de visión artificial, basado en un modelo de bloques y, en especial, orientado a la tarea de video-vigilancia". Authors: M. Rincón, E. Folgado, E. Carmona & M. Bachiller (under consideration).

Publications

- **Journal Papers:**

Journal title	Reference	Status	Impact
Pattern Recognition	[27]	Published	2.153
Image and Vision Computing	[24]	Published	1.383
Int. Journal of Human-Computer Studies	[73]	Published	1.348
Expert Systems with Applications	[16]	Under review	1.236
	[17] [26] [52] [67]	Published	
Pattern Recognition Letters	[9] [37]	Under review	1.138
	[28] [31]	Published	
Fuzzy Sets and Systems	[54]	Under review	1.039
Neurocomputing	[40] [43]	Under review	0.790
Robotics and Autonomous Systems	[1] [61]	Under review	0.777
Applied Intelligence	[44]	Under review	0.569
Lecture Notes in Computer Science	[4] [15] [19] [29] [34] [38] [49] [50] [74]	Published	0.402
Lecture Notes in Artificial Intelligence	[33]	Published	0.302
Scientiae Mathematicae Japonicae	[47]	Published	none
Inteligencia Artificial	[32]	Published	none

- **Conference Papers**

Eleventh International Conference on Computer Aided Systems Theory, EUROCAST-2007	[25] [55] [66]
IEEE International Conference on Advanced Video and Signal based Surveillance, AVSS-2006	[75]
7th International Conference on Intelligent Data Engineering and Automated Learning, IDEAL-2006	[45]
Campus Multidisciplinary in Perception and Intelligence, CMPI-2006	[2] [3] [5] [10] [41] [49] [56] [62]
The Fifth International Workshop on Pattern Recognition in Information Systems, PRIS-2005	[20] [21]
4th Indian Conference on Computer Vision, Graphics and Image Processing, ICVGIP-2004	[35]
IEEE 4th International Conference on Intelligent Systems Design and Application, ISDA-2004	[53]

- **Master Thesis:** [8] [11] [14] [42] [59] [60] [63] [64] [69] [70] [71] [72]
- **Technical Reports:** [18] [39] [51]

Research Collaborations

- Prof. J. Mira collaborates with Dr. Ángel Sánchez and Dr J.J. Pantrigo of the Universidad Rey Juan Carlos in knowledge modelling for a complex object tracking system based on video sequences. Visual tracking is based on a synergic combination of strategies coming from particle filters and population-based metaheuristics.
- Dr. A. Fernández-Caballero collaborates with Dr. J. Pavón of the Universidad Complutense de Madrid and Dr. J.A. Botía of the Universidad de Murcia in visual surveillance scenario modelling by INGENIAS agent-based methodology.

- Dr. A. Fernández-Caballero belongs to the SIMILAR European Network of Excellence ("The European taskforce creating human-machine interfaces SIMILAR to human-human communication").
- Dr. A. Fernández-Caballero is a member of to the Spanish Agent Network agentcities.es ("Creación de un entorno innovador para la comunicación de agentes inteligentes").
- Dr. A. Fernández-Caballero pertains to the recently created Spanish Physical Agent Network RedAF.

Other Activities

- **Conferences and Workshops Organized**
  - Prof. J. Mira served as the General Chairman of 1st International Work-Conference on the Interplay between Natural and Artificial Computation, IWINAC 2005.
  - Dr. A. Fernández-Caballero and Dr. Heinz Hügli organized the Special Session "Natural and Artificial Aspects of Visual Attention" at IWINAC-2005.
- **Invited Speakers**
  - Prof. J. Mira has served as Invited Speaker at ISDA-2004 [53], IDEAL-2006 [45], and EVI-2006 [46].

4 References

[1] Álvarez J.R., Mira J., De la Paz F. & Cuadra J.M. (2007). Centre of area method as basic mechanism for representation and navigation in situated robotics. Under review, Robotics and Autonomous Systems.

[2] Álvarez J.R., Mira J. & De la Paz F. (2006). Reflexiones sobre la utilización de robots autónomos en tareas de vigilancia. CMPI-2006, 10-14 July 2006, Albacete (Spain), pp. 322-331.

[3] Álvarez J.R., Mira J. & De la Paz F. (2006). El método del centro de áreas como mecanismo básico de representación y navegación en robótica situada. CMPI-2006, 10-14 July 2006, Albacete (Spain), pp 103-117.

[4] Álvarez J.R., De la Paz F. & Mira J. (2005). A robotics inspired method of modeling accessible open space to help blind people in the orientation and traveling tasks. Lecture Notes in Computer Science, 3561, pp. 405-415.

[5] Bachiller M., Martínez R., Mira J. & Rincón M. (2006) Niveles de descripción para la interpretación de secuencias de vídeo en tareas de vigilancia. CMPI-2006, 10-14 July 2006, Albacete (Spain), pp 450-462.

[6] Bolles B. & Nevatia R. (2004). A hierarchical video event ontology in OWL. Final Report 2004, ARDA Project.

[7] Bremond F., Maillot N., Thonnat M. & Vu V.T. (2004). Ontologies for video events. INRIA, Research Report, 5189.

[8] Caldas A. (2007). Integración de PDAs en un sistema multiagente de vigilancia. Master Thesis, Advisor: Fernández-Caballero A., June 2007.

[9] Carmona E., Martínez-Cantos J. & Mira J. (2007). A new video segmentation method of mobile objects based on blob-level knowledge. Under review, Pattern Recognition Letters.

[10] Carmona E., Martínez-Cantos J. & Mira J. (2006). Posprocesamiento morfológico adaptativo basado en algoritmos genéticos y orientado a la detección robusta de humanos. CMPI-2006, 10-14 July 2006, Albacete (Spain), pp. 249-261.

[11] Ciudad J.M. (2005). Desarrollo de la ontología del movimiento en secuencias de imágenes. Master Thesis, Advisors: Fernández-Caballero A. & López M.T., September 2005.

[12] Collins R.T., Lipton A.J. & Kanade T. (eds.) (2000). Special Issue on Video Surveillance, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22:8.

[13] Collins R.T., Lipton A.J., Kanade T., Fujiyoshi H., Duggins D., Tsin Y., Tolliver D., Enomoto N., Hasegawa O., Burt P. & Wixson L. (2000). A system for video surveillance and monitoring. Technical Report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University.

[14] Cuenca A.M. (2006). Estudio de las capacidades paramétricas de la segmentación basada en ILCA para secuencias de vídeo. Master Thesis, Advisors: Fernández-Caballero A. & López M.T., December 2006.
[15] Fernández M.A., López-Valles J.M., Fernández-Caballero A., López M.T., Mira J. & Delgado A.E. (2005). Permanency memories in scene depth analysis. Lecture Notes in Computer Science, 3643, pp. 531-536.
[16] Fernández-Caballero A., López M.T. & Saiz-Valverde S. (2007). Dynamic Stereoscopic Selective Visual Attention (DSSVA): Integrating motion and shape with depth in video segmentation. Under review, Expert Systems with Applications.
[17] Fernández-Caballero A., López M.T., Mira J., Delgado A.E., López-Valles J.M. & Fernández M.A. (2007). Modelling the stereovision-correspondence-analysis task by lateral inhibition in accumulative computation problem-solving method. In press, Expert Systems with Applications, 34 (4), December 2007.
[18] A. Fernández-Caballero & López M.T. (2006). D_2^2: Lenguaje multinivel para la interpretación de eventos de video. Technical Report.
[19] Fernández-Caballero A., Mira J., Delgado A.E., Fernández M.A. & López M.T. (2005). Lateral interaction in accumulative computation: Motion-based grouping method. Lecture Notes in Computer Science, 3704, pp. 399-408,
[20] Fernández-Caballero A. (2005). Knowledge-based silhouette detection. The Fifth International Workshop on Pattern Recognition in Information Systems, PRIS 2005, 24-25 May 2005, Miami (USA), pp. 114-123.
[21] Fernández-Caballero A. (2005). Motion direction detection from segmentation by LIAC, and tracking by centroid trajectory calculation. The Fifth International Workshop on Pattern Recognition in Information Systems, PRIS 2005, 24-25 May 2005, Miami (USA), pp. 213-218.
[22] Haritaoglu I., Harwood D. & Davis L.S. (2000). W4: Real-time surveillance of people and their activities. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8), pp. 809-830.
[23] Howarth R.J. & Buxton H. (2000). Conceptual descriptions from monitoring and watching image sequences. Image and Vision Computing, 18, pp. 105-135.
[24] López M.T., Fernández M.A., Fernández-Caballero A., Mira J. & Delgado A.E. (2007). Dynamic visual attention model in image sequences. In press, Image and Vision Computing, 25.
[25] López M.T., Fernández-Caballero A., Fernández M.A., Mira J. & Delgado A.E. (2007). Real-time motion detection from ALI model. Eleventh International Conference on Computer Aided Systems Theory, EUROCAST 2007, 12-16 February 2007, Las Palmas de Gran Canaria (Spain).
[26] López M.T., Fernández-Caballero A., Mira J., Delgado A.E. & Fernández M.A. (2006). Algorithmic lateral inhibition method in dynamic and selective visual attention task: Application to moving objects detection and labelling. Expert Systems with Applications, 31 (3), pp. 570-594.
[27] López M.T., Fernández-Caballero A., Fernández M.A., Mira J. & Delgado A.E. (2006). Visual surveillance by dynamic visual attention method. Pattern Recognition, 39 (11), pp. 2194-2211.
[28] López M.T., Fernández-Caballero A., Fernández M.A., Mira J. & Delgado A.E. (2006). Motion features to enhance scene segmentation in active visual attention. Pattern Recognition Letters, 27 (5), pp. 469-478.
[29] López M.T., Fernández-Caballero A., Fernández M.A. & Delgado A.E. (2005). Sensitivity from short-term memory vs. stability from long-term memory in visual attention method. Lecture Notes in Computer Science, 3562, pp. 448-457.
[30] López M.T. (2004). Modelado computacional de los mecanismos de atención selectiva mediante redes de interacción lateral. PhD Dissertation, ETSII, UNED.
[31] López-Valles J.M., Fernández M.A. & Fernández-Caballero A. (2007). Stereovision depth analysis by two-dimensional motion charge memories. Pattern Recognition Letters, 28 (1), pp. 20-30, Enero 2007.
[32] López-Valles J.M., Fernández M.A. & Fernández-Caballero A. (2006). Conceptos y técnicas de estereovisión por computador. Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial, 9(27), pp. 35-62. AEPIA.
[33] López-Valles J.M., Fernández M.A., Fernández-Caballero A., López M.T., Mira J. & Delgado A.E. (2005). Motion-based stereovision model with potential utility in robot navigation. Lecture Notes in Artificial Intelligence, 3533, pp. 16-25.

[34] López-Valles J.M., Fernández M.A., Fernández-Caballero A. & Gómez F.J. (2005). Stereovision disparity analysis by two-dimensional motion charge map inspired in neurobiology. Lecture Notes in Computer Science, 3704, pp. 459-468.
[35] López-Valles J.M., Fernández M.A. & Fernández-Caballero A. (2004). Bidimensional motion charge map for stereovision disparity analysis. 4th Indian Conference on Computer Vision, Graphics and Image Processing, ICVGIP 2004, 16-18 December 2004, Kolkata (India).
[36] López-Valles J.M. (2004). Modelo de disparidad de carga: Un enfoque con inspiración biológica. PhD Dissertation, Escuela Politécnica Superior de Albacete, UCLM.
[37] Martínez R., Rincón M., Bachiller M. & Mira J. (2007). A model of surveillance task based on the use of a hierarchy of interpretation levels. Under review, Pattern Recognition Letters.
[38] Martínez R., Álvarez J.R. & Mira J. (2005). Key aspects of the Diagen conceptual model for medical diagnosis. Lecture Notes in Computer Science, 3561, pp. 481-488.
[39] Martínez R., Rincón M. & Bachiller (2005). D_2^1: Visión cognitiva y vigilancia. Technical Report.
[40] Martínez-Cantos J., Carmona E., Fernández-Caballero A. & López M.T. (2007). Parametric improvement of lateral interaction in accumulative computation in motion-based segmentation. Under review, Neurocomputing.
[41] Martínez-Cantos J., Carmona E., Fernández-Caballero A. & López M.T. (2006). Mejora paramétrica de la interacción lateral en computación acumulativa. CMPI-2006, 10-14 July 2006, Albacete (Spain), pp.262-273.
[42] Martínez-Cantos J. (2004). Interacción lateral y computación acumulativa en secuencias de vídeo: Implementación y mejora paramétrica. Master Thesis, Advisors: Fernández-Caballero A. & López M.T., September 2004.
[43] Mira J., Delgado A.E., López M.T., Fernández-Caballero A. & Miguel A. Fernández (2007). A unified approach with two neural mechanisms to cope with the visual selective attention task. Under review, Neurocomputing.
[44] Mira J., Delgado A.E., Fernández-Caballero A. & López M.T. (2007). Real-time motion detection starting from the formal model underlying Algorithmic Lateral Inhibition (ALI). Under review, Applied Intelligence.
[45] Mira J. (2006). On some of the neural mechanisms underlying adaptive behavior. 7th International Conference on Intelligent Data Engineering and Automated Learning, IDEAL-2006, Burgos (Spain), pp. 1-15.
[46] Mira J. (2006). La inteligencia artificial como ciencia y como ingeniería. In: 50 Años de la Inteligencia Artificial, XVI Escuela de Verano de Informátiva, EVI-2006, Albacete (Spain), pp 1-12.
[47] Mira J & Delgado A.E. (2006). A cybernetic view of artificial intelligence. Scientiae Mathematicae Japonicae, 64 (2), pp. 331-349.
[48] Mira J., Delgado A.E., López M.T., Fernández-Caballero A. & Fernández M.A. (2006). Una arquitectura bioinspirada para el modelado computacional de los mecanismos de atención visual selectiva. CMPI-2006, 10-14 July 2006, Albacete (Spain), pp. 425-437.
[49] Mira J. (2005). On the physical formal and semantic frontiers between human knowing and machine knowing. Lecture Notes in Computer Science, 3643, pp 1-8.
[50] Mira J. (2005). On the use of the computational paradigm in neurophysiology and cognitive science. Lecture Notes in Computer Science, 3561, pp. 1-15.
[51] Mira J. & Fernández-Caballero A. (2005). D_1^1 + D_1^2: Descripción del modelado de tareas en AVISA. Technical Report.
[52] Mira J., Delgado A.E., Fernández-Caballero A. & Fernández M.A. (2004). Knowledge modelling for the motion detection task: The algorithmic lateral inhibition method. Expert Systems with Applications, 27 (2), pp. 169-185.
[53] Mira J. & Delgado A.E. (2004). From modeling with words to computing with numbers. Proceedings of the IEEE 4th International Conference on Intelligent Systems Design and Application, ISDA-2004. Plenary Lecture.
[54] Moreno-Garcia J., Rodriguez-Benitez L., Fernández-Caballero A. &. López M.T. (2007). Motion analysis based on permanency values and fuzzy logic. Under review, Fuzzy Sets and Systems.

[55] Moreno-Garcia J., Fernández M.A., Fernández-Caballero A. & López M.T. (2007). Tracking deformable objects by grouping permanency values. Eleventh International Conference on Computer Aided Systems Theory, EUROCAST 2007, 12-16 February 2007, Las Palmas de Gran Canaria (Spain).
[56] Moreno-Garcia J., Rodríguez-Benítez L., Fernández-Caballero A. & López M.T. (2006). Análisis del movimiento basado en valores de permanencia y lógica difusa. CMPI-2006, 10-14 July 2006, Albacete (Spain), pp. 725-738.
[57] Nagel H.H. (2004). Steps towards a cognitive vision system. AI Magazine, 25 (2), pp. 31-50.
[58] Neuman B. & Weiss T. (2003). Navigating through logic-based scene models for high-level scene interpretations. 3rd International Conference on Computer Vision Systems, ICVS-2003. Lecture Notes in Computer Science, 2626, pp. 212-222.
[59] Ortiz J.L. (2006). Mejoras a los procesos de segmentación y seguimiento en atención selectiva visual dinámica. Master Thesis, Advisors: López M.T. & Fernández-Caballero A., July 2006.
[60] Parra I. (2006). Visualización y grabación de imágenes digitales con la librería ActiveMIL de Matrox. Master Thesis, Advisors: López M.T. & Fernández-Caballero A., December 2006.
[61] Pavón J., Gómez-Sanz J., Fernández-Caballero A. & Valencia-Jiménez J.J. (2007). Development of intelligent multi-sensor surveillance systems with agents. Under review, Robotics and Autonomous Systems.
[62] Pavón J., Gómez-Sanz J., Valencia-Jiménez J.J. & Fernández-Caballero A. (2006). Desarrollo de un sistema inteligente de vigilancia multi-sensorial con agentes software. CMPI-2006, 10-14 Julio 2006, Albacete (España), pp.377-388.
[63] Pérez D. (2005). Diseño e implementación de la interfaz gráfica de una centralita de alarmas. Master Thesis, Advisors: Fernández-Caballero A. & López M.T., September 2005.
[64] Pérez-Jiménez R. (2005). Análisis y comparación de algoritmos para la detección de movimiento en secuencias de imágenes. Master Thesis, Advisors: Fernández-Caballero A. & López M.T., July 2005.
[65] Regazzoni, C.S., Fabri, G. & Breñaza, G. (eds.) (1999). Advanced Video-Based Surveillance Systems, Kluwer Academic Publishers, Dordrecht.
[66] Rincón M. & Martínez-Cantos J. (2007). An annotation tool for video understanding. Eleventh International Conference on Computer Aided Systems Theory, EUROCAST 2007, 12-16 February 2007, Las Palmas de Gran Canaria (Spain).
[67] Rincón M., Bachiller M. & Mira J. (2005). Knowledge modeling for the image understanding task as a design task. Expert Systems with Applications, 29 (1), pp. 207-217.
[68] Robertson N. & Reid I. (2006). A general method for human activity recognition in video. Computer Vision and Image Understanding, 104, pp. 232-248.
[69] Rodríguez J.M. (2007). Descripción conceptual de objetos en movimiento en secuencias de vídeo. Master Thesis, Advisor: Fernández-Caballero A., July 2007.
[70] Ruiz R.C. (2007). Navegación autónoma del robot Pioneer 3AT. Master Thesis, Advisor: Fernández-Caballero A., July 2007.
[71] Saiz I. (2006). Biblioteca de algoritmos de vigilancia para el procesador de visión Odyssey. Master Thesis, Advisors: López M.T. & Fernández-Caballero A., July 2006.
[72] Saiz-Valverde S. (2005). Atención visual selectiva con imágenes estéreo. Master Thesis, Advisors: Fernández-Caballero A. &. López M.T., December 2005.
[73] Taboada M., Martínez D. & Mira J. (2005). Experiences in reusing knowledge sources using Protégé and Prompt. International Journal of Human-Computer Studies, 62 (5), pp. 597-618.
[74] Taboada M., Des J., Martínez D. & Mira J. (2005). Aligning reference terminologies and knowledge bases in the health care domain. Lecture Notes in Computer Science, 3561, pp. 437-446.
[75] Valencia-Jiménez J.J. & Fernández-Caballero A. (2006). Holonic multi-agent systems to integrate independent multi-sensor platforms in complex surveillance. IEEE International Conference on Advanced Video and Signal based Surveillance, AVSS-2006, 22-24 November 2006, Sydney (Australia), pp. 49.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

ASILAN: Interfaces gráficas y algoritmos para aplicaciones basadas en la simulación de luz artificial y natural TIN2004-07672-C03

Ignacio Martín Campos *
Instituto de Informática y Aplicaciones
Universidad de Girona

Abstract

This project is aimed at making the most of the experience and knowledge in the field of global illumination acquired in previous projects by the participant research groups. The goal is to design and develop graphics interfaces applied to sustainable lighting design. Our proposal aims at applying the know-how of the research groups to the improvement of lighting design at different levels: from interior lighting design to urban environments. In order to develop usable prototypes it is very important to move away from the simplifications that are usually assumed in global illumination computations. These simplifications usually lead to unacceptable results that cannot be used in real-world applications.

The project deals with the problem of simultaneous and efficient use of artifical and natural light in a way that the tools developed allow for a optimum use of natural light resources, and at the same time, to improve the luminaire design process to optimize its performance. This will bring benefits in energy savings and light pollution reduction.

Another important issue is the use of optical properties captured from real-world materials that are used in the simulations. This is the only way to guarantee the usability of the results. There are two main aspects to consider: the obtention of real-world data from measurements, and the efficient representation of this data.

Finally, due to the high computational cost of most of the techniques proposed, the project also deals with the use of parallel processing and super-computing technologies.
Keywords:

1 Project Goals

The project goals can be summarized as follows:

1. Design and development of software tools for inverse reflector design. The idea is to improve the design process, and the optimization of the lighting efficiency. The benefits should be energy savings and lighting pollution reductions. This goal has three tasks devoted:

*Email: ignacio.martin@udg.es

- Task Z-1: Inverse reflector design based on a theoretical inverse reflector calculation using a direct search based technique.
  Schedule: months 1-36
- Task Gi-2.2: the use of near-field light sources [5] is a must if we want reliable results from the lighting simulations. This task aims at simplifying the near-field representation to make them usable into global illumination algorithms.
  Schedule: months 1-24
- Task Gi-1.1: starting from the work developed by [4] the goal is to overcome the simplifications assumed in this work. The goal is to include near-field light sources, complex material representations, and more flexible surface representations.
  Schedule: months 1-36

2. Design and development of software tools for integration if artificial and natural lighting. The goal is to provide tools for sustainable building, with applications ranging from interior design to large area urban design. This goal has three tasks devoted:

   - Task Gr-1.1: use of Monte-Carlo techniques for lighting simulation in arquitectonic environments, and its application to artificial and natural lighting.
     Schedule: months 1-6
   - Task Gr-1.2: lighting component analysis for complex environments: direct solar illumination, artificial lighting, indirect skylight illumination, ... etc.
     Schedule: months 6-24
   - Task Gi-1.2: natural lighting in complex environments. The goal is to produce semi-interactive tools for speeding up the design process in building design.
     Schedule: months 1-36

3. Light transport optimization using high-performance computing (parallel computing, GPUs, GRID technologies, ...). The idea is to optimize the lighting computations previously developed.

   - Task Gr-2.2: Study of time efficiency and implementation of several optimized algorithms for advanced density estimation in photon-mapping related algorithms. This task includes both theoretical complexity analysis and heuristic time cost evaluation. Implementation of different variants and evaluation of benefits.
     Schedule: months 1-18
   - Task Gr-2.4: Design and implementation of parallell algorithms for global illumination Monte-Carlo techniques, using PC clusters and advanced graphics hardware. Evaluation of different options for load balancing and algorithm vectorization in the context of Global illumination algorithms.
     Schedule: months 1-36
   - Task Z-2: Code optimization for Global Illumination algorithms, and parallelization on shared and distributed memory systems.
     Schedule: months 1-36

4. Optimization of complex material models in the context of global illumination methods.

   - Task Gr-2.1: efficient importance sampling of Bidirectional Reflectance Distribution Functions (BRDFs) in order to use arbitrary material representations.
     Schedule: months 1-18

5. Design and development of graphical interfaces to allow users the use of the techniques developed into the project. Assess of usability.

   - Task Gr-3.1: design and implementation of a interface to allow multiple scene formats to be loaded and edited, including optical properties (BRDFs), and multiple types of light sources.
     Schedule: months 6-18
   - Task Gr-3.2: development of software tools to integrate all subsystems for light source design, scene modeling, lighting simulation and lighting assessment.
     Schedule: months 18-36

2 Achievements

The project achievements are herein summarized and classified into the global project goals:

2.1 Inverse Reflector Design

Task Z.1: Inverse reflector design based on a direct search technique

This task is scheduled from month 1 to 36. The goal of this task is to design and implement an algorithm for the inverse reflector design, starting from a direct search technique. Direct searches carry out sequential checks of possible solutions, defining a given strategy to compute the next trial solution. In this way, they try to get a feasible solution in a faster manner. The task accomplish the search, research and implementation of a feasible direct search method, together with a theoretical inverse reflector calculation to provide a test bed for the validation of the algorithm.

- *SubTask Z-1.1: Theoretical inverse reflector calculation*

  In order to validate the direct search method for the inverse reflector design, trial reflectors have to be computed. In this subtask we have focused on obtaining theoretical solutions. We have selected some near-field light sources and scene settings in order to calculate in a theoretical way the shape of the reflector.

  Since the theoretical inverse calculation for a general reflector is complex, we have made some simplifications and assumptions in the settings of the testes. We calculate axially symmetrical reflectors. In this way we deal with a usual shape in industrial reflector design. Working with the profile of a reflector of revolution, together with a near-field light source, we can also spend more efforts in obtaining more accurate solutions. For light sources we use isotropic point light sources, which do not occlude the lighting conditions of the reflector to calculate.

We have used two theoretical methods to inverse calculate the shape of the reflectors, one analytical and other numerical. Analytical solutions are computed by integrating the reflection's law. Numerical solutions are computed comparing the available fluxes, computed from the light source, with the needed fluxes, defined for the near-field light source.

The outcome of this subtask is a theoretical computed test bed which will act as the income of the SubTask Z-1.4, validating the direct search technique selected in SubTask Z-1.2 and implemented in SubTask Z-1.3.

- *SubTask Z-1.2: Search of a numerical method for the inverse reflector design*

  Parallel to the carrying out of the SubTask Z-1.1, SubTask Z-1.2 was performed. The aim of this task is the search of a feasible algorithm to solve the inverse reflector design, which is an optimization problem.

  The selected algorithm was the *Hooke-Jeeves* method, also named *Pattern Search*. The algorithm try to minimize (optimize) an objective function, which computes the error between the desired near-field light source and the computed by the trial solution. We use a mean squares error function since it is a very common function in data comparison.

  The optimization process done by the Hooke-Jeeves method is achieved by the sequential modification of a given initial guess reflector. We use NURBS surfaces to represent the reflectors due to two reasons. First, the use of NURBS are widely extended in the industrial CAD design. Second, the use of a NURBS instead of a triangle mesh to represent a reflector, simplifies the operation of the Hooke-Jeeves algorithm. In this way the algorithm only has to deal with few control points which represent the NURBS, and not with thousand vertices in a triangle mesh.

  In spite of the speedup achieved by using NURBS, we have introduced some simplifications in order to trade-off between accuracy and time of computation. For the moment, the algorithm deals only with axially symmetrical reflectors, defined as a revolution NURBS surface from a given NURBS curve (the revolution profile). In this way, a profile curve of three to ten control points is enough to design a detailed reflector of revolution. Again, isotropic light sources are used to avoid occluding lighting. Finally, the reflector is designed with a perfect specular material.

  In order to compute the near light field we use global illumination techniques, specifically a method named Photon Mapping. This technique is a physically based simulation of the light transport.

- *SubTask Z-1.3: Implementation of the numerical method for the inverse reflector design*

  The outcome of this subtask is the implementation and integration of the selected optimization Hooke-Jeeves method into a rendering library for fast physically based simulation of the global illumination. The design, optimization and parallelization of this rendering system is described in detail in Task Z-2. In this SubTask Z-1.3 we also implement a comprehensive library to manage NURBS surfaces and curves. The programming language selected was C++, giving a cross-platform support for its compilation.

  The implementation of the optimization method is finished, but we are enhancing it by implementing it as a generic solver for optimization problems, decoupling the solver from the kind of optimization problem to solve.

- *SubTask Z-1.4: Validation and parallelization of the numerical method for the inverse reflector design*

  The goal of this subtask is to validate the direct search method, called Hooke-Jeeves, by the comparison of the obtained approximate results with the theoretical calculated results. Therefore, this subtask will have as income the test bed calculated in SubTask Z-1.1. The development of this subtask also includes the parallelization of the Hooke-Jeeves algorithm.

  As initial steps in this subtasks we are on the research of the behavior of the Hooke-Jeeves algorithm. After this early stage a complete validation and the parallelization, both at CPU level (with threads) and at Cluster level (with MPI), will have to be performed.

Task Gi-2.2: use of near-field representations

This task is scheduled from month 1 to 24, and the final goal was to develop a new representation for near-field measurements of light sources. The goals has been fully accomplished and the results are shown here [1]. This report presents a new technique for data compression for near-field measurements. The result is a much more compact representation of the lighting distribution, and allows the light source to be used in different global illumination algorithms.

Task Gi-1.1 : inverse reflector design

This task is scheduled from month 1 to 36. The main goal is to design an implement a prototype for inverse reflector design, overcoming some of the limitations of the state of the art techniques. In a first step, we have evaluated the current solutions in this area. The results have been published here [21].The conclusions presented in this report show that the state of the art solutions to the problem of inverse reflector design are quite simple, and all of them make strong assumptions that simplify the problem in order to make it computationally affordable.

We have been working in a more general approach to the problem in order to make as less simplifications as possible. This work was finished by the end of 2005 and the latest results have been published here [4].

In a joint work with University of Zaragoza, the algorithm presented in [4] has been re-implemented to work in a parallel computer. Inverse design are computationally very expensive and one of the way to improve the speed of the calculations is to modify the algorithm so it can work in super-computers. The results of this work have been published here [13].

However, the algorithm developed in this task is still based on some simplifications that are not suitable for industry-level results. Those simplifications are:

1. Use of far-field light sources. This type of representations do not yield accurate results when the illuminated surfaces are close to the source, and this is happens when dealing with reflectors.

2. Use of low resolution polygonal meshes as reflector representation.

In the last year we have been working on finding out solutions for these limitations. In special, we are considering the use of near-field representation for the bulbs, and the use of NURBS as a surface representation for the reflector. These two issues have come up because our relationship with HELLA [6], a major manufacturer of light and electronics for the automotive industry. After

several meetings, the University of Girona has signed a confidentiality agreement with this company to exchange data and knowledge in order to cooperate to develop new solutions for inverse reflector design. We have just received the first data in December 2006 and we have just begun to test it with our new ideas.

In this task we are slightly delayed by two reasons: we have spent too many time on Task Gi-2.2, and our cooperation with HELLA has changed slightly our work plan, but in a good sense. We have now a very precise idea of what is needed in the industry, and how far a big company as HELLA has gone in the area of inverse reflector design. The main conclusions that have to guide our work in the last year are:

- The use of near-field light sources is a must, Otherwise the results cannot be accurate.
- The use of NURBS as a surface representation is almost mandatory since it is the standard in high-end CAD software used by these companies. That means that we cannot longer use our previous approach of vertex optimizing [4].
- The lighting restrictions for automotive optical sets are very specific. We will have to develop new error criteria to meet these specific needs.

2.2 Natural and Artificial Lighting

Task Gi-1.2: natural lighting in complex environments

In this task we have focused our work on providing tools for the improving the lighting design process of architectural environments. In particular, we want to develop new techniques that effectively can be used into building design to optimize the use of natural lighting. The impact of such tools is clear: less energy consumption, less costs, and improved visual comfort.

The first approach was to develop a scan-line based method that projects light from sky patches and produces an hierarchical subdivision of the surfaces that adapts to the illumination changes. The results of the work were published here [22].

However, the surface hierarchy hast to be computed using a software scan-line algorithm that makes the computation too slow (tens of minutes) for an interactive tools. Thus, the next goal is to come up with a new method that do not compute illumination, but visibility. This will allow to use the visibility information with different sky configurations without any recomputation. Moreover, we want to perform this visibility computations in less than a minute, so the geometry of the building can be also changed and at the same time keep the whole process interactive.

2.3 High-Performance Computing

Task Z-2: Code optimization for Global Illumination algorithms, and parallelization on shared and distributed memory systems.

For this task we have started from our existing rendering system ALEPH. Its main features are:

- basic magnitude in ALEPH is spectral radiance, and color is handled with a spectral representation
- can manage scenes defined with parametric surfaces, from triangles to bicubic patches, defined implicitly

- the Radiance Equation is solved using MonteCarlo methods, where sampling is performed with rays
- results are numerical values for spectral radiance or irradiance

From this system we built a new system called SIRIG, where we can introduce all this enhancements. To compare with the old system we choose a standard model from the Stanford Scanning repository and rendered it with the old and the new system.

- *SubTask Z-2.1: Algorithmic and code-level optimization of GI algorithms, focused on MonteCarlo ray tracing and photon mapping*

  Many data structures in ALEPH have been redesigned or reworked in order make them safe for the next steps (optimization, parallelization). Some restrictions have also been imposed to the design, like abandoning parametric surfaces in favor of triangles. This allows to better streamline ray tracing operations and to simplify them.

  Some new data structures have been introduced, like the 'light-cache', to allow new simulation algorithms like photon mapping.

  And other have been added in parallel with existing ones to build algorithms better than those already present in ALEPH, like the kd-tree for spatial indexing, which behaves much better than voxel grids or octrees.

  We performed some tests to see if it would be feasible to use vectorized instructions offered by nowadays processors (SSE,MMX, 3DNow). This preliminary tests showed that this instruction sets were very powerfully but also very limited in their use. They give their full potential only for certain problems and data structures, and were not suitable for our design.

  With this software enhancements we got speedups in the 5x to 12x range.

  Research work using this results has been published in international journals ([7],[8],[9]), international congresses ([12],[14],[15], [16],[17],[18]) or is still in review process ([10],[11]).

- *SubTask Z-2.2: Parallelization on shared memory systems*

  There are two main time-consuming algorithms in the new system: the photon shooting pass and the ray tracing pass. We have parallelized both on shared memory systems, using the available system facilities.

  To test our implementation, we have used both Linux and Windows, basing the implementation on the native multiprocessing facilities (POSIX Threads for Linux and native WinThreads for Windows). The implementation spawns several threads for each parallel task we want to perform, and the work for this task is split in subtasks or jobs, that all the threads can pull from a shared work queue (in a master-slaves design). Some SMP-safe data structures have been implemented for this task, like concurrent queues, locks or monitors (on top of system calls).

  The system has been evaluated on several hardware types, to detect possible bottlenecks like excessive locking. We have tested it on different parallel architectures Intel HyperThreading Pentium4, Intel Core Duo, AMD Turion Dual Core, and AMD Quad Opteron. Efficiency of the parallel version of the software highly depends on the hardware. In our experiments it could go from 60% to 95%.

If we join software optimization and parallel processing, we can get up to a 50x speedup factor on a Quad Opteron box.

Research work using this results has been published in international congresses ([13], [19]) or is still in review process ([20]).

- *SubTask Z-2.3: Parallelization on distributed memory systems*

  The data structures and the load leveling algorithms have been prepared to be used on a distributed system.

  The hardware system has been designed. We choose to use the same systems described in SubTask Z-2.2 to build an initial cluster, that will grow in the near future.

  Some initial tests have been done for the parallelization over a cluster. The system is built for INDAL, SA, It is build of one front-end and six worker nodes, which use Pentium4 processors. The parallelization just spreads the rendering work as individual frames for an animation, no intra-frame parallelism is used. Our main goal is to use intra-frame parallelism to higher frame rates.

2.3.1 Task Gr-2.2. Theoretical and empirical study of the different Density Estimation algorithms,

In this task we focused on advanced density estimation techniques for photon-mapping related algorithms like density estimation on the tangent plane (DETP) [25],[26] or ray-maps [32]. While these algorithms have shown to greatly improve radiance reconstruction from photon histories, they introduce high computation time costs because they require a huge number of ray-disc or ray-ball intersection tests, instead of the simpler point-in-ball query which is needed for basic photon-mapping.

In order to reduce computation time of these algorithms we have designed and implemented two techniques: *Sphere-Cache*, which on a list of spheres which allow fast discarding of potential ray-disc intersection tests, and *Disc Indexing*, a more standard approach which uses octrees to build a spatial index with discs referenced from leaf nodes.

A study of the time complexity of raw DETP, Sphere Cache and Disc Indexing has been made. Integral geometry was used to achieve results which are scene-independent. The formula for the expected computation time of each technique as a function of the complexity of the scene, the illumination and other parameters of the simulation allows us to choose the most efficient technique in each case. Testing the algorithms for different scenes show a good correlation between expected and measured computing times. Results have been presented in WSCG'2006 conference [29], and in the Eurographics Spanish chapter annual conference [28].

This study was later applied in the context of interactive rendering, by using a software system capable of radiosity computation at interactive rates for scenes with moving objects. The characteristics of interactive rendering give us opportunities for more efficient algorithms by reusing information from previous frames [27]. The theoretical study proved useful to pinpoint the most suitable algorithm in this context. Results have been presented in WSCG'2007 conference [30].

2.3.2 Task Gr-2.4 - Usage of advanced hardware and computer networks

Generating photo-realistic images is a compute-intensive task which requires efficient techniques and software/hardware architectures to be accomplished in an affordable manner. In this task

we deal with high computation time involved in physically based rendering. We have explored the feasibility of solutions which take advantage of modern high performance hardware and fast computer networks. This are the three hardware environments we have addressed:

- Single instruction multiple data (SIMD) capabilities in standard CPUs (SSE instructions sets)
- Advanced programmable graphics hardware (GPUs) with multiple floating-point units (FPUs)
- High bandwidth local area networks with concurrently processing PCs (PCs clusters)

Regarding GPUs, we have implemented the sphere cache technique for advanced density estimation [25],[26] using modern programmable graphics cards. We also developed reference and highly optimized CPU implementation, which makes use of SSE instructions and coherent memory accesses.

We have analyzed execution times from both implementations. GPU implementation takes advantage of multiple FPUs, however it suffers from (a) relatively slow data transfer rate between RAM and GPU memory and (b) vectorized programming requirements, which makes it hard to adapt algorithms originally designed for sequential CPUs. Design of GPU programs should focus on minimization of data transfers and efficient data layouts in texture memory.

Our GPU implementation run times are comparable to optimized, SSE-based CPU implementation ones for cards manufactured in 2005 (nVidia 7800 series). However, GPUs vendors are nowadays increasing the performance of these class of hardware, for instance by increasing the number of FPUs per card, or by increasing GPUs cache memory size and bandwidth. We have run our GPU implementation on a recent card model (nVidia 8800 series), and for this implementation, we have obtained a 5x speedup factor with respect to the optimized CPU implementation, thanks to improved hardware capabilities. As a result, we conclude that GPU implementation is a valuable tool for producing high time efficient rendering systems, as it is the case for our density estimation algorithm.

Regarding PC clusters, we have developed a rendering system which runs concurrently on multiple PCs running Linux. Processes communicate via MPI library. At each node, the system takes advantage of SSE instructions for efficient ray-triangle intersections. We have a cluster where each node is equipped with a GPU, thus we plan to take advantage of that in order to further reduce running times.

2.4 Complex Materials

2.4.1 Task Gr-2.1: Efficient Importance sampling of BRDFs

This task is related to the problem of BRDF sampling in Monte-Carlo based rendering algorithms (both those based on path tracing from the observer and those based on photon simulation from light sources). These class of algorithms need to use a probability density function (PDF) which correctly matches each surface BRDFs. This means that, in absence of information about irradiance, the PDF function used should be proportional to or almost proportional to the BRDF, in order to avoid high variance or error.

Moreover, a time efficient direction generator algorithm is also required. This algorithm must output sampling directions distributed according to the PDF. There is no single analytic PDF which is well fitted for arbitrary BRDFs (including a wide set of analytically defined BRDFs and tabulated or measured BRDFs).

We have designed and implemented a new sampling method with reduced variance, which is based in a PDF which is constructed for each BRDF in the scene, as other authors have done [33]. In our case, the PDF is stored as a quadtree with more resolution in those areas of the sphere where BRDF exhibits more changes. Direction generation algorithm uses a traversal process over this tree, and this traversal can be efficiently implemented so its time cost is minimized. At leaf nodes, efficient rejection sampling is done (with an *a priori* bounded number of trials because of BRDFs variation is bounded in each leaf node). This ensures that the PDF is exactly proportional to the BRDF and can be sampled fast.

We have implemented this PDF and integrated it in a Monte-Carlo rendering system. We have compared the error (due to variance) produced by our algorithm with that obtained from standard simple analytic PDFs, for scenes with various well known analytic BRDFs. Results from these simulations show that, for any given fixed output error, computing time is equal or even less with these new algorithm than the one obtained with standard PDFs which had been manually fitted to the material BRDFs.

2.5 Graphical User Interfaces

2.5.1 Task Gr-3.1: Interface for editing scene properties

In this task we sought solutions to the needs for software which enables its users to describe the reflective properties of any given 3D scene. While nowadays there are a number of both free and commercial 3D modeling packages (like Blender or 3D Studio Max) available, these packages usually lack the ability to assign advanced material properties which are required as an input for physically based renderers. Each rendering system uses its own format for describing material properties. Thus, we needed a system which enables us to assign BRDFs, textures, etc... to the objects in a scene. Moreover, users should be capable of visually evaluating the impact of different values of BRDFs parameters on the look of final images, thus allowing them to select the values which more fits their needs.

In this line, we have already implemented and tested these software systems:

1. **BRDFs assignment tool**
   We have developed a Java3D multiplatform application which imports 3D scene files (i.e. Autocad .dxf, Kinetix 3D Studio, VRML, Wavefront .obj, etc.) and allows its users to edit the BRDF properties and also to assign different instances of the function to different surfaces in the scene. Output scene file is generated with this tool in a format suited for our renderers, that also includes definitions of extended light sources, with information about their location and characteristics.

2. **BRDFs visualization tool**
   In addition another tool has been implemented to visualize the BRDF applied to a scene in real time using programmable graphics hardware (GPU). Reflectance functions have been implemented as GLSL (OpenGL Shading Language) shaders. The application allows users to interactively modify shader's parameters, while an image is displayed which shows the impact of these changes on the look of the geometry.

3. **BRDFs parameters editing tool**
   A BRDF parameters editor is included in the previously described system (although this editor

can be used as a standalone application). This software allows to interactively define and control the appearance of a material with more than fifteen different reflectance functions. A display of BRDFs shape is available to users, both as a flat 2D plot, or as a full 3D depiction. BRDFs parameters can be interactively modified, with immediate feedback to users, who can evaluate the impact of that change on BRDF shape.

2.6 Other Activities

- **University of Girona**

  There has been a research line that began before the project started, and even it is not strictly related to any task, it has produced very good results, and it has provided knowledge and new ideas for Task 1.1 (inverse reflector design). We have worked in interactive reflection and refractions on curved objects. This work has produced two publications [23, 24].These two publications are a joint work with REVES research group of the INRIA Sophia Antipolis center, in France. This work has allowed us to get a deep understanding of the GPUs architecture and has the method presented can be adapted to provide a very efficient solution for inverse reflector lighting calculations.

- **University of Granada**

  From February to July 2005 Miguel Lastra did a five months research stay at Technical University of Vienna, where he collaborated with members of Computer Graphics and Algorithms Research department leaded by Werner Purgathofer, in the context of research project entitled *treelumination*. Miguel did adapted our previously implemented DETP software to GPU architectures, in close collaboration with Dr. Michael Wimmer and Dr. Stefan Jeshke.

  Starting at December 2006, Rosana Montes has done a three months research stay in the Institute of Software and Multimedia-Technology (Technische Universitat Dresden) in collaboration with Computer Graphics and Visualization chair. Collaboration has been closer to Professor Stefan Gumhold and Sören König, extending their scanner 3d setup for dynamic acquisition and also working in an estimate of the BRDF function of simple objects.

3 Results

3.1 Students

- University of Girona

  There are three Ph.D. students working for the project right now. Pau Estalella got a FI grant from Ministerio de Educación y Ciencia on 2003 and he has been involved in the project since the beginning. He is expecting to defend his thesis on September 2007. Albert Mas got a grant from the University of Girona and he began to work in the project since January 2006, and his work is related to tasks Gi-1.1 and Gi-2.2. He is expected to finish hi thesis on 2008. Carles Bosch has been hired by the project since June 2005 and he is working in task 1.2. He is writing his Ph.D now while working in the project.

- University of Granada

There are currently two Granada University's Ph.D. students involved in the project, both are University staff members, and they are involved in undergraduate teaching at the *Departamento de Lenguajes y Sistemas Informáticos*, in Computer Science related degrees. Rosana Montes is *Profesora Asociada* (although she has been just accredited to become *Profesora Colaboradora*), she is working on BRDF sampling algorithms, which is her Ph.D.'s main subject. She plans to defend her Ph.D. on 2007.Rubén García is currently *Profesor Ayudante*, his research is related to the study of time complexity and efficiency of several algorithms for advanced density estimation for photon-mapping related Global Illumination algorithms. He expects to finish his Ph.D. on 2008.

- University of Zaragoza

  There are two Ph.D. students working for the project right now. Adolfo Muñoz is Profesor Ayudante of 2nd year, and he has been involved in the project since the beginning, and works mainly in the Global Illumination part of the project. He is working on his thesis and expects to finish it by end of 2008. Oscar Ansón is Profesor Ayudante of 2nd year, he has been involved in the project since the beginning, and works mainly in the Inverse Reflector Design part of the project. He is working on his thesis and expects to finish it by end of 2008.

3.2 Technology Transfer

- University of Girona

  The University of Girona has signed two collaboration agreements with three companies.

  Espacio Solar is a SME company that elaborates natural lighting projects and also is a manufacturer of passive lighting devices. At this moment we are developing a tool for interarctive natural lighting design.

  Fundación Ductil Benito is a company that manufactures luminaires for urban lighting. We are collaborating in the field of inverse reflector design.

  HELLA is a world leader in automotive lighting manufacturing and we are collaborating to find new specific solutions for inverse reflector design in teh car industry.

- University of Zaragoza

  The University of Zaragoza mantains a collaboration agreement with the company INDAL Iluminación Técnica, SA (Valladolid). As a result of this collaboration, a small PC cluster has been built for INDAL to test and transfer paralleization algorithms and technology on an industrial lighting system.

3.3 Research Collaboration outside the project

- University of Girona

  We have collaborated with the REVES group of the INRIA Sophia-Antipolis research center in France. Group leader George Drettakis has participated in two of the publications of the project [2, 3].

- University of Granada

  José Miguel Mantas is involved in active collaboration with other research groups in order to explore the benefits of PC clusters and graphics hardware for general purpose computation.

  In close collaboration with members of the research project *Nonlinear diffusion and kinetic equations: asymptotic behaviour and numerical approximation* leaded José Antonio Carrillo (ICREA-UAB) we have developed an strategy to obtain efficient implementations of Implicit -Explicit Runge-Kutta methods (IMEXRK) for PC clusters. Results have been published in the Euro-par conference [31].

  Members of EDANYA (Differential Equations, Numerical Analysis and Applications Research Group - University of Málaga) have developed optimized parallel numerical software based on finite volume techniques to simulate shallow water systems on PC clusters and using SSE instructions sets. We have jointly analyzed ways to implement that software on GPUs and GPUs clusters, which can further reduce computation times.

- University of Zaragoza

  We have established a contact for collaboration with the MOAR (Mother Of All Renderers) Project. This project aims to develop the next–generation rendering system, to supercede other systems like Radiance. Main leadership of the project is carried by Alan Chalmers (currently in Bristol University) and Greg Ward. The group showed a big interest about our system.

References

[1] *Compression and Importance Sampling for Near-Field Light Sources.* Albert Mas, Ignacio Martín, Gustavo patow. Submitted to *Computer Graphics Forum*.

[2] *Accurate Interactive Specular Reflections on Curved Objects.* P.Estalella, I.Martin, G.Drettakis, D.Tost, O.Devillers, F.Cazals. Proceedings of the Vision, Modeling and Visualization 2005, Erlangen, Germany, November 2005.

[3] *A GPU-driven Algorithm for Accurate Interactive Reflections on Curved Objects.* P.Estalella, I.Martin, G.Drettakis, D.Tost. Eurographics Symposium on Rendering 2006, Cyprus, June 2006.

[4] *User-Guided Inverse Reflector Design.* Gustavo Patow, Alvar Vinacua, Xavier Pueyo Accepted for publication at *Computer & Graphics*

[5] *Near-Field Photometry Measuring and Modeling Complex 3-D Light Sources.* Ian Ashdown. ACM SIGGRAPH '95 Course Notes - Realistic Input for Realistic Images

[6] Hella KGaA Hueck & Co. *http://www.hella.com*

[7] *Simulation of Atmospheric Phenomena.* D. Gutiérrez , F.J. Serón, A. Muñoz, O. Ansón. Computer & Graphics, vol. 30, no. 6, pp. 994–1010, 2006.

[8] *A Survey on Participating Media Rendering Techniques.* E. Cerezo, F. Pérez, X. Puedo, F.J. Serón, F.X. Sillion. The Visual Computer, vol. 21, no. 5, pp. 303-328, 2005.

[9] *Dust and light: predictive virtual archaeology*. D. Gutiérrez, V. Sundstedt, F. Gomez, A. Chalmers. Journal of Cultural Heritage. Accepted on 2006, pending of publication.

[10] *Visualizing ocean optics*. D. Gutiérrez , F.J. Serón, A. Muñoz, O. Ansón. IEEE Transaction on Visualization and Computer Graphics. Submitted, in review.

[11] *Simulating Global Illumination*. J. López, D. Gutiérrez , J. Fandós, F.J. Serón, E. Reinhard. SIGGRAPH-07. Submitted, in review.

[12] *Efficient Physically-Based Simulation of Non-Linear Media*. A. Muñoz, D. Gutierrez, F. Serón. GRAPHITE 2006, 4th International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia, pp. 97-105, 2006. ISBN: 1-59593-564-9. Ed: Stephen N. Spencer, ACM SIGGRAPH.

[13] *Parallelization of Inverse Design of Luminaire Reflectors*. J.A. Magallón, G. Patow, F. Serón, X. Pueyo. Eurographics Symposium on Parallel Graphics and Visualization, pp. 99-107, 2006. ISBN: 3-905673-40-1. ISSN: 1727-348X. Ed: A. Heirich, B. Raffin and L.P. dos Santos, Eurographics Association.

[14] *Efficient selective rendering of participating media*. O. Anson, V. Sundstedt, D. Gutiérrez, A. Chalmers. ACM Symp. on Applied Perception ob Graphics and Visualization (APGV06), pp. 135-142, 2006. Ed: E. Reinhard.

[15] *Non-Linear volume photon mapping*. D. Gutiérrez, A. Muñoz, O. Ansón, F.J. Serón. Eurographics Symposium on Rendering, pp. 291-300, 2005. ISBN: 3-905673-23-1. Ed: K. Bala, Ph. Dutré in cooperation with ACM SIGGRAPH, Eurographics Association.

[16] *Rendering Ghost Ships and Other Phenomena in Arctic Atmospheres*. D. Gutiérrez, F.J. Serón, A. Muñoz, O. Ansón. WSCG-05, (short paper) pp. 73-76, 2005. ISBN: 80-903100-9-5. Ed: University of West Bohemia.

[17] *Perception-Based Rendering: Eyes Wide Bleached*. D. Gutiérrez, O. Ansón, A. Muñoz, F.J. Serón. Eurographics 05, (short paper) pp. 49-52, 2005. ISSN: 1017-4656. Ed: Eurographics Association.

[18] *Predictive sunlight simulation for the remodelling of the Real Madrid stadium*. F. Gómez, F.J. Serón. J. Aporta, D. Gutiérrez. Graphite 2005, (sketche) pp. 139-142, 2005. ISBN: 1-59593-201-1. Ed: D. Arnold, M. Billinghurst in cooperation with ACM SIGGRAPH. Ed: ACM SIGGRAPH.

[19] *Tools for Simulation and Visualization of Complex Illumination Environments*. E.J. Sobreviela, F.J. Serón, S. Baldassarri, J.A. Magallón, D. Gutiérrez. Research in Interactive Design, Proceedings of Virtual Concept, in CD, 2005. ISBN: 2-287-28772-8. Ed: Springer-Verlag.

[20] *The need for parallel processing in Industrial Lighting Simulation*. J.A. Magallón, F. Serón, E. Sobreviela, A. Muñoz, D. Gutiérrez. Eurographics Symposium on Parallel Graphics and Visualization, 2007. Submitted, in review.

[21] *A Survey of Inverse Surface Design From Light Transport Behaviour Specification.* Gustavo Patow, Xavier Pueyo. Computer Graphics Forum, vol 24 (2005), number 4, pp. 773-789.

[22] *A Hierarchical Algorithm for Radiosity Daylighting*. Gonzalo Besuievsky, Ignacio Martin. Congreso Español de Informática Gráfica, 2005, Granada.

[23] *Accurate Interactive Specular Reflections on Curved Objects*. P.Estalella, I.Martin, G.Drettakis, D.Tost, O.Devillers, F.Cazals. Proceedings of the Vision, Modeling and Visualization 2005, Erlangen, Germany, November 2005.

[24] *A GPU-driven Algorithm for Accurate Interactive Reflections on Curved Objects*. P.Estalella, I.Martin, G.Drettakis, D.Tost. Eurographics Symposium on Rendering 2006, Cyprus, June 2006.

[25] *A Particle-Path Based Method for Monte-Carlo Density Estimation* M. Lastra, C. Ureña, J. Revelles, R. Montes Poster at: 13th EUROGRAPHICS Workshop on Rendering. Pisa, Italy, June 26-28, 2002.

[26] *A Density Estimation Technique for Radiosity*. M. Lastra, C. Ureña, J. Revelles, R. Montes. 1st Ibero-American Symposium in Computer Graphics (SIACG'2002). Guimaraes, Portugal, July 2-5, 2002.

[27] *Interactive Global Illumination for Quasi-Static Scenes*. R.J. García, C. Ureña, M. Lastra, R. Montes, J. Revelles. Proceeedings of the Computer Graphics International Conference 2004. pp.128-131. Heraklion. Crete (Greece). June 16-19, 2004.

[28] *Optimizaciones en Estimación de Densidades para Iluminación Global*. R. J. García, C. Ureña, M. Lastra, R. Montes y J. Revelles Actas del XV Congreso Español de Informática Gráfica, CEIG2005, pp.33-42 ISBN: 84-9732-431-5.

[29] *Density Estimation Optimizations for Global Illumination*. R. Garcia, C. Ureña, J. Revelles, M. Lastra, R. Montes WSCG'2006 - Short Communications Proceedings. ISBN: 80-86943-05-4.

[30] *A study of incremental update of global illumination algorithms* R. J. García, C. Ureña, M. Lastra, R. Montes y J. Revelles WSCG'2007 - Short Communications Proceedings.

[31] *Parallelization of Implicit-Explicit Runge-Kutta Methods for cluster of PCs* J.M. Mantas, P. González, J.A. Carrillo. Euro-Par 2005. Lisboa, Portugal Aug 30 - Sept 2, 2005. Published in Lecture Notes in Computer Science 3648, 815-825.

[32] *Ray Maps for Global Illumination* V. Havran, J. Bittner, R. Herzog, H.P. Seidel. In Eurographics Symposium on Rendering, pages 43-54. 2005.

[33] *Adaptive Numerical Cumulative Distribution Functions for Efficient Importance Sampling* J. Lawrence, S. Rusinkiewicz, R. Ramamoorthi Eurographics Symposium on Rendering, June 2005.

Jornada de Seguimiento de Proyectos en Tecnologías Informáticas, 2003
Programa Nacional de Tecnologías Informáticas

AE3: Aprendizaje, Evolución y Estadística Extrema TIN2004-07676-C02

José R. Dorronsoro *
Instituto de Ingeniería del Conocimiento, UAM

Alberto Suárez †
Departamento de Ingeniería Informática, UAM

Abstract

A common feature in many hard pattern recognition problems is the fact that the object of interest is statistically overwhelmed by others. The overall aim of the "Learning, Evolution and Extreme Statistics" (AE3 being its Spanish acronym) project is to study those problems in the following concrete areas:

1. Natural image statistics and applications.
2. New classification techniques in extreme sample problems.
3. Evolutionary machine learning.
4. Machine learning and evolutionary computing in finance.

AE3 is a coordinated project between a research group at the Instituto de Ingeniería del Conocimiento (IIC) and another at the Escuela Politécnica Superior (EPS), both in the Universidad Autónoma de Madrid (UAM).

Keywords: Machine Learning. extreme statistics, natural images, margin classifiers, neural networks, information.

1 Project objectives

A frequent trait in many difficult prediction and classification problems is that the behaviour of a given item may be hidden under that of many others, possibly less relevant but statistically dominant. For instance, fraudulent card behaviour takes place within an overwhelming majority of legal traffic. On the other hand, among the many possible image blocks of a given size, the number of natural blocks with high information content (for instance, those having

*Email: jose.dorronsoro@uam.es
†Email: alberto.suarez@ii.uam.es

edges) is much smaller. Or relevant statistical structure in financial time series may be obscured by other stochastic factors. The consequence is that learning the interesting model or class becomes very difficult.

The overall goal of the AE3 project is to study such "extreme sample" modelling and classification problems in four application areas:

1. Natural image statistics and applications.

2. New classification techniques in extreme sample problems.

3. Evolutionary machine learning.

4. Machine learning and evolutionary computing in finance.

AE3 is a coordinated project between a research group at the Instituto de Ingeniería del Conocimiento (IIC) and another at the Escuela Politécnica Superior (EPS), both in the UAM. The IIC group will lead the work done in the first two areas, natural image statistics, where high information content structures such as edges lie in low dimensional submanifolds, and pattern recognition problems where the class of interest (the positives, such as credit card fraudulent operations or patients with a concrete disease) is hidden within the much more common negative cases. The EPS team will lead work done in the other two areas, with the overall aim being to develop novel machine learning algorithms, with emphasis on the use of evolutionary computation techniques and applications to finance. Most machine learning problems can be formulated as an optimization problem. By making use of genetic algorithms and/or evolution strategies, one can address machine learning problems that require the minimization of cost functions that are discrete or discontinuous. The algorithms developed are being applied to financial problems involving time series modeling, portfolio optimization, derivatives pricing and risk analysis.

Next we briefly review concrete objectives in the above areas.

1.1 Natural image statistics and applications

Natural image statistics has a clear importance in questions such has compression, filtering, transmission, enhancing or, even, the understanding of the visual system. Nevertheless, its study has been mostly restricted to 2×2 blocks or, partially, to 3×3 blocks. A direct statistical study is extremely difficult for 4×4 blocks: the best compression rates available require about 2.5 bits per pixel for 8–bit grey scale images, which means that a reasonable sample would require about 2^{40} such blocks, or about 16×10^6 1024×1024 images. This means that statistics collections has to be done on compressed images; however, the more widely used methods such as wavelets may result in different structural representations for different images, which makes difficult uniform statistics collection. To avoid this, a method that combines fractal compression with hash key computations will be used on 4×4 natural image blocks. Once a good understanding of the statistical structure is obtained, possible applications to predictive compression or searching on image databases are to be considered.

1.2 New classification techniques in extreme sample problems

There are many ways in which a given sample could be considered "extreme": large pattern dimension, large sample sizes, imbalances between class sample sizes or between sample size and pattern dimension or high overlapping between sample classes. Some of these characteristics relate to the statistical properties of the problem under consideration; others result in severe complexity restrictions to the algorithms that may be applied. These circumstances will be considered in the project. For problems with large sample overlapping, we will first consider "target free" methods, that is, methods that do not construct classification models trying to approximate predefined targets. Another option are methods that try to identify overlapping patterns, for instance, viewing them has having "noisy" class labels, and that construct classifiers over a "noisy pattern" free sample. A possibility for this is to consider methods that try to optimize a certain margin: once computed, it may be used to identify noisy label patterns (for instance, those with a large wrong margins). A first example are the well known support vector machine methods; another option, using a different margin concept, are the so–called parallel perceptrons. An advantage of these methods is their simple structure and, therefore, their relatively light complexity requirements. They are promising thus not only for class overlapping problems but also for those showing large sample size o high pattern dimensions. A third possibility, classifier ensembles, is discussed next.

1.3 Learning with ensembles

Learning devices that pool the decisions of collections of learners of the same type (bagging, boosting, random forests, etc.) can perform better than a single learner, also of the same type, in many regression and classification tasks. These methods take advantage of the instabilities in the algorithms that generate the base classifiers or regressors to construct an ensemble of diverse learners whose joint action may improve the classification performance of a single classifier. In this project, several new methods to generate ensembles are developed. We have also investigated the influence of the order in which new units are incorporated into the ensemble. If individual learners are aggregated taking into account their complementary performance, it is possible to identify subensembles which, besides being smaller than the original complete ensembles, may also have better generalization performance.

1.4 Evolutionary machine learning

Traditional methods of supervised dimensionality reduction proceed by maximizing one or another measure of class separation. Quite often, the rank deficiency of the involved covariance matrices precludes the application of this classical approach to real situations. Besides, the number of projections can not be chosen freely but it is bounded to be equal to the number of classes minus one. During the last few years, one of our main objectives has been the design of a dimensionality reduction method capable of obtaining two-dimensional renderings of data sets independently of the number of classes of the problem. We have found in evolutionary computation a means of skipping the use of covariance matrices.

As a second objective, we have tried to answer whether general minimization problems, not only classification problems, can benefit from the use of one dimensional projections. After all, classical minimization algorithms proceed by successive one dimensional line minimization

algorithm. We have been investigating whether these successive directions can be found by evolutionary means and, in this way, improve the classical minimization approaches. Our results are very promising as we report in next section.

1.5 Machine learning and evolutionary computing in finance

One of the goals of this project is the application of Artificial Intelligence techniques to the solution of problems of practical interest in the area of quantitative finance. The increasing complexity and global character of financial markets has led to a surge in the demand for the use of sophisticated quantitative and computational tools. There are many tasks in this area, such as the pricing of derivative products with a complex structure, quantification and management of financial risks and time series modelling, that would greatly benefit from the application of recent advances in the field of Artificial intelligence: Intelligent data analysis, global optimization methods, inductive learning, and so on.

The work on financial applications developed in the AE3 project focuses on four different areas chosen because of their relevance to financial companies, in particular to some of the EPO's involved in this project, and because they are within the area of expertise of our research group.

- Use of evolutionary and other heuristic tools for the solution of financial optimization problems with realistic market constraints (single-period portfolio optimization, index tracking, etc.)
- Modeling and analysis of financial time series from the point of view of inductive learning. In particular, we have designed and implemented models within the "mixture of experts" paradigm for the analysis of financial time series, with applications to the quantification of market risk.
- Advanced statistical tools for the modelling of operational risk loss data. This type of analysis is of practical interest because financial institutions are required to quantify their risk exposure originating in operational losses. Companies whose main area of activity lies outside of finance also need to measure risk to become more efficient by improving their risk management practices.
- Design of Direct Reinforcement learning methods to solve multi-period portfolio optimization with different risk-adjusted performance measures and realistic market constraints.

We turn next to the description of the results obtained up to this point, both with respect to the above described objectives and with respect of other research or application areas.

2 Goals accomplished

The goals accomplished at this moment of the project (ending of its second year) are described next. This section is structured in 8 subsections. The first five correspond to the original objectives, while section 6 others describe scientific results in areas not considered originally but that represent outgrowth of other work done in AE3, section 7 gives technology transfer actions and section 8 describes collaboration with other research groups. This section ends with a brief summary of technology transfer actions that have taken place within AE3.

2.1 Natural image statistics and applications

We have first obtained an approximate statistical model [6] of the distribution of 4×4 pixel natural image patches B. Given the huge size of the sample space, we have collected the required statistics over a fractal compression inspired representation of B, namely by a triplet (D^B, μ_B, σ_B) with σ_B the patch's contrast, μ_B its brightness and D^B a codebook representation of the mean-variance normalization of B : $(B - \mu_B)/\sigma_B$. While not coinciding exactly with the true natural patch density $p(B)$, the density $\hat{p}(B) = p(D^B, \mu_B, \sigma_B)$ should give an adequate approximation of $p(B)$, because $B \simeq \sigma_B D^B + \mu_B$. We have shown that the probability density $p(D, \mu, \sigma)$ can be factored as $p(D, \mu, \sigma) \simeq p(D)p(\mu)p(\sigma)\Phi(||\nabla B||)$, with Φ being a high contrast correction. Furthermore, we have also studied the structure of the $p(\sigma)$ and $p(D)$ two factors (the brightness term $p(\mu)$ is largely irrelevant), showing that $p(\sigma)$ follows an exponential distribution and that $p(D)$ is uniformly distributed with respect to volume in image space. Finally we have shown that the above results are largely independent of the codebook used.

The second activity area has been new methods in image compression [3, 20] and representation, more precisely on extensions for that purpose of the Hopfield networks. The main advantage of these networks is their noise-resistance. but an important drawback is their low capacity and their instability due to the recurrent character of the network. Another drawback, concerning images is its predominantly long-range connectivity. To overcome these difficulties, a sparse representation and spatial-dependent connectivity of the network were chosen. Additionally, an asymmetry between the retrieval and learning activities is imposed.

The above setting leads to the formation of well-defined spatial patterns – bumps that extract some part of the image. They are noise resistant, image representative and able to reconstruct the image from tiny initial overlap [33]. Thus the bumps can be seen as very noise-resistant signatures of the images they represent. The behaviour of these networks has been studied analytically and by simulations. These networks posses some minimal properties to develop bump activity [3] and the capacity of the networks increase to more than one pattern per synaptic link [2].

2.2 New classification techniques in extreme sample problems

This area seeks to solve classification problems that may combine a high imbalance between class sample sizes, a possibly large overlap among classes and very large sample sizes. There have been three main research lines in this area. The first one concentrates on the application of natural gradient techniques for the training of nonlinear discriminant analysis networks. It has been shown by Amari and his co–workers that the Fisher information matrix defines for maximum log–likelihood estimation a Riemannian metric in weight space and that the resulting natural gradient greatly accelerates on–line multilayer perceptron (MLP) training. We have applied the same ideas to Non Linear Discriminant Analysis (NLDA) networks, variants of standard multilayer perceptrons where Fisher's linear discriminant analysis criterion is used instead of the usual square error function (that, as it is known, tends to place a bigger weight on the larger class). We have introduced natural gradients for NLDA networks observing that the gradient $\nabla J(W)$ of the NLDA criterion function can be written as the expectation $E[Z(X, W)]$ of a certain random vector Z; defining then the Fisher information matrix as $\mathcal{I} = E[Z(X, W)Z(X, W)^t]$, it formally coincides with the square error MLP information matrix.

Although very simple, the proposed approach shows [10] much faster convergence than that of standard NLDA gradient descent, even when its costlier complexity is taken into account.

Returning to MLPs, batch natural gradient descent counterpart also improves on standard gradient descent (as it gives a Gauss–Newton approximation to mean square error minimization), but it is no longer be competitive with more advanced gradient–based function minimization procedures. However, and as a consequence of the previous work, it has been observed [29, 19] that conjugate gradient descent can be also defined on a Riemannian setting, showing encouraging results for multilayer perceptron training, leading to faster convergence to better minima than that achieved by standard euclidean CG descent, even when cost simplifying variants are applied (a drawback of full natural gradient is its larger computational cost). We are presently further pursuing these properties.

In any case, the complexity of methods using an MLP architecture makes them unsuitable for problems with either very large sample sizes or very high pattern dimensionality. This has lead the IIC research group to study simpler methods based on the original Rosenblatt's perceptron (pcp) architecture, where two alternative lines are being followed. The first one works with the so–called parallel perceptrons, that consist in an odd number of standard pcps that are trained independently in parallel fashion and whose output is computed by majority voting. Besides weight adaptation, their training also adjusts a margin value for each R-pcp. Because of this, we have studied their applicability to imbalanced classification problems, where we try to remove possibly label–noisy patterns, either directly [15] or through a boosting procedure [13, 16]. Moreover, while the output weights of standard PPs are fixed to 1 (i.e., they simply add the outputs of their individual pcps), in [18] it is shown how to improve their performance allowing their output weights to have real values that are computed by applying Fisher's linear discriminant analysis to them. This improves the final performance of the resulting classifiers over that of standard PPs, making it comparable to that of the more complex and costlier to train multilayer perceptrons.

The second approach, that has just been started, works with pcps in a support vector machine (SVM) setting. The good classification properties of SVMs are well known; on the other hand, their training may be quite costly, particularly for large databases. Moreover, a large number of final support vectors makes their application quite slow. Simpler training procedures, closer to Rosenblatt's rule for pcps have been recently introduced that obtain optimal margins while having smaller training costs. The basic idea of these procedures is to recast SVM training as a minimum norm problem on the sample's convex hull. In [23] we have shown that Rosenblatt's rule can be rewritten so that the successive weights lie in that convex hull. Moreover, in [41] (and partially in [35]) it is shown how to relate the coefficients in the weight's convex combination representation with the margins of the corresponding support vector, which, in turn, allows an efficient selection of the final support vectors to be used and, therefore, a faster application of the final SVM.

2.3 Ensemble learning

Several aspects of classification and regression involving ensembles of predictors have been investigated:

2.3.1 New algorithms for the generation of classifier ensembles

- Class switching ensembles are built by generating each classifier with a different perturbed version of the training set. These surrogate training sets are obtained by randomly switching the class labels of some of the examples in the original set. The base classifiers can be either decision trees [1] or neural networks [22] whose architecture is determined by standard selection techniques. The individual decision trees in these ensembles should be trained to achieve 0-error in the perturbed training set. Provided that relatively high class switching rates and large ensemble sizes are used, the error accuracies achieved are better than bagging and comparable or better than boosting in the benchmark problems investigated.

- In [37] a new ensemble learning algorithm that combines features of boosting and bagging is introduced. The method uses boosting ensembles as the base classifier of a bagging-type ensemble to finally obtain an ensemble of ensembles (or committees). This method preserves the robustness of bagging in noisy datasets and the good generalization performance of boosting.

2.3.2 Ensemble pruning

The generalization error of classification and regression ensembles can sometimes be reduced by eliminating some of their members. However, it can be demonstrated that the problem of selecting the optimal subensemble from the original pool of learners is an NP-complete problem and therefore intractable for ensembles that are typically composed of 100 or more elements. In the framework of this project several algorithms with polynomial time complexity have been developed to approximately solve the problem of selecting a subensemble that is near-optimal. These algorithms are based in reordering the original classifiers of the ensemble and selecting the first τ elements for aggregation according to a pruning rule. The resulting subensembles besides being smaller (which implies faster classification time and reduced storage requirements) can improve the classification performance of the original ensemble. Different heuristics have been designed to order the classifiers generated on different bootstrap samples of the training set [8, 30]. Experiments carried out in several benchmark classification problems show that that proposed methods are able to reduce the generalization error of the ensembles using between 15% and 30% of the elements in the original ensemble. The approach used in [30] for classification has also been successfully applied to regression problems [31].

As a result of the success of the pruning methods developed for bagging ensembles, additional research was made to assess whether these techniques can be extended to other ensemble learning algorithms, such as boosting. In [24] several pruning techniques for classification ensembles and a genetic algorithm devised to extract the optimal subensemble from a complete initial AdaBoost ensemble are compared.

2.4 Evolutionary machine learning

Evolutionary Discriminant Analysis (EDA) [7] is capable of reducing the number of dimensions of a classification problem by the direct minimization of the number of misclassified patterns. No matrices are involved whatsoever and the number of projections can be chosen without restrictions. This allows to obtain two dimensional renderings of problems with any number

of classes, such as the 19 class UCI soybean problem. We have also succeeded in training multilayer perceptrons without the use of output codes [27, 11].

We have also developed a new estimation of distribution algorithm (ED2) [36] which takes advantage of one dimensional subproblems such as EDA does. In Estimation of Distribution Algorithms (EDAs), minimization proceeds by sampling new candidate solutions from a probability distribution function (pdf) which is improved along the process. One of the simplest pdfs for continuous domains is a product of one dimensional Gaussian functions. Not surprisingly, this pdf fails to yield global minima in problems with correlated variables. Unlike most EDAs, which try to circumvent this limitation by adapting the structural form of the pdf to the problem at hand, our algorithm adapts the problem itself to the sampling mechanism. More precisely, our algorithm samples lines of the target function from a product of one dimensional Gaussian functions. Basically, this amounts to evolving correlations between the variables instead of estimating them out of the best samples as it is usually the case. In order to assign fitness to each line, the algorithm resorts to a classical line minimization algorithm. Brent's line minimization method is shown to yield state of the art results for functions such as the Rosenbrock function.

2.5 Machine learning and evolutionary computing in finance

2.5.1 Evolutionary approaches to optimization problems in finance

- In [32] the problem of selecting an optimal portfolio for investment is addressed using a hybrid optimization strategy. This strategy separates the combinatorial optimization aspect of the problem from the quadratic part. A genetic algorithm with specially designed crossover and mutation operators is used to select which assets should be included in the portfolio. Quadratic programming is then used to solve the problem of finding the optimal portfolio, within the Markowitz model, given the assets selected for investment.

- The optimization strategy designed in [32] can also be applied to the related problem of tracking a financial index using a portfolio composed of a limited number assets, selected among the ones that make up the index. The objective is to minimize the tracking error, which can be defined as the quadratic deviation between the index and the tracking portfolio returns. This hybrid strategy is an effective and efficient approach to the index tracking problem [38].

2.5.2 Time series models in finance

In [21] we compare the performance of competitive and collaborative strategies for mixtures of autoregressive experts with normal innovations for conditional risk analysis in financial time series. Experiments indicate that the best performance for risk analysis is obtained by mixtures with soft competition, where the experts have a probability of activation given by the output of a gating network of softmax units.

2.5.3 Advanced statistical tools for the modelling of operational risk loss data

Operational risk has been defined as the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. The regulatory framework outlined

by the *Basel Committee on Banking supervision* establishes three methods for calculating operational risk capital charges, of which the Advanced Measurement Approaches (AMA) has the highest level of sophistication and risk sensitivity. The Lost Distribution Approach (LDA, which is widely used in actuarial analysis, is one of the implementations with stronger formal support. However its application to operational loss data, which is typically strongly heavy-tailed is plagued by difficulties. Using synthetic data, we analyze in [12] the risk associated with the selection incorrect models from limited empirical evidence.

2.5.4 Portfolio Selection by Direct Reinforcement Learning

In this project, a collaboration with professor John Moody at the ICSI (Berkeley, CA, USA), we address different aspects of the problem of portfolio management using extensions of Direct Reinforcement learning. Portfolio selection with realistic market constraints and transaction costs is a complex stochastic control problem, which combines combinatorial and continuous optimization aspects. The problem is a testbed to investigate several extensions of Direct Reinforcement learning of practical interest. The results of this work will be presented to the EPO (Supervising Industrial Partner) *Analistas Financieros Internacionales, AFI* upon completion (est. june 2007).

2.6 Other research areas

Besides AE3 main research lines, work has been also carried out in some other areas, that are briefly described next.

2.6.1 Text mining

Compression and hashing are frequently related in that compression codes give short data representations that, in turn, can be used for fast searches. Moreover, compression procedures also reflect data's underlying statistical structure. Under that light, work has started to exploit compression information for text analysis and in the contributions [5, 34] a procedure has been developed to measure and identify text similarity through compression information that allows to establish a common source for text and binary data that works, for instance, over different languages and alphabet coding.

2.6.2 Bioinformatics

Bioinformatics is clearly one of the most promising areas of data mining. The IIC group has just started a collaboration with the Systems Biology Research Group at the University of Ulster with a six month stage of professor Ana González in that group. Pattern recognition problems involving genomic data typically involve relatively few sample patterns that, on the other hand, may have very large dimensions, making feature selection mandatory. In [39] a study has been carried out on the most frequently used statistical techniques for genomic feature selection and a new procedure has been proposed and tested on several genomic data bases.

2.6.3 Graph connectivity properties

The properties of processes working on graphs are greatly influenced by their connectivity characteristics. This has been already mentioned for recurrent graphs used in image compression and in the references [4, 9, 14, 17, 28] it is studied how the overall connectivity patterns and the individual node connectivity density may define different attractor regions and produce specific spike bursting patterns.

2.6.4 Subadditive penalties in decision tree pruning

Recent work on decision tree pruning has brought to the attention of the machine learning community the fact that, in classification problems, the use of subadditive penalties in cost-complexity pruning has a stronger theoretical basis than the usual additive penalty terms. In [25] cost-complexity pruning algorithms with general size-dependent penalties are implemented to confirm that the family of pruned subtrees selected by pruning with a subadditive penalty of increasing strength is a subset of the family selected using additive penalties. Consequently, this family of pruned trees is unique, it is nested and it can be computed efficiently. However, in spite of the better theoretical grounding of cost-complexity pruning with subadditive penalties, we found no systematic improvements in the generalization performance of the final classification tree selected by cross-validation using subadditive penalties instead of the commonly used additive ones.

2.7 Technology transfer

The technology transfer activities derived from work done by AE3 researches encompasses the following three areas:

- Wind energy prediction.
- Credit card fraud detection.
- Operational risk (in collaboration with Risklab Madrid).

2.7.1 Wind energy prediction

The IIC group has acquired in the past years a large experience in wind energy prediction, both for individual wind farms as well as for joint prediction for farm groups. This has resulted in a demonstration service currently under operation that captures numerical weather predictions and provides to about 25 farms with two daily predictions in a time window ranging from 12 to 96 hours after the 0 hour of the weather forecasts. Besides this, the IIC is presently working on global predictions for Spain's wind energy production, working with data representing about 80 % of Spain's installed power. In 2005 IIC asked for the registration in Spain's Registro de la Propiedad Intelectual of a first version of this software with the reference number M-000051/2005. Among its authors appears two AE3 researchers, J. Dorronsoro and A. Suárez. A new inscription covering the system present functionality will be made in the first semester of 2007.

2.7.2 Credit card fraud detection

IIC also has a long experience in credit card fraud detection, having developed a solution named Lynx currently being used by several banks and credit card service providers. Credit card fraud is a clear example of a problem with very large sample sizes and where class sample imbalance and class overlaps are also very high. Lynx successive versions have been inscribed in Spain's Registro de la Propiedad Intelectual and a new inscription is going to be asked in the first semester of 2007. Some of the techniques being currently studied in the AE3 problem are being considered for potential application in Lynx, although not in the immediate future. The AE3 researches which will be included are Carlos Santa Cruz and Pedro Pascual.

2.7.3 Operational risk – in collaboration with Risklab Madrid

RiskLab-Madrid is a research institute in quantitative and computational finance. This institute specializes in the measurement, analysis and modelling of financial risks. In collaboration with some of the researchers in this group RsikLab has carried out extensive research on the modelling of heavy tails, skewness and correlations in financial data and on the impact of these factors in the risk profile of an institution. As a result of this collaboration, some of the results of this project have been presented in the "VI^{a} Jornada de riesgos financieros", jointly organized by RiskLab Madrid and BBVA in Madrid on OCtober 19, 2006. [http://www.risklab-madrid.uam.es/es/jornadas/2006/index.html] This work [12] has also been published in *Estabilidad Financiera*, a publication of the *Bank of Spain* whose goal is to disseminate information and participate in debates on issues relevant to the stability of the financial system. The focus of this journal is on regulatory and supervision practices.

2.8 Collaborations with international groups

2.8.1 Researcher exchanges

- Ana González has spent six months from July 2006 to January 2007 at the Systems Biology Research Group at the University of Ulster, directed by professor W. Dubitzky, working on the statistical analysis of feature selection for genomic data. Professor González intends to make a second stage with the same group in 2007.

- Alejandro Sierra spent four months from July 2006 to October 2006 at the Bionik und Evolutionstechnik Laboratory of the Technical University of Berlin, Germany. The project entitled "Multilevel Evolutionary Computation" consisted in the development of a new Estimation of Distribution Algorithm based on the sampling of directions for minimization problems. This stay was sponsored by a fellowship from the Spanish Ministry of Education.

- Alberto Suárez has been the 2005 Visitor Senior Scientist Program at the International Computer Science Institute (ICSI), affiliated to the University of California, Berkeley. The project developed involves a collaboration with professor John Moody on applications of Artificial Intelligence in finance. It involves two stays at the ICSI, from June 2006 to August 2006 and from March 2007 to May 2007.

2.8.2 Student exchanges

- Daniel Hernández-Lobato spent 3 months (september 2006-dec. 2006) in Prof. Bert Kappen's group at the Biophysics department of the Radboud University in Nijmegen, The Netherlands to receive formation in Bayesian methods in machine learning.
- José Miguel Hernández-Lobato spent 3 months (september 2006-dec. 2006) in Prof. Tom Heskes' group at the Department of Information and Knowledge Systems (IRIS) at RadBoud University in Nijmegen, The Netherlands. During his stay, José Miguel Hernández-Lobato developed a Bayesian model to describe relationships in biological time series for gene expression.

3 Main results indicators

This section will contain first numerical indicators of the results obtained so far.

3.1 Personnel formation

- Ph. D. Theses:
  - Gonzalo Martínez, Clasificación mediante conjuntos, February 2006.
- Completed 2nd. year doctorate works read at the EPS–UAM PhD Program:
  - Iván Cantador, Aplicación de Perceptrones Paralelos y AdaBoost a Problemas de Clasificación de Muestra Extrema. June 2005.
- Ph.D. students and status:
  - Alejandro Echeverría (Ph.D. thesis to be read during the spring of 2008).
  - Daniel Hernández Lobato (2nd year Ph.D. Student, recipient of an *FPI fellowship (2005)* from the *Comunidad Autónoma de Madrid*)
  - José Miguel Hernández Lobato (2nd year Ph.D. Student, recipient of a *FPU fellowship (2005)* from the *Universidad Autónoma de Madrid*)
  - Daniel Hernández Lobato (2nd year Ph.D. Student) (2nd year Ph.D. Student, recipient of an *FPU fellowship (2006)* from the *Universidad Autónoma de Madrid*)
  - Daniel García (2nd. year student, recipient of an *Tercer Ciclo fellowship (2007)* from the *Universidad Autónoma de Madrid*)
  - Alvaro Barbero (1st. year student), recipient of an IIC Doctoral Fellowship (2007) from the *Instituto de Ingeniería del Conocimiento, Universidad Autónoma de Madrid.*
  - Jorge López (1st. year student), recipient of an *Tercer Ciclo fellowship (2007)* from the *Universidad Autónoma de Madrid*

  It deserves mention that three of these students have received the "Mención Honorífica" of the EPS–UAM after graduation, as having the best grades of their classes.

3.2 Publications

- JCR–covered Journals: 10 (7 IIC group + 3 EPS group)
- Other international journals: 1 (0 IIC group + 1 EPS group)
- National journals: 1 (0 IIC group + 1 EPS group)
- Lecture Notes in Computer Science: 13 (9 IIC group + 4 EPS group)
- International Conference proceedings: 11 (5 IIC group + 6 EPS group)
- National Conference proceedings: 1 (0 IIC group + 1 EPS group)
- Other (submitted work, preprints, reports): 4 (3 IIC group + 1 EPS group)

3.3 Technology transfer

- "Motor EA2 (Energía Eólica en Áreas Amplias)": inscription with reference number M-000051/2005 in Spain's Registro de la Propiedad Intelectual in February 2005; a new inscription to be made in the first semester of 2007.
- "Lynx" credit card fraud detection system: a new inscription to be made in the first semester of 2007.

References

JCR–covered Journals

[1] G. Martínez-Muñoz and A. Suárez. Switching class labels to generate classification ensembles. *Pattern Recognition*, vol. 38(10), pp. 1483–1494, 2005.

[2] K. Koroutchev, E. Korutcheva, Conditions for the emergence of spatially asymmetric retrieval states in attractor neural networks, *Central European Journal of Physics* 3 (2005) 409–419.

[3] K. Koroutchev, E.Korutcheva, Bump formation in a binary attractor network, *Physical Review E* 73 (2006) 026107.

[4] C. Aguirre, D. Campos, P. Pascual, E. Serrano, Synchronization effects using a piecewise linear map-based spiking-bursting neuron model, *Neurocomputing* 69, 1116–1119, 2006

[5] K. Koroutchev and M. Cebrian, Detecting Translations of the Same Text and Data with Common Source, *J. Stat. Mech.*, (2006),P10009

[6] K. Koroutchev, J.R. Dorronsoro, Factorization and structure of natural 4 x 4 patch densities *Journal of Optical Engineering* 45 (2006), 12-24.

[7] A. Sierra and A. Echeverría, Evolutionary discriminant analysis. *IEEE Transactions on Evolutionary Computation*, 10 (1), 2006, pp. 81–92.

[8] G. Martínez-Muñoz and A. Suárez. Using boosting to prune bagging ensembles. *Pattern Recognition Letters*, vol. 28(1), pp. 156–165, 2007.

[9] D.Dominguez, K.Koroutchev, E.Serrano, F.B. Rodríguez, Information and Topology in Attractor Neural Networks, *Neural Computation*, in press.

[10] A. González, J.R. Dorronsoro, Natural learning in NLDA networks, accepted in September 2006 for publication in *Neural Networks*.

Other International Journals

[11] A. Sierra and A. Echeverría, Neural networks trained by distance to means. *WSEAS Transactions on Information Science and Applications*, vol. 2 (9), 1446-1453, 2005.

National journals

[12] S. Carrillo Menéndez, A. Suárez. Medición efectiva del riesgo operacional, *Estabilidad Financiera*, vol. 11, páginas 61–90, 2006.

Lecture Notes in Computer Science

[13] I. Cantador, J.R. Dorronsoro, Balanced Boosting with Parallel Perceptrons, *Proceedings of the IWANN 2005, International Work Conference in Neural Networks*, Lecture Notes in Computer Science 3512, 208–216, Springer Verlag 2005 .

[14] C. Aguirre, D. Campos, P. Pascual, E. Serrano A Model of Spiking-bursting Neuronal Behaviour Using a Piecewise Linear Twodimensional Map, *Proceedings of the IWANN 2005, International Work Conference in Neural Networks*, Lecture Notes in Computer Science 3512, 130–135, 2005.

[15] I. Cantador, J.R. Dorronsoro, Parallel Perceptrons, Activation Margins and Imbalanced Training Set Pruning, *Proceedings of the IBPRIA 2005, Iberian Pattern Recognition and Image Analysis Conference*, Lecture Notes in Computer Science 3523, 43–50, Springer Verlag 2005.

[16] I. Cantador, J.R. Dorronsoro, Boosting Parallel Perceptrons for Label Noise Reduction in Classification Problems, *Proceedings of the IWINAC 2005, International Work Conference in Neural Networks*, Lecture Notes in Computer Science 3562, 586–593, Springer Verlag 2005.

[17] C. Aguirre, D. Campos, P. Pascual, E. Serrano Neuronal Behaviour With Sub-threshold Oscillations And Spiking/bursting Activity Using a Piecewise Linear Two-dimensional Map, *Proceedings of the ICANN 2005 International Conference in Artificial Neural Networks*, Lecture Notes in Computer Science 3696, 103–108, 2005.

[18] A. González, I. Cantador, J.R. Dorronsoro, Discriminant Parallel Perceptrons, *Proceedings of the ICANN 2005 International Conference in Artificial Neural Networks*, Lecture Notes in Computer Science 3697, 15–20, 2005.

[19] A. González, J.R. Dorronsoro, Natural Conjugate Gradient Training of Multilayer Perceptrons. *Proceedings of the ICANN 2006 International Conference in Artificial Neural Networks*, Lecture Notes in Computer Science, 4131, 169-177, 2006.

[20] K. Koroutchev, E.Koroutcheva, Improved Storage Capacity of Hebbian Learning Attractors, *Proceedings of the ICANN 2006 International Conference in Artificial Neural Networks*, Lecture Notes in Computer Science, 4131, 234–242, 2006.

[21] J. M. Hernández-Lobato and A. Suárez. Competitive and collaborative mixtures of experts for financial risk analysis. *Lecture Notes in Computer Science, Proceedings of the 2006 International Conference on Artificial Neural Networks (ICANN 2006)*, Athens (Greece), September 10-14, 2006. Part II vol. 4132, pp. 691–700, 2006.

[22] G. Martínez-Muñoz, A. Sánchez-Martínez, D. Hernández-Lobato, and A. Suárez. Building ensembles of neural networks with class-switching. *Lecture Notes in Computer Science, Proceedings of the 2006 International Conference on Artificial Neural Networks (ICANN 2006)*, Athens (Greece), September 10-14, 2006. Part I, vol. 4131, pp. 178–187, 2006.

[23] D. García, A. González, J.R. Dorronsoro Convex Perceptrons. *Proceedings of the 7th International Conference on Intelligent Data Engineering and Automated Learning - IDEAL 2006*, Lecture Notes in Computer Science 4224, 578–585.

[24] D. Hernández-Lobato, J. M. Hernández-Lobato, R. Ruiz-Torrubiano, and A. Valle. Pruning adaptive boosting ensembles by means of a genetic algorithm. *Lecture Notes in Computer Science Proceedings of the 7th Intelligent Data Engineering and Automated Learning (IDEAL 2006)*, Burgos (Spain), September 2006., pp. 322–329, 2006.

[25] S. García-Moratilla, G. Martínez-Muñoz, and A. Suárez. Evaluation of decision tree pruning with subadditive penalties. *Lecture Notes in Computer Science, Proceedings of the 7th Intelligent Data Engineering and Automated Learning (IDEAL 2006)*, Burgos (Spain), September 2006. vol. 4224, pp. 995–1002, 2006.

International congress communications

[26] K. Korutchev, C. Miravet, F. Rodríguez, Super-resolution using codebooks, *Proceedings of the Intl. Conf. On Multimedia, Image Processing and Computer Vision, IADAT-micv2005*, 128–131, Madrid, 2005.

[27] A. Sierra and A. Echeverría, Perceptrons without output codes. *Proceedings of the 5th WSEAS International Conference on Simulation, Modelling and Optimization*, Corfu, Greece, pp. 331-336, 2005.

[28] K. Koroutchev, K. Korutcheva, Spatial asymmetric retrieval status in attractor neural networks, *Proceedings of the Fluctuation and Noise* Conference, 602–608, American Institute of Physics, 2005.

[29] A. González, J.R. Dorronsoro, A Note on Conjugate Natural Gradient Training of Multilayer Perceptrons, *Proceedings of the 2006 International Joint Conference in Neural Networks*, Vancouver 2006.

[30] G. Martínez-Muñoz and A. Suárez. Pruning in ordered bagging ensembles. *Proceedings of the 23rd International Conference on Machine Learning (ICML 2006)*, Carnegie Mellon University, Pittsburgh, Pennsylvania (EE.UU.), June 25-29, 2006, ACM press, New York, (ISBN: 1-59593-383-2), pp. 609–616, 2006.

[31] D. Hernández-Lobato, G. Martínez-Muñoz, and A. Suárez. Pruning in ordered regression bagging ensembles. *Proceedings of the 2006 International Joint Conference on Neural Networks (IJCNN 2006)*, *IEEE World Congress on Computational Intelligence (WCCI, 2006)* Vancouver, BC (Canada), July 16-21, 2006, IEEE ed. (ISBN: 0-7803-9489-5), pp. 1266 – 1273, 2006.

[32] R. Moral-Escudero, R. Ruiz-Torrubiano and A. Suárez. Selection of Optimal Investment Portfolios with Cardinality Constraints. *Proceedings of the 2006 Congress on Evolutionary Computation (CEC 2006)*, *IEEE World Congress on Computational Intelligence (WCCI, 2006)*, Vancouver, BC (Canada), July 16-21, 2006, IEEE ed. (ISBN: 0-7803-9489-5), pp. 2382-2388, 2006.

[33] K, Koroutchev, E.Koroutcheva, Bump formations in attractor neural network and their application to image reconstruction, *Proceedings of 9th Granada Seminar of Statistical Physics*, Granada, Spain, Sept. 2006.

[34] K. Koroutchev, M. Cebrian, Detecting the same text in different languages, *IEEE ITW 2006*, October 2006 Chengdu (China), 1-4244-0067-8,pp. 337-340

[35] Daniel García, Ana González, José R. Dorronsoro, Optimal support vector selection for kernel perceptrons, *Proceedings of Learning 2006*, Vilanova i la Geltrú, Spain, October 2006.

[36] A. Sierra and I. Santibáñez, Evolution of Descent Directions. *Proceedings of 7th EU/MEeting on Adaptive, Self-Adaptive, and Multi-Level Metaheuristics*, University of Málaga, Spain, 2006.

National congress communications

[37] G. Martínez-Muñoz and A. Suárez. Comités de árboles IGP. *Actas del I simposio de inteligencia computacional CEDI-SICO-2005, 13 al 16 de septiembre 2005, Granada*, pp. 277–283. Thomson Press, 2005.

Other: submitted work, preprints, reports

[38] R. Ruiz-Torrubiano and A. Suárez A Hybrid Optimization Approach to Index Tracking. Submitted to *Annals of Operations Research, Special Issue on Applied Mathematical Programming and Modeling, APMOD 2006.*

[39] A. González, D. Berrar, F. Azuaje, W. Dubitzky, J. Dorronsoro, Are corrections for multiple testing adequate for high–dimensional genomic data sets? Submitted to *Bioinformatics*, January 2007.

[40] L. Vázquez, P. Pascual, S. Jiménez, Charge Density Waves, *Encyclopedia of Nonlinear Science*, 131–133, Routledge Taylor & Francis Group, New York, 2005

[41] Support vector selection for kernel perceptrons, Daniel García, Ana González, José R. Dorronsoro, December 2006, (preprint).

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

High-Performance Computing IV
TIN2004-07739-C02-01

Mateo Valero Cortés[*]
Dep. de Arquitectura de Computadores
Universidad Politécnica de Catalunya (UPC)
Barcelona, Spain

Victor Viñals Yúfera[**]
Dep. de Informática e Ingeniería de Sistemas
Universidad de Zaragoza
Zaragoza, Spain

Abstract

This project targets at improving the efficiency of high-performance computing systems. To that end, it is necessary to explore different approaches with a certain level of cooperation and integration: microarchitecture and multiprocessor architecture, compilers and programming models, operating and execution environments, performance analysis, visualization and prediction tools, algorithms and applications. This cooperation and integration will contribute to make simple and elegant design decisions. When measuring efficiency, and besides the classical approach that just includes the execution time, the project uses metrics that consider design aspects (such as cycle time, area and power dissipation of the processor and memory hierarchy, scalability of the microarchitecture and multiprocessor organization), portability and easy-to-use of the programming models and performance when running on multi-user, multi-programmed and distributed environments, among others. This is an ongoing project (Computación de Altas Prestaciones I, II, III and IV) initiated more than 10 years ago with the support of three previous CICYT projects.

Keywords: processor microarchitecture, memory hierarchy, special-purpose architectures, parallel programming models, execution environments for parallel, J2EE and embedded systems, resource management, application optimization, tools for performance analysis.

1 Project objectives

The main objective of this project is to advance the state-of-the-art in High-performance Computing, specifically in the following areas, at a macroscopic level:

- Computer architecture (microarchitecture, memory hierarchy and multiprocessor systems)
- System software for high-performance computing (compilers and programming models, operating environments for a variety of target workloads, tools for performance analysis, optimization and prediction)
- Algorithms and application for high-performance computing (scientific and engineering applications, signal processing, multimedia)

To that end, the project aims at conducting research in these three areas with a certain level of cooperation and integration. This cooperation and integration is contributing to make simple and elegant design decisions.

[*] Email: mateo@ac.upc.edu
[**] Email: victor@unizar.es

The research activities during these two initial years of the project are a continuation of the research that has been carried out in the last 20 years in high-performance computing in the group. The activity during all these years has been supported by and financed with a number of basic research projects from the Spanish Ministry of Science and Technology, as shown in Figure 1. In addition, a number of European projects under the umbrella of the CEPBA (European Center for Parallelism of Barcelona) also supported the potential of the group in carrying the research in these areas.

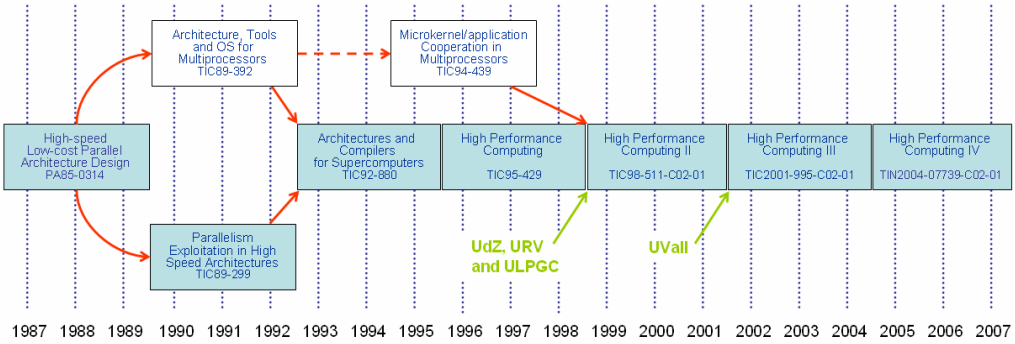


Figure 1: Current and previous CICYT projects related with high-performance computing systems.

The project is organized as a coordinated project with two subprojects:

- **Subproject 1: Architecture, compilers, operating systems, tools and applications.** In this subproject researchers belong to the following universities and groups:
  - Computación de Altas Prestaciones (CAP) group from the Departamento de Arquitectura de Computadores (DAC) at the Universitat Politècnica de Catalunya (UPC).
  - Researchers from the Departamento de Informática y Sistemas (DIS) of the Universidad de las Palmas de Gran Canaria (ULPGC in Figure 1).
  - Researchers from the Universitat Rovira i Virgili (URV in Figure 1).
- **Subproject 2: High-performance memory hierarchy.** In this subproject researchers belong to the following universities and groups:
  - Grupo de Arquitectura (gaZ) at the Universidad de Zaragoza (UdZ in Figure 1).
  - Researchers from the Universidad de Valladolid (UVall in Figure 1).

The project is structured around 3 workpackages, each one with some objectives in mind. We briefly describe them in the remainder of this section.

WP 1: Computer Architecture (subproject 1 and 2)

- **Efficiency and scalability of microprocessor architectures**, including superscalar, VLIW, multithreaded and clustered architectures. Cycle time, area as well as energy consumption are considered metrics to evaluate the overall efficiency. In addition, back-end code generation techniques should also be considered.

- **Reducing latency and increasing bandwidth in the memory hierarchy**, for single- and multi-processor architectures. The research proposal includes both hardware (organization of the physical and virtual registers, centralized vs. distributed organizations, hardware prefetching, …) and software (loop transformations, prefetching, ….) issues for all levels in the memory hierarchy (registers and cache memory in each processor and coherence in multiprocessor systems).
- **Architectures for multimedia applications**, taking into account performance and energy consumption. This includes efficient architectures to exploit data-level parallelism (for example SIMD and vector architectures), and special-purpose architectures for 3D graphics.
- Development of models and tools oriented towards **formal verification in VLSI architectures**. This includes complex synchronous and asynchronous systems.

WP2: System software for high-performance computing (subproject 1)

- **Scalability of programming models**, including shared-memory (such as OpenMP for SMP and ccNUMA architectures), message-passing (such as MPI for highly scalable systems) and GRID. The project also proposes shared-memory in clusters using a software DSM (Distributed Shared Memory) approach. The achievement of this objective requires work in compilers/source-to-source translators and the proposal of intelligent runtime libraries.
- **Resource management for parallel and GRID architectures**, including coordination with the runtime supporting the execution of the programming model above and the execution environment below. Resources considered are CPU and memory.
- **Execution environments for parallel and embedded systems**. This objective includes work at the level of kernel, monitoring and interception tools at the library level, and at the virtual machine level. Specialization of components and code reduction are also considered an objective when targeting embedded systems.
- **Improving the behavior of J2EE application servers**. To achieve this objective the group plans to develop a platform for the analysis of application servers, as a basis to understand their behavior and how the interaction with specialized self-managed resource management systems.
- **Scalability of tools for performance analysis and prediction of parallel systems**. The objective here is to extend the tools already developed (instrumentation, visualization in Paraver and simulation in Dimemas) in different directions, including more powerful analysis, scalability, portability and modeling of a large variety of target architectures.

WP3: High-performance applications and algorithms (subproject 1)

- **Optimization of numerical and signal processing applications**. This objective included the proposal of novel algorithms to solve systems of equations (dense and sparse) and techniques for automatic differentiation. Also the parallelization of applications in the field of signal processing, mainly voice processing.
- **Optimization of information management systems**, including the database engine, algorithms for data mining, sorting, …

The project also defines a number of tasks to achieve the objectives above, all of them with a duration of 3 years. Table 1 lists the tasks and includes information about the executing subproject as well as the objective that the task contributes.

Task	Title	Subproject	Contribution to objective:
WP1: Computer architecture			
AC1.1	Simple out-of-order architectures	1	1
AC1.2	VLIW architectures	1	1
AC1.3	Low-power architectures	1	1
AC1.4	Speculative architectures	1	1
AC1.5	QoS in SMT architectures	1	1
AC1.6	Hw-Sw co-design for reconfigurable architectures	1	1
AC1.7	Scalable components design	1	1
AC1.8	High-performance fetch engines in future architectures	1	1
AC1.9	Code generation for VLIW architectures	1	1
AC1.10	Clustered microarchitectures	1	1
AC1.11	Application specific architectures: Network processors	1	1
AC1.12	Branch predictors	1	1
AC1.13	Resource pressure in SMT architectures	1	1
AC1.14	Highly out-of-order architectures	1	1
AC1.15	Detection of repetitive code structures	1	1
AC2.1	Register file design	1	2
AC2.2	Efficient register renaming technique	2	2
AC2.3	Multilevel prefetch and energy consumption	2	2
AC2.4	Distributed first level caches	2	2
AC2.5	Coherence optimization in multiprocessor systems	2	2
AC3.1	Architectures for multimedia applications	1	3
AC3.2	Architectures for 3D graphic applications	1	3
AC3.3	Exploitation of DLP in general-purpose architectures	1	3
AC3.4	Multimedia extensions for general-purpose architectures	1	3
AC3.5	Vector units in superscalar processors	1	3
AC4.1.1	Formal models to specify synchronous and asynchronous systems	1	4
AC4.1.2	Definition of techniques for system abstraction	1	4
AC4.2.1	Efficient exploration of the state space	1	4
AC4.2.2	System falsification	1	4
AC4.2.3	Reduction gears of the complexity: hierarchy, composition and abstraction	1	4
AC4.3.1	Verification of systems with relative timing	1	4
AC4.3.2	Automatic extraction of the causality in the temporary dominion	1	4
AC4.3.3	Falsification of temporary systems	1	4
WP2: System software for high-performance computing			
SS1.1	OpenMP extensions for Fortran and C	1	5
SS1.2	OpenMP execution in multithreaded systems	1	5
SS1.3	OpenMP execution in sDSM systems	1	5
SS1.4	OpernMP benchmarking	1	5
SS1.5	Dynamic optimization of OpenMP applications	1	5
SS1.6	Dynamic optimizations of .net applications	1	5
SS1.7	MPI scalability	1	5
SS1.8	Definition and implementation of GRID programming models	1	5

Table 1: Workpackages and tasks in High-Performance Computing IV.

Task	Title	Subproject	Contribution to objective:
SS2.1	Resource management for OpenMP and MPI applications	1	6
SS2.2	Resource management in GRID environments	1	6
SS2.3	Memory management in Java environments	1	6
SS3.1	System tool towards efficient execution environments	1	7
SS3.2	Execution environments for embedded systems	1	7
SS4.1	Performance analysis and visualization of J2EE environments	1	8
SS4.2	J2EE application servers modeling	1	8
SS4.3	Cooperation between the execution environment and J2EE servers	1	8
SS4.4	Dynamic monitoring and management of J2EE servers	1	8
SS5.1	Semantic module in Paraver	1	9
SS5.2	Scalable instrumentation	1	9
SS5.3	Scalable analysis and visualization	1	9
SS5.4	Other instrumentation packages	1	9
SS5.5	Application optimization methodology	1	9
SS5.6	Performance models	1	9
SS6.1	Algorithm optimization for SMT architectures	1	9
WP3: High-performance applications and algorithms			
AA1.1	Automatic differentiation	1	10
AA1.2	Resolution of sparse systems of equations	1	10
AA1.3	Voice processing	1	10
AA2.1	Improving Join operations	1	11
AA2.2	Improving queries with more than 100 joins	1	11
AA2.3	XML databases	1	11
AA2.4	Tool for file fusion	1	11
AA2.5	Data management in file fusion systems	1	11
AA2.6	Arrangement for data sets in memory	1	11
AA2.7	On-the-fly data sorting	1	11
AA2.8	Memory management in the database engine	1	11
AA2.9	Database engine, operating system and architecture	1	11

Table 1 (continuation): Workpackages and tasks in High-Performance Computing IV.

2 Success degree in achieving project objectives

Measuring success in a big project like High-Performance Computing IV is not a trivial task. First because there is a large number of researchers involved coming from different Universities and research groups. Second, the research activities in this project are a continuation of the research that has been carried out, at least, over the last decade in high-performance computing in the different groups; therefore it is very difficult to associate the results with this individual project and have to be seen as the results of a research trajectory supported by and financed with a number of previous basic research projects from the Spanish Ministry of Science and Technology (as shown in Figure 1).

In this section we provide global numbers that allow the reader to measure how the overall project has advanced and produced scientific results. A detailed description of the results achieved in each one of the tasks as well as the associated publications has not been included in order to avoid a too large document. This document is available at [3]. An exhaustive list of the publications of the project, as well as many of the articles in electronic format, can be found in [1][2]. The quantitative results of our research group are summarized in Table 2.

	2005-06	Last 5 years	Result of coordinated activities (2005-06)
PhD thesis	13	31	2
Journals	24	69	3
High-quality conference proceedings	22	65	3
Other conferences	90	223	7
Book chapters	3	10	0

Table 2: Thesis and publications derived from High-Performance Computing IV (period 2005-06) and III (period 2002-04). Last column indicated results that have been generated as a result of the coordinated research activities between different groups in High-Performance Computing IV.

When showing the global numbers we will differentiate between the project period (2005-06) and the last five years (2002-06). Other relevant indicators of success will be commented in Section 3 (projects and collaboration with external researchers). In this section we will also comment about the success of the coordinated activities done in the two subprojects. After that, we will comment the major difficulties in pursuing the research activities initially proposed.

It is important to mention that the collaboration between researchers of the three groups in the project has been very successful. We can measure the success in this direction in terms of number of PhD thesis and publications done in collaboration. This information is shown in the last column in Table 2.

Finally, some deviations from the initial workplan have happened:

- First, by the end of 2005 two researchers originally in the project decided to abandon it, to create their own research group and ask for independent CICYT projects. This mainly affects the development of tasks AA2.1-AA2.9.
- Second, a small number of researchers change their priorities in research because they moved to a different campus in UPC with new global research priorities. This mainly affects to 5 researchers, who were in charge of conducting the tasks AC4.1.1-AC4.3.3.
- And third, other topics of research, that were not considered when writing the project proposal, have been started. The most relevant are:
  - Research activities in multi-core systems and transactional memory.
  - Parallelization of applications and the development of programming models and execution environments for heterogeneous multi-core architectures like the IBM/Sony/Toshiba Cell BE chip.

3 Project results

In this section we detail the results produced by the researchers in the project during the two first years of High-Performance Computing IV.

PhD thesis

During the period 2005-206 the following PhD thesis have been presented:

- Yolanda Becerra. Mejora del rendimiento de las aplicaciones Java usando cooperacion entre el sisttema operativo y la maquina virtual de Java. Universitat Politècnica de Catalunya (UPC), September 2006. Advisors: Toni Cortés, Jordi Garcia and José I. Navarro.
- Josep R. Herrero. A Framework for Efficient Execution of Matrix Computations. Universitat Politècnica de Catalunya (UPC), July 2006. Advisor: Juan J. Navarro.
- Adrian Cristal. Kilo Instruction Processors. Universitat Politècnica de Catalunya (UPC), April 2006. Advisors: Josep Llosa and Mateo Valero.
- Juanjo Noguera. Energy-efficient hardware/software co-design for dynamically reconfigurable architectures. Universitat Politècnica de Catalunya (UPC), December 2005. Advisor: Rosa M. Badia.
- Carlos Molina, "Microarchitectural Techniques to Exploit Repetitive Computations and Values", Universitat Politècnica de Catalunya (UPC), December 2005. Advisors: Antonio González and Jordi Tubella.
- Jordi Guitart. Performance Improvement of Multithreaded Java Applications Execution on Multiprocessor Systems. Universitat Politècnica de Catalunya (UPC), November 2005. Advisors: Jordi Torres and Eduard Ayguadé.
- Alex Pajuelo. Speculative Vectorization for Superscalar Processors. Universitat Politècnica de Catalunya (UPC), November 2005. Advisor: Antonio González and Mateo Valero.
- Enric, Gibert, "Clustered Data Cache Designs for VLIW Processors", Universitat Politècnica de Catalunya (UPC), November 2005. Advisors: Antonio González and Jesús Sánchez.
- Francisco J. Cazorla. Quality of Service for SMT processors. Universitat Politècnica de Catalunya (UPC), October 2005. Advisors: Mateo Valero, Alex Ramírez and Enrique Fernandez.
- Jaume Abella, "Adaptive and Low-Complexity Microarchitectures for Power Reduction", Universitat Politècnica de Catalunya (UPC), July 2005. Advisor: Antonio González.
- E. Torres. Alternativas de Diseño en Memoria Cache de Primer Nivel Multibanco. Dpto. Informática e Ingeniería de Sistemas, U. de Zaragoza. Junio 2005. Advisors: Pablo Ibañez and Victor Viñals.
- Oliverio J. Santana. Advanced Stream Prediction. Universitat Politècnica de Catalunya (UPC), May 2005. Advisors: Mateo Valero and Alex Ramírez.
- Ayose Falcón. Fetch improvement mechanisms for next-generation processors. Universitat Politècnica de Catalunya (UPC), February 2005. Advisors: Alex Ramírez and Mateo Valero.

Publications in international journals

During the period 2005-206 we have published 24 articles in international journals, 16 of them in indexed journals in the JCR with the impact factors that are shown in the Table 3. Some of them (12) include research results done in collaboration with researchers from other groups outside the project (international collaborations).

JOURNAL	Number 2005-06	Impact Factor
IEEE TRANSACTIONS ON COMPUTERS	4	1.875
IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS	2	1.462
ACM TRANSACTIONS ON PROGRAMMING LANGUAGES AND SYSTEMS	1	1.404
IEEE MICRO	3	1.238
IBM JOURNAL OF RESEARCH AND DEVELOPMENT	2	1.039
JOURNAL OF PARALLEL AND DISTRIBUTED COMPUTING	2	0.900
IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS	1	0.860
CONCURRENCY AND COMPUTATION-PRACTICE & EXPERIENCE	1	0.535

Table 3: Journals indexed in JCR where researchers in the project have published their results.

The articles published in international journals in the 2005-06 period are listed below:

- Javier Verdu, Jorge García, Mario Nemirovsky and Mateo Valero. The Impact of Traffic Aggregation on the Memory Performance of Networking Applications. Journal of Embedded Computing, pp. 77-82, vol. 2, no. 1, October 2006.
- J. Alastruey, J. L. Briz, P. Ibañez, V. Viñals, Software Demand, Hardware Supply. IEEE Micro, vol. 26, n. 4, Jul/Aug. 2006.
- Francisco J. Cazorla, Peter Knijnenburg, Rizos Sakellariou, Enrique Fernandez, Alex Ramírez and Mateo Valero. Predictable Performance in SMT processors: Synergy Between the OS and SMT. IEEE Transactions on Computers, pp. 785-799, vol. 55, no. 7, July 2006.
- Juanjo Noguera and Rosa M. Badia. System-Level Power-Performance Trade-Offs for Reconfigurable Computing. IEEE Transactions on VLSI, pp. 730-739, vol. 14, no. 7, July 2006.
- Raül Sirvent, Josep M. Pérez, Rosa M. Badia and Jesús Labarta. Automatic Grid workflow based on imperative programming languages. Concurrency and Computation, pp. 1169-1186, vol. 18, no. 10, July 2006.
- Eduard Ayguadé, Marc González, Xavier Martorell and Gabriele Jost. Employing Nested OpenMP for the Parallelization of Multi-Zone Computational Fluid Dynamics Applications. Journal of Parallel and Distributed Computing, pp. 686-697, vol. 66, no. 5, May 2006.
- Juan Jose Costa, Toni Cortés, Xavier Martorell, Eduard Ayguadé and Jesús Labarta. Running OpenMP application efficiently on an everything-shared SDSM. Journal of Parallel and Distributed Computing, pp. 647-658, vol. 66, no. 5, May 2006.
- Ronald D. Barnes, Erik M. Nystrom, John W. Sias, Sanjay J. Patel, José I. Navarro and Wen-mei W. Hwu. Beating Beating In-Order Stalls with "Flea-Flicker" Two-Pass Pipelining. IEEE Transactions on Computers, pp. 18-33, vol. 55, no. 1, January 2006.
- Tomer Morad, Uri Weiser, Avinoam Kolodny, Mateo Valero and Eduard Ayguadé. Performance, Power Efficiency and Scalability of Asymmetric Chip Multiprocessors. Computer Architecture Letters, pp. 14-17, vol. 5, no. 1, January 2006.
- María Jesús Garzarán, Milos Prvulovic, José M. Llabería, Victor Viñals, Lawrence Rauchwerger and Josep Torrellas. Tradeoffs in buffering speculative memory state for thread-level speculation in multiprocessors . ACM Transactions on Architecture and Code Optimization, pp. 247-279, vol. 2, no. 3, September 2005.

- Xavier Vera, Jaume Abella, Josep Llosa and Antonio González. An accurate cost model for guiding data locality transformations. ACM Transactions on Programming Languages and Systems, pp. 946-987, vol. 27, no. 5, September 2005.
- Carlos Álvarez, Jesús Corbal and Mateo Valero. Fuzzy Memoization for Floating-Point Multimedia Applications. IEEE Transactions on Computers, pp. 922-927, vol. 54, no. 7, July 2005.
- Julita Corbalán, Xavier Martorell and Jesús Labarta. Performance-Driven Processor Allocation. IEEE Transactions on Parallel and Distributed Systems, pp. 599-611, vol. 16, July 2005.
- Esther Salamí and Mateo Valero. Dynamic Memory Interval Test vs. Interprocedural Pointer Analysis in Multimedia Applications. ACM Transactions on Architecture and Code Optimization, pp. 199-219, vol. 2, no. 2, June 2005.
- Alex Pajuelo, Antonio González and Mateo Valero. Speculative Execution for Hiding Memory Latency. ACM, Computer Architecture News, pp. 49-56, vol. 33, no. 3, June 2005.
- Javier Verdu, Jorge García, Mario Nemirovsky and Mateo Valero. The Impact of Traffic Aggregation on the Memory Performance of Networking Applications. ACM, Computer Architecture News, pp. 57-62, vol. 33, no. 3, June 2005.
- Adrian Cristal, Oliverio J. Santana, Francisco J. Cazorla, Marco Galluzzi, Tanausu Ramirez, Miquel Pericas and Mateo Valero. Kilo-instruction Processors: Overcoming the Memory Wall. IEEE Micro, pp. 48-57, vol. 25, no. 3, May 2005.
- Javier Larrosa Bondía, Enric Morancho and David Niso. On the Practical use of Variable Elimination in Constraint Optimitzation Problems: 'Still-life' as a Case Study. Journal of Artificial Intelligence Research, pp. 421-440, vol. 23, April 2005. [Cited by 5]
- Marisa Gil and Ruben Pinilla. Thread Coloring: A Scheduler Proposal from User to Hardware Threads. ACM Operating Systems Review, pp. 54-70, vol. 39, no. 2, April 2005. Linux, HT, multithreading.
- George Almasi, Charles Archer, Jose Castaños, John Gunnels, Chris Erway, Philip Heidelberger, Xavier Martorell, Jose Moreira, Kurt Pinnow, Joseph Ratterman, Burkhard Steinmacher-Burow, William Gropp and Brian Toonen. Design and Implementation of Message-Passing Services for the Blue Gene/L Supercomputer. IBM Journal of Research and Development, pp. 393-406, vol. 49, no. 2, March 2005.
- Xavier Martorell, Nils Smeds, Robert Walkup, Jose Brunheroto, George Almasi, John Gunnels, Luiz DeRose, Jesús Labarta, Francesc Escalé, Judit Gimenez, Harald Servat and Jose Moreira. Blue Gene/L Performance Tools. IBM Journal of Research and Development, pp. 407-424, vol. 49, no. 2, March 2005.
- Joan-Manuel Parcerisa, Julio Sahuquillo, Antonio González and José Duato. "On-Chip Interconnects and Instruction Steering Schemes for Clustered Microarchitectures". IEEE Transactions on Parallel and Distributed Systems, 16(2):130-144, IEEE Computer Society, February 2005.
- Ayose Falcón, Jared Stark, Alex Ramírez, Konrad Lai and Mateo Valero. Better Branch Prediction through Prophet/Critic Hybrids. IEEE Micro, pp. 80-89, vol. 25, no. 1, January 2005.
- Alex Ramírez, Josep-L. Larriba-Pey and Mateo Valero. Software Trace Cache. IEEE Transactions on Computers, pp. 22-35, vol. 54, no. 1, January 2005.

Publications in international conferences

In the period 2005-06 we have published 112 papers in international conferences. There is an unbalance between publications in journals (24) and publications in conferences (112). This unbalance is justified. The average time of publication of an article in the journals listed above is superior to 2 years. The average time of publication of an article in a conference is around 6 months. If we take this difference and the speed of change in our research area into account, it is natural that the forums with more impact to present new ideas are the conferences.

In order to determine the quality of publications in international conferences, in a similar way it is done in the SCI Journal Citation Report, we have used the ranking in http://www.cs-conference-ranking.org/conferencerankings/topicsiii.html, which is updated every three months. The ranking is based on the evaluation of EIC (Estimated Impact of Conference), which is normalized to be in the range 0.00-1.00 (the closer the number to 1.00, the better the conference). The current ranking just includes conferences with EIC above 0.50.

Of all the conferences in the above mentioned ranking, we have just selected those that the Universitat Politècnica de Catalunya (UPC) qualifies as high-quality conferences ("Congresos Notables"). The conditions that the UPC imposes to qualify a conference as "Congreso Notable" are the following:

- Minimum of 2 editions and 5 years old.
- Having 10 cites to articles published in the conference in journals situated in the first third of its category of the Journal Citation Reports (JCR). The cites will have to be in a period not superior to 1 year, and exclude self-citations.
- Papers are reviewed by international referees.

In addition, we can add that in the remarkable conferences of our research area:

- The paper revision is done by 4-6 international experts.
- The revision is based on the full paper.
- In many cases it is a blind revision, the referees do not know author's name.
- The final selection is carried out by a program committee.
- The acceptance ratio is less than 25%.

Table 4 shows the 10 international conferences in our research area that either belong to the "Congresos Notables" UPC list and their associated EIC. In addition, the table also shows the number of publications in these conferences generated by researchers in this project.

CONFERENCE	Number 2005-06	EIC
ISCA: ACM/IEEE Symposium on Computer Architecture	1	0.99
MICRO: International Symposium on Microarchitecture	1	0.96
SUPER: ACM/IEEE Supercomputing Conference	1	0.96
HPCA: IEEE Symposium on High-Performance Computer Architecture	0	0.96
IPDPS: International Parallel and Distributed Processing Symposium	4	0.91
PACT: IEEE Intl Conference on Parallel Architectures and Compilation Techniques	2	0.88
ICS: International Conference on Supercomputing	7	0.74
ICPP: International Conference on Parallel Processing	3	0.61
PLDI: ACM Conference on Programming Language Design and Implementation	2	na
ICCD: IEEE International Conference of Computer Design	1	na

Table 4: Contributions to the most relevant international conferences in our research area.

The papers published in international journals in the 2005-06 period are listed below:

- Josep M. Pérez, Rosa M. Badia and Jesús Labarta. Including SMP in Grids as execution platform and other extensions in GRID superscalar. In IEEE International Conference on e-Science and Grid Computing (eScience 2006), December 2006.
- Elzbieta Krepska, Thilo Kielmann, Raül Sirvent and Rosa M. Badia. A Service for Reliable Execution of Grid Applications. In CoreGRID Integration Workshop, Integrated Research in Grid Computing, December 2006.
- Francesc Guim, Ivan Rodero, Julita Corbalán, Jesús Labarta, A. Oleksiak, K. Kurowski and J. Nabrzyski. Integrating the Palantir Grid Meta-Information System with GRMS. In CoreGRID Integration Workshop, Integrated Research in Grid Computing, December 2006.
- Ivan Rodero, Francesc Guim, Julita Corbalán, Jesús Labarta, A. Oleksiak, K. Kurowski and J. Nabrzyski. Integration of the eNANOS execution framework with GRMS for GRID purposes. In CoreGRID Integration Workshop, Integrated Research in Grid Computing, December 2006.
- Rosa M. Badia, Raül Sirvent, Marian Bubak, Wlodzimierz Funika, Cezary Klus, Piotr Machner and Marcin Smetek. Performance Monitoring of GRID superscalar with OCM-G: Integration Issues. In CoreGRID Integration Workshop, Integrated Research in Grid Computing, December 2006.
- Raül Sirvent, Rosa M. Badia, Natalia Currle-Linde and Michael Resch. GRID Superscalar and Gricol: Integrating Different Programming Approaches. In CoreGRID Integration Workshop, Integrated Research in Grid Computing, December 2006.
- Pieter Bellens, Josep M. Pérez, Rosa M. Badia and Jesús Labarta. CellSs: A Programming Model for the Cell BE Architecture. In ACM/IEEE SC 2006 Conference, November 2006.
- Josep R. Herrero and Juan J. Navarro. A Framework for Accurate Measurements with Low Resolution Clocks. In The 10th IASTED International Conference on Software Engineering and Applications (SEA 2006), November 2006.
- Jordi Roca, Victor Moya, Carlos González, Chema Solis, Agustín Fernández and Roger Espasa. Workload Characterization of 3D Games. In 2006 IEEE International Symposium on Workload Characterization (IISWC-2006), October 2006.
- Ruben Gran, Enric Morancho, Angel Olivé and José M. Llabería. An Enhancement for a Scheduling Logic Pipelined over two Cycles. In XXIV IEEE International Conference on Computer Design (ICCD-2006), October 2006.
- L.M. Ramos, J.L. Briz, P. Ibáñez and V. Viñals. "Data prefetching in a cache hierarchy with high bandwith and capacity".In Procs. of the 7th MEDEA Workshop. pp. 37-44, Sep. 2006
- Oliverio J. Santana, Ayose Falcón, Alex Ramírez and Mateo Valero. Branch Predictor Guided Instruction Decoding. In Intl. IEEE Conference on Parallel Architectures and Compilation Techniques (PACT-2006), September 2006.
- Pau Bofill and Enric Monte Moreno. Towards Underdetermined Source Reconstruction from a Clap-and-Play Binaural Live Recording. In European Signal Processing Conference EUSIPCO, September 2006.

- A. de Dios, B. Sahelices, Pablo Ibáñez, Victor Viñals and José M. Llabería. Speeding-Up Synchronizations in DSM Multiprocessors. In Euro-Par 2006, September 2006.
- Francesc Guim, Ivan Rodero, M. Tomas, Julita Corbalán and Jesús Labarta. The Palantir grid meta-information system. In 7th IEEE/ACM International Conference on Grid Computing (GRID2006), September 2006.
- Ramon Nou, Ferran Julià, David Carrera, Kevin Hogan, Jordi Caubet, Jesús Labarta and Jordi Torres. Monitoring and analysing a grid middleware node. In 7th IEEE/ACM International Conference on Grid Computing (GRID2006), September 2006.
- José Luis González and Toni Cortés. Evaluating the Effects of Upgrading Heterogeneous Disk Arrays. In International Symposium on Performance Evaluation of Computer and Telecommunication Systems , July 2006. Best paper award.
- L. C. Aparicio, J. Segarra, J. L. Villarroel, and V. Viñals. "WCET Computation in Presence of Caches". In Proc. Advanced Computer Architecture and Compilation for Embedded Systems (ACACES 05), pp. 205-208, Jul. 2006.
- Isaac Gelado, Enric Morancho and José I. Navarro. Hybrid Hardware/Sotfware Co-execution using Continuations. In 2006 Advanced Computer Architecture and Compilation for Embedded Systems (ACACES-06), July 2006.
- Manoj Gupta, Josep Llosa and Fermín Sánchez. Cluster Level Multithreading for VLIW Processors. In 2006 Advanced Computer Architecture and Compilation for Embedded Systems (ACACES-06), July 2006.
- Nikolaos Galanis and Josep Llosa. Statistical Simulation of Large-Scale Multicomputer Systems. In 2006 Advanced Computer Architecture and Compilation for Embedded Systems (ACACES-06), July 2006.
- Ramon Bertran, Marisa Gil, Javier Cabezas, Victor Jimenez, Lluis Vilanova, Enric Morancho and José I. Navarro. An Experimental Framework for Whole System Optimization. In 2006 Advanced Computer Architecture and Compilation for Embedded Systems (ACACES-06), July 2006.
- Ramon Nou, Jordi Guitart, David Carrera and Jordi Torres. Experiences with Simulations - a Light and Fast Model for Secure Web Applications. In 12th International Conference on Parallel and Distributed Systems (ICPADS'06), July 2006.
- Ruben Gran, Enric Morancho, Angel Olivé and José M. Llabería. Non-Speculative Enhancements for a Pipelined Scheduling Logic. In 2006 Advanced Computer Architecture and Compilation for Embedded Systems (ACACES-06), July 2006.
- Matteo Monchiero, Ramon Canal and Antonio González, Design Space Exploration for Multicore Architectures: A Power/Performance/Thermal View, In 20th ACM International Conference on Supercomputing (ICS'06), June 2006.
- David Oro, Ramon Canal, Antonio González and James E. Smith, Power/Performance/Thermal trade-offs in microarchitecture, Intel Academic Forum 2006 Dublin (Ireland), June 2006.
- Matteo Monchiero, Ramon Canal and Antonio González, Design Space Exploration for MulticoreArchitectures: A Power/Performance/Thermal View, Intel Academic Forum 2006 Dublin (Ireland), June 2006.
- Josep R. Herrero and Juan J. Navarro. Sparse Hypermatrix Cholesky: Customization for High Performance. In International MultiConference of Engineers and Computer Scientists (IMECS 2006), June 2006.

- Montse Farreras, Toni Cortés, Jesús Labarta and George Almasi. Scaling MPI to short-memory MPPs such as BG/L. In 2006 International Conference on Supercomputing (ICS06), June 2006.
- Eduardo.Quiñones, Joan M. Parcerisa, Antonio Gonzalez. Selective Predicate Prediction for Out-of-Order Processors. In ICS'06: Proceedings of the 20th annual international conference on Supercomputing, June 2006.
- Isaac Gelado, Enric Morancho and José I. Navarro. Experimental Support for Reconfigurable Application-Specific Accelerators. In 2nd Workshop on the Interaction between Operating Systems and Computer Architecture (WIOSCA), June 2006.
- Ramon Bertran, Marisa Gil, Javier Cabezas, Victor Jimenez, Lluis Vilanova, Enric Morancho and José I. Navarro. Building a Global System View for Optimization Purposes. In 2nd Workshop on the Interaction between Operating Systems and Computer Architecture (WIOSCA), June 2006.
- Josep R. Herrero and Juan J. Navarro. Using nonlinear array layouts in dense matrix operations. In Workshop on state-of-the-art in scientific computing (PARA'06), June 2006.
- Josep R. Herrero and Juan J. Navarro. A Study on Load Imbalance in Parallel Hypermatrix Multiplication using OpenMP. In 11th International Workshop on High-Level Parallel Programming Models and Supportive Environments, June 2006.
- Josep R. Herrero and Juan J. Navarro. Adapting Linear Algebra Codes to the Memory Hierarchy Using a Hypermatrix Scheme. In 11th International Workshop on High-Level Parallel Programming Models and Supportive Environments, June 2006.
- Roberto Morales, Francesc Altisent, José I. Navarro and Marisa Gil. Hacia una Comunicación Simple y Eficiente entre Dispositivos Inalámbricos Heterogéneos. In II Congreso IberoAmericano sobre Computación Ubicua (CICU'06), June 2006.
- Yolanda Becerra, Jordi Garcia, Toni Cortés and José I. Navarro. Java Virtual Machine: the key for accurated memory prefetching. In The 2006 International Conference on Programming Languages and Compilers, June 2006.
- Ivan Rodero, Francesc Guim, Julita Corbalán and Jesús Labarta. How the JSDL can Exploit the Parallelism?. In 6th International Symposium on Cluster Computing and the Grid (CCGrid2006), May 2006.
- Francesc Guim, Ivan Rodero, Julita Corbalán, A. Oleksiak, D. Szejnfeld and J. Nabrzyski. Uniform Job Monitoring using the HPC-Europa Single Point of Access. In International Workshop on Grid Testbeds, in conjunction with CCGrid2006, May 2006.
- Josep R. Herrero and Juan J. Navarro. Compiler-Optimized Kernels: An Efficient Alternative to Hand-Coded Inner Kernels. In International Conference on Computational Science and its Applications (ICCSA 2006), May 2006.
- Tanausu Ramirez, Adrian Cristal, Oliverio J. Santana, Alex Pajuelo and Mateo Valero. Kilo-Instruction Processors, RunAhead and Prefetch. In Conference on Computing Frontiers, May 2006.
- J. Alastruey, T. Monreal, V. Viñals and M. Valero, "Speculative Early Register Release". ACM Int'l Conf. on Computing Frontiers (ICCF 06), May 2006.
- Xavier Martorell, Marc González, Alejandro Duran, Jairo Balart, Roger Ferrer, Eduard Ayguadé and Jesús Labarta. Techniques Supporting threadprivate in OpenMP. In 11th

International Workshop on High-Level Parallel Programming Models and Supportive Environments, April 2006.

- X. Vera, J. Abella, O. Unsal, A. González, E. Ergin. Checker Backend for Soft and Timing Error Detection and Recovery. En el Second Workshop on System Effects of Logic Soft Errors (SELSE'06), Abril 2006.
- Christopher Barton, Calin Cascaval, George Almasi, Yili Zheng, Montse Farreras, Siddhartha Chatterjee and José Nelson Amaral. Shared Memory Programming for Large Scale Machines. In 2006 Conference on Programming Language Design and Implementation (PLDI 2006), March 2006.
- Victor Moya, Carlos González, Jordi Roca, Agustín Fernández and Roger Espasa. ATTILA: A Cycle-Level Execution-Driven Simulator for Modern GPU Architectures. In IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS-2006), March 2006.
- Pau Bofill and Enric Monte Moreno. Underdetermined Convoluted Source Reconstruction Using LP and SOCP, and a Neural Approximator of the Optimizer. In International Conference on Independent Component Analysis and Blind Source Separation, ICA'06, March 2006.
- Enric Morancho, José M. Llabería and Angel Olivé. Predicting L2 Misses to Increase Issue-Queue Efficacy. In 4th workshop on Memory Performance Issues (WMPI-2006), February 2006.
- Alejandro Duran, Roger Ferrer, Juan Jose Costa, Marc González, Xavier Martorell, Eduard Ayguadé and Jesús Labarta. A proposal for error handling in OpenMP. In 2nd International Workshop on OpenmP (IWOMP 2006), January 2006.
- Friman Sanchez, Esther Salamí, Alex Ramírez and Mateo Valero. Performance Analysis of Sequence Alignment Applications. In 2006 IEEE International Symposium on Workload Characterization (IISWC-2006), January 2006.
- Javier Vera Gómez, Francisco J. Cazorla, Alex Pajuelo, Oliverio J. Santana, Enrique Fernández and Mateo Valero. A Novel Evaluation Methodology to Obtain Fair Measurements in Multithreaded Architectures. In Workshop on Modeling Benchmarking and Simulation, January 2006.
- Ramon Bertran, Marisa Gil, Javier Cabezas, Victor Jimenez, Lluis Vilanova, Enric Morancho and José I. Navarro. Opportunities for Global Optimization: Breaking the Boundaries Across System Components. In EuroSys 2006, January 2006.
- Tanausu Ramirez, Alex Pajuelo, Oliverio J. Santana and Mateo Valero. A Simple Speculative Load Control Mechanism for Energy Saving. In MEDEA Workshop, January 2006.
- Beatriz Otero, José M. Cela, Jesús Labarta and Rosa M. Badia. Performance Analysis of Domain Decomposition Applications using Unbalanced Strategies in Grid Environments. In The fourth International Conference on Grid and cooperative computing (GCC05), November 2005.
- Róbert Lovas, Raül Sirvent, Gergely Sipos, Josep M. Pérez, Rosa M. Badia and Péter Kacsuk. GRID superscalar enabled P-GRADE portal. In First CoreGrid Integrated Research in Grid Computing Workshop, November 2005.

- Rosa M. Badia, Marian Bubak, Wlodzimierz Funika and Marcin Smetek. Performance monitoring of Grid superscalar applications with OCM-G. In First CoreGrid Integrated Research in Grid Computing Workshop, November 2005.
- Rosa M. Badia, Olav Beckmann, Marian Bubak, Denis Caromel, Vladimir Getov, Stavros Isaiadis, Vladimir Lazarov, Maciek Malawski, Sofia Panagiotidi and Jeyarajan Thiyagalingam. Lightweight Grid Platform: Design Methodology. In First CoreGrid Integrated Research in Grid Computing Workshop, November 2005.
- Raül Sirvent, Andre Merzky, Rosa M. Badia and Thilo Kielmann. GRID superscalar and SAGA: forming a high-level and platform-independent Grid programming environment. In First CoreGrid Integrated Research in Grid Computing Workshop, November 2005.
- Victor Moya, Carlos González, Jordi Roca, Agustín Fernández and Roger Espasa. A Single (Unified) Shader GPU Microarchitecture for Embedded Systems. In First International Conference on High Performance Embedded Architectures and Compilers (HiPEAC 2005), November 2005.
- Victor Moya, Carlos González, Jordi Roca, Agustín Fernández and Roger Espasa. Shader Performance Analysis on a Modern GPU Architecture. In The 38th Annual IEEE/ACM International Symposium on Microarchitecture, 2005 (MICRO-38), November 2005.
- Stefan Bieschewski, Joan Manuel Parcerisa, Antonio González, "Memory Bank Predictors". International Conference on Computer Design, San Jose, USA, October, 2005.
- Beatriz Otero, José M. Cela, Rosa M. Badia and Jesús Labarta. Data Distribution Strategies for Domain Decomposition Applications in Grid Environments . In The 6th International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP-2005), October 2005.
- Mauricio Alvarez, Esther Salamí, Alex Ramírez and Mateo Valero. A performance characterization of high definition digital video decoding using H.264/AVC. In 2005 IEEE International Symposium on Workload Characterization (IISWC 2005), October 2005.
- Vicenç Beltran, David Carrera, Jordi Guitart, Jordi Torres and Eduard Ayguadé. A Hybrid Web Server Architecture for Secure e-Business Web Applications. In International Conference on High Performance Computing and Communications (HPCC'05), September 2005.
- Oliverio J. Santana, Alex Ramírez and Mateo Valero. Multiple Stream Prediction. In 6th International Symposium on High Performance Computing (ISHPC-VI 2005), September 2005. Distinguished Paper Award.
- Sebastián Reyes, Camelia Muñoz-Caro, Alfonso Niño, Rosa M. Badia and José M. Cela. Development and performance of a grid computing approach to the massive exploration of potential energy hypersurfaces for spectroscopic studies. In The Nineteenth Colloquium on High Resolution Molecular Spectroscopy, September 2005.
- E. Gibert, J. Abella, J. Sánchez, X. Vera, A. González, "Variable-Based Multi-Module Data Caches for Clustered VLIW Processors", to appear in Proceedings of 14th International Conference on Parallel Architectures and Compilation Techniques, Sept. 2005
- Francisco J. Cazorla, Peter Knijnenburg, Rizos Sakellariou, Enrique Fernandez, Alex Ramírez and Mateo Valero. Architectural Support for Real-Time Task Scheduling in SMT

Processors. In 2005 ACM/IEEE International Conference on Compilers, Architecture, and Synthesis for Embedded Systems (CASES-2005), September 2005.

- C. Molina, J. Tubella and A. González, "Reducing Misspeculation Penalty in Trace-Level Speculative Multithreaded Architectures", In Proceedings of the 6th International Symposium on High Performance Computing, September 2005.
- Gladys Utrera, Julita Corbalán and Jesús Labarta. Dynamic Load Balancing in MPI Jobs. In 6th International Symposium on High Performance Computing (ISHPC-VI 2005), September 2005.
- Javier Verdu, Jorge García, Mario Nemirovsky and Mateo Valero. Workload Characterization of Stateful Networking Applications. In 6th International Symposium on High Performance Computing (ISHPC-VI 2005), September 2005.
- Juanjo Noguera and Rosa M. Badia. Performance and Energy Analysis of Task-Level Graph Transformation Techniques for Dynamically Reconfigurable Architectures. In Field-Programmable Logic and Applications 2005 (FPL05), August 2005.
- Mauricio Alvarez, Esther Salamí, Alex Ramírez and Mateo Valero. Performance Evaluation of High Definition Digital Video Decoding Using the H.264/AVC Standard. In 2005 Advanced Computer Architecture and Compilation for Embedded Systems (ACACES-2005), July 2005.
- David Carrera, Vicenç Beltran, Jordi Torres and Eduard Ayguadé. A Hybrid Web Server Architecture for e-Commerce Applications. In The 11th International Conference on Parallel and Distributed Systems (ICPADS 2005), July 2005.
- Carmelo Acosta, Ayose Falcón, Alex Ramírez and Mateo Valero. Complexity-Effectiveness in Multithreading Architectures. In 2005 Advanced Computer Architecture and Compilation for Embedded Systems (ACACES-2005), July 2005.
- Friman Sanchez, Esther Salamí, Alex Ramírez and Mateo Valero. Parallel Processing in Sequence Matching. In 2005 Advanced Computer Architecture and Compilation for Embedded Systems (ACACES-2005), July 2005.
- Josep R. Herrero and Juan J. Navarro. Efficient Implementation of Nearest Neighbor Classification. In International Conference on Computer Recognition Systems (CORES 2005), July 2005.
- Tanausu Ramirez, Marco Galluzzi, Adrian Cristal and Mateo Valero. Different approaches using Kilo-instruction Processors. In 2005 Advanced Computer Architecture and Compilation for Embedded Systems (ACACES-2005), July 2005.
- D. Suárez, T. Monreal, and V. Viñals. Improvements on wire delay tolerant caches. In Proc. Advanced Computer Architecture and Compilation for Embedded Systems (ACACES-05), Jul. 2005..
- A. Aletà, J.M. Codina, A. González and D. Kaeli. “Demistifying On-the-fly Spill Code”, in Procs. of the Conf. on Programming Languages, Design and Implementation, June 2005.
- Alejandro Duran, Marc González, Julita Corbalán, Xavier Martorell, Eduard Ayguadé, Jesús Labarta and Raul Silvera. Automatic Thread Distribution for nested Parallelism in OpenMP. In The 19th ACM International Conference on Supercomputing (ICS05), June 2005.

- Carmelo Acosta, Ayose Falcón, Alex Ramírez and Mateo Valero. A Complexity-Effective Simultaneous Multithreading Architecture. In The 2005 International Conference on Parallel Processing (ICPP 2005), June 2005.
- Esther Salamí and Mateo Valero. A Vector-uSIMD-VLIW Architecture for Multimedia Applications. In The 2005 International Conference on Parallel Processing (ICPP 2005), June 2005.
- George Almasi, Charles Archer, Charles Erway, Philip Heidelberger, Xavier Martorell, Jose Moreira, Burkhard Steinmacher-Burow and Yili O. Zheng. Optimization of MPI Collective Communication on Blue Gene/L Systems. In The 19th ACM International Conference on Supercomputing (ICS05), June 2005.
- Rubén González, Adrian Cristal, Miquel Pericas and Mateo Valero. An asymetric Clustered Processor Based on Value Content. In The 19th ACM International Conference on Supercomputing (ICS05), June 2005.
- Gladys Utrera, Julita Corbalán and Jesús Labarta. Another approach to backfilled jobs: applying Virtual Malleability to expired windows. In The 19th ACM International Conference on Supercomputing (ICS05), June 2005.
- E. F. Torres, Pablo Ibáñez, Victor Viñals and José M. Llabería. Store Buffer Design in First-Level Multibanked Data Caches. In The 32nd Annual International Symposium on Computer Architecture (ISCA 2005) , June 2005.
- Jairo Balart, Alejandro Duran, Marc González, Xavier Martorell, Eduard Ayguadé and Jesús Labarta. Experiences parallelizing a Web Server with OpenMP. In 1st International Workshop on OpenMP (IWOMP 2005), June 2005.
- Jordi Guitart, David Carrera, Vicenç Beltran, Jordi Torres and Eduard Ayguadé. Session-Based Adaptive Overload Control for Secure Dynamic Web Applications. In The 2005 International Conference on Parallel Processing (ICPP 2005), June 2005.
- Josep R. Herrero and Juan J. Navarro. Intra-Block Amalgamation in Sparse Hypermatrix Cholesky Factorization. In International Conference on Computational Science and Engineering (ICCSE 2005), June 2005.
- Ayose Falcón, Alex Ramírez and Mateo Valero. Effective Instruction Prefetching via Fetch Prestaging. In The 19th International Parallel & Distributed Processing Symposium (IPDPS 2005), April 2005.
- Vasilis Dialinos, Rosa M. Badia, Raül Sirvent, Josep M. Pérez and Jesús Labarta. Implementing Phylogenetic Inference with GRID superscalar. In IEEE International Symposium on Cluster Computing and the Grid 2005 (CCGrid 2005), May 2005.
- Jordi Guitart, Vicenç Beltran, David Carrera, Jordi Torres and Eduard Ayguadé. Characterizing Secure Dynamic Web Applications Scalability. In The 19th International Parallel & Distributed Processing Symposium (IPDPS 2005), April 2005.
- David Rodenas, Xavier Martorell, Eduard Ayguadé, Jesús Labarta, George Almasi, Calin Cascaval, Jose Castaños and Jose Moreira. Optimizing NANOS OpenMP for the IBM Cyclops Multithreaded Architecture. In The 19th International Parallel & Distributed Processing Symposium (IPDPS 2005), April 2005.
- Alex Pajuelo, Antonio González and Mateo Valero. Control-Flow Independence Reuse via Dynamic Vectorization. In The 19th International Parallel & Distributed Processing Symposium (IPDPS 2005), April 2005.

- Raimir Holanda, Javier Verdu, Jorge García and Mateo Valero. Performance Analysis of a New Packet Trace Compressor based on TCP Flow Clustering. In 2005 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS 2005), March 2005.
- Friman Sanchez, Mauricio Alvarez, Esther Salamí, Alex Ramírez and Mateo Valero. On the Scalability of 1- and 2-Dimensional SIMD Extensions for Multimedia Applications. In 2005 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS 2005), March 2005.
- C. Molina, A. González and J. Tubella, "Compiler Analysis for Trace-Level Speculative Multithreaded Architectures", In Proceedings of the 9th Annual Workshop on Interaction between Compilers and Computer Architectures, San Francisco, United States, February 2005.
- David Carrera, David García, Jordi Torres, Eduard Ayguadé and Jesús Labarta. WAS Control Center: An Autonomic Performance-Triggered Tracing Environment for WebSphere. In The 13th Euromicro Conference on Parallel, Distributed and Network-based Processing (PDP 2005), February 2005.
- Ivan Rodero, Julita Corbalán, Rosa M. Badia and Jesús Labarta. eNANOS Grid Resource Broker. In Advances in Grid Computing (EGC 2005), February 2005.
- Jordi Guitart, David Carrera, Jordi Torres, Eduard Ayguadé and Jesús Labarta. Tuning Dynamic Web Applications using Fine-Grain Analysis. In The 13th Euromicro Conference on Parallel, Distributed and Network-based Processing (PDP 2005), February 2005.
- Friman Sanchez, Esther Salamí, Alex Ramírez and Mateo Valero. Parallel Processing in Biological Sequence Comparison using General Purpose Processors . In 2005 IEEE International Symposium on Workload Characterization (IISWC 2005), January 2005.
- Javier Verdu, Jorge García, Mario Nemirovsky and Mateo Valero. Architectural Impact of Stateful Networking Applications. In 1st Symposium on Architecture for Networking and Communications Systems (ANCS-I 2005), January 2005.

Participation in international projects

In the period 2005-06 the group has started a number of projects in the framework of the European Union and research contracts with companies. A brief summary of these projects and contracts follows:

- Research Project "Many-core architectures" with Microsoft on architecture and programming models for multi-core systems. Researchers from Microsoft Cambridge and Redmont participate in this project. The project is centered on the comparison and implementation of hardware and software based transactional memory systems. It also includes new proposals of programming models for imperative and functional languages base don this concept.
- Research projects with different groups in IBM T.J. Watson Research Center:
    - Cell BE Compiler Department. With this project we are contributing to the definition of new programming model proposals based in tasks for the Cell BE architecture, their implementation and evaluation.
    - High Volume web Serving Department. With this project the group is contributing to the design of adaptive middleware systems for distributed eBusiness execution environments.

  - Power Architecture Department. The goal of this project is to study the interaction between the operating system and the SMT/CMP architectures to improve the distribution of hardware resources. In particular, this study is focused on a real operating system, such as Linux, and a real SMT/CMP processor, such as IBM Power5.
  - Blue Gene/L team. The project with this team is focused in the scalability study of the MPI and UPC programming models and in the development of HPC Challenge applications to evaluate the performance and scalability of UPC.
- Research project with IBM Toronto Laboratory. In this project, we work with an implementation of OpenMP allowing dynamic adaptivity (self-tuning), in such a way that the runtime system can adapt to the application characteristics and the execution environment where it is executing. We have recently started a new project focused in the implementation of a software Distributed Shared Memory (DSM) to support OpenMP applications based on the IBM XLSMO library, for Power5/PowerPC cluster under Linux.
- Research projects with Intel, including the following topics: Operating Systems for Simultaneous Multithreading processors, Heterogeneity-Aware Multithreading Architectures, High performance Fetch for SMT Processors, Register File Organization for Future Superscalar Processors and Advanced Optimizations Enabled by the Stream Predictor.
- SARC (Scalable ARChitectures) European Project. The group is working in the definition and implementation of programming models and runtime systems for the Chip-Multiprocessor that will also be developed inside the project.
- ACOTES (Advanced Compiler Technologies for Embedded Streaming) European Project. The group is working in the development of a programming model, runtime system and abstract description machine oriented to streaming applications and architectures.
- HiPEAC European Network of Excellence in High-Performance Embedded Architecture and Compilation. The group is participating in a large number of topics related with static and dynamic compilation and code optimization tailored to embedded architectures, quality of service in both soft and hard-real time SMT scenarios, reconfigurable computing in embedded systems (for instance, network and multimedia processors), runtime optimization techniques for a large variety of applications, running on both general purpose and specific processors and collaborative embedded systems (services based on smartphones, and vehicle to vehicle communication) and their related resource management.
- CoreGRID European Network of Excellence in Grid computing. The group is contributing with the integration of Grid middlewares, and the design of the GRID superscalar programming model and runtime system.
- BEinGRID European IP Project. The research group will contribute with the integration of GRID superscalar with the GridWay metascheduler of the Universidad Complutense de Madrid. The system will also be evaluated with end-used GRID applications.
- SORMA European STREP Project. In this project, we are exploring new resource management models based on economic algorithms.

- BREIN European STREP Project. In this project, we are exploring the extension of resource management by using semantic web. In addition, we propose using virtualization techniques for full resource adaptation to the workload diversity.
- XtreemOS European Project. In this project the group is contributing in the design of services to manage the GRID. From CoreGRID we will obtain traces to evaluate the proposed policies. In addition we will integrate part of the implementation we will do in other already available systems that have been developed in other projects. The group will also contribute with the implementation of a file system for the GRID based on the usage of autonomic policies based on access patterns, device performance, and cooperation with the execution management systems to mange file replicas.

Collaborations with other researchers

In addition to the collaboration in the framework of the projects and contracts listed above, our research group also collaborates with a number of researchers:

- Daniel A. Jiménez from Rutgers University. The collaboration is established to improve the branch predictors in processors with large instruction windows, as the kilo-instruction processor.
- James Smith of the University of Wisconsin-Madison. The collaboration is kept in aspects related to the scalability of the microarchitectural structures of superscalar processors and in the implementation of hardware virtual machines.
- Alex Veidenbaum, University of California, Irvine. The collaboration focuses on the design of Value-driven architectures, that is, clustered architectures where the clusters depend on the bit-size of the operands.
- Mike O'Boile, University of Edinburgh. The collaboration focuses on the design of Value-driven architectures. He is involved in a project that tries to design a simple, decoupled embedded architecture that needs compiler help to generate code that can be executed efficiently.
- Peter Knijnenburg from the University of Amsterdam (Holland) and Rizos Sakellariou from the University of Manchester (United Kingdom) collaborate in the topic of resource allocation between the hardware contexts in a SMT architecture.
- Per Stenström, Computer Engineering Department, Chalmers University of Technology. He will collaborate in the design of chip multiprocessor for commercial and emergent workloads, implicit transactions and kilo-instruction multiprocessors.
- Ramon Beivide, Valentin Puente y Enrique Vallejo, Grupo de Arquitectura y Tecnología de Computadores, Universidad de Cantabria. Collaboration in implicit transactions and transactional memory.
- Oguz Ergin, TOBB University of Economics and Technology. Collaboration on utilizing transactional memory systems to implement fault-tolerant architectures.
- Mario Nemirovsky, Consentry Networks Inc. The collaboration is oriented to the support offered for validating new "stateful" processor architecture designs, as well as to validate problems and models related with network security.
- Roger Espasa and Victor Moya from Intel Labs Barcelona. The collaboration is established in specialized architectures for 3D graphics.
- Wen-mei Hwu, UIUC, USA. This collaboration is related to compilation for heterogeneous multi-core systems, and it includes research on the cooperation between

the operating system and the compiler to expose the hardware accelerators and the reconfigurable logic available to the applications.

- OpenMP Architecture Review Board. Several members of our research team are participating in the regular meetings of the ARB, contributing with proposals, discussions, and implementations, that allow to decide the cost of their inclusion in the OpenMP standard.
- Patrick Geoffray from Myricom. This cooperation Intend. To research on new dynamic routing mechanisms, needed to dynamically optimize the behavior of MPI at large installations.
- We have started some research with Prof. Stamatis Vassiliadis group on operating systems support for reconfigurable computing elements, taking into account power issues, too.
- Maria S. Pérez y Alberto Sanchez from Universidad Politécnica de Madrid. With this group we are currently cooperating in studying the viability of using Markov chains to predict the best storage systems where to write a file or replica.
- Database and Distributed Systems group (Samuel Kounev), Darmstadt University of Technology. In order to take autonomous decisions inside middleware, this collaboration will enable the analytic modeling and the simulation of runtime environments.
- Dependable System Group (Luis Silva) Coimbra University. The collaboration is focused on Self-healing topic, proposing early detection of potential problems and software rejuvenation techniques at level of service.
- Allan Snavely, San Diego Supercomputing Center (SDSC). We cooperate on the proper estimation of the length of computation phases between communications. Their estimates are fed into Dimemas for a better prediction of application performance on different platforms. Support for other programming models (LAPI) and I/O is also part of the collaboration.
- Bernd Mohr, Central Institute for Applied Mathematics (ZAM) en Juelich. We have jointly Developer translators from the trace format (EPILOG) used in his tool (KOJAK) to Paraver. We are also providing mechanisms for the integration of both tools, such that KOJAK can drive Paraver as well as the structure identification we will perform can drive KOJAK analysis.
- Phil Mucci, ICL (University of Tennessy at Kentucky). Cooperation will address issues of hardware counters and instrumentation techniques and environments.

4 References

[1] An exhaustive list of the publications for subproject 1, as well as many of the articles in electronic format, can be found in: "http://www.ac.upc.es/CAP/hpc/HPC-Publications.html".
[2] An exhaustive list of the publications for subproject 1, as well as many of the articles in electronic format, can be found in: "http://webdiis.unizar.es/gaz/publ.html".
[3] Annex to this document with a detailed description of the results achieved in all tasks. Download in electronic format from "http://www.ac.upc.es/CAP/hpc/HPC-Publications.html"

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

MIDHARD: Middleware and hardware solutions in high performance computation: Application to multimedia and simulation TIN2004-07797-C02

Javier Díaz Bruguera *
Departamento de Electrónica e Computación
Universidade de Santiago de Compostela

Ramón Doallo Biempica †
Departamento de Electrónica e Sistemas
Universidade de A Coruña

Abstract

The objectives of this project are organized into two large areas: (1) Development of middleware tools to improve the performance of applications for parallel and distributed systems, and to facilitate their administration/configuration taking into account the fault tolerance capacity, including the development of run time and compile time tools for extracting parallelism. The proposals will be validated with those applications that both groups are developed in previous projects, such as computer graphics, multimedia processing or numerical simulation (microelectronic devices, atmosphere pollution, etc.) (2) Development of application specific hardware and hardware for general purpose processors, to improve the performance for multimedia applications (architectures for improving the efficiency in the transmission of data from the PC to the GPU, design of a reconfigurable video processor and design of functional units).
Keywords: Parallel and distributed systems. Run–time and compile–time frameworks. Functional unit design. Application specific processors. High Performance Computing.

1 Project objectives

The development of parallel and distributed systems in the last years has lead to the development of middleware to help in an efficient management of the system's resources and the introduction of application specific hardware and the design of new functional units to improve the performance. However, some of these software and hardware tools have low performance and their application can be improved. This project tackles this topics from two points of

*Email: bruguera@dec.usc.es
†Email: doallo@udc.es

view: (1) Improvement of several middleware tools for parallel and distributed systems, including Grids. (2) Hardware for multimedia applications. In the following, these general objectives are detailed and decomposed in several more specific sub–objectives.

1.1 Middleware and compilation–time and run–time frameworks for parallel and distributed systems

1. Middleware for distributed systems.
   (a) Middleware architecture for the administration of distributed systems. This architecture is based on an object–oriented data model with some interesting characteristics: abstraction, classification, etc. This model can represent dependencies and relations between elements as well.
   (b) Middleware for process and data distribution, fault tolerance and resources planning in Grids. We will use our experience in analysis and determination of performance prediction models to tackle the automatic dynamic distribution of data and process, and to develop new monitoring and scheduling tools. On the other hand, we will develop a tool to carry out a source-to-source transformation of a MPI code in an application level fault-tolerant program.
   (c) Efficient communications in Java for cluster with SCI networks. We propose to develop a communication library for SCI and its integration in KaRMI, a Java library for communication over Myrinet networks. Some improvements on fault–tolerance, load balancing and processing distribution will be implemented.
2. Compile–time and run–time frameworks for parallel systems.
   (a) Compile–time framework for automatic parallelism detection. Our group has recently developed a compilation framework based on the GSA representation that it is able to identify a large amount of computational kernels. In this project we propose to enlarge the set of computational and to integrate it on the GNU GCC compiler.
   (b) Run–time tools for a efficient implementation of irregular codes. Irregular codes can be analyzed and optimized only at run–time. We propose to tackle the optimization of irregular codes, not only taking into account the parallelization of the code, but other problems as load balancing and memory hierarchy. Those techniques will be tailored to different architectures.
   (c) Iterative framework for the development of parallel applications. We propose the implementation, evaluation and optimization design for a parallel programming framework based on HTAs (a framework developed by our group in cooperation with other groups), that includes the development of a MATLAB toolbox and the implementation on a free software tool.

1.2 Hardware for multimedia

1. Information compression in the communication between host and GPU. The representation of complex objects and scenes in computer graphics requires a huge amount of

information, that has to be send from the CPU (host) to the GPU to carry out the rendering. The objective is to reduce the information to be send to the GPU. This reduction can be achieved by (1) Elimination of repeated computations in the NURBS tessellation, (2) Algorithms for mesh compression in point rendering applications and (3) Efficient coding of tetrahedral meshes.

2. Reconfigurable processor for video compression. The new standard H.264 for video compression requires special architectures. We propose the design of a low cost programmable architecture based on a FPGA combined with a general purpose processor. Then, new specialized functional units to deal with the most costly operations will be optimized. This process requires the development of new algorithms and architectures.

3. General purpose processor functional unit design. The development of multimedia applications demand more and more computational resources in the actual processors. As a consequence, new functional units are incorporated in the architecture. Then, we propose to improve some units that are widely used in computer graphics and multimedia applications: low latency floating–point MAF unit and calculation of the normalization operation.

2 Scheduling chart and available resources

In this section we show the scheduling chart proposed for the project. The objectives are listed here and table 1 shows the planning for the activities of the project:

1. Middleware architecture for distributed systems administration: (1) Development of a model-driven framework based for the systematic development of integrated system administration applications. (2) Design of a high–level administration tool

2. Middleware for Grid: (1) Prototype deployment of a Grid infrastructure using the computing resources of both subprojects. (2) Improvements of the performance models. (3) Process and data distribution and migration, task scheduling to improve performance on Grid environments. (4) Libraries design for fault tolerant MPI applications. (5) Compilation techniques for automatable checkpointing

3. Design of efficient communications in Java for clusters with SCI network: (1) Development of a high-performance sockets library for SCI. (2) Development of a high-performance Java RMI library for SCI

4. Compiler framework for automatic parallelism detection: (1) Extension of the collection of kernels recognized by XARK and its applicability to C, C++ or Java. (2) Improvement of current optimizing/parallelizing compilers using the XARK compiler

5. Compile–time and run–time solutions for irregular codes: (1) Improvement of the IARD representation. (2) Usage of the IARD representation and the analytical modeling based on probabilistic miss equations to increase locality in irregular accesses.

6. Iterative framework for the development of parallel applications: (1) Development of a MATLAB toolbox for an environment based on HTAs. (2) Analysis and implementation of optimizations for the HTA compiler, and migration to an open source environment

7. Application specific hardware for computer graphics: (1) Elimination of redundant computations in NURBS tessellation. (2) Algorithms and architectures for point rendering. (3) Efficient and fast coding of tetrahedral meshes.
8. Video compression processors: (1) Architecture of a basic video compression/decompression system. (2) Optimization of functional units
9. Functional unit design: (1) Architecture for the low-latency computation of the floating–point MAF operation. (2) Improvement of the vector normalization
10. Development of applications

Task	year 1						year 2						year 3					
1.1	X	X	X	X	X	X	X	X	X									
1.2				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2.1	X	X	X															
2.2	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
2.3				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2.4	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X			
2.5										X	X	X	X	X	X	X	X	X
3.1	X	X	X	X	X	X	X	X	X	X	X	X	X					
3.2								X	X	X	X	X	X	X	X	X	X	X
4.1	X	X	X	X	X	X	X	X	X	X	X	X	X					
4.2							X	X	X	X	X	X	X	X	X	X	X	X
5.1	X	X	X	X	X	X												
5.2	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
6.1	X	X	X	X	X	X												
6.2							X	X	X	X	X	X	X	X				
7.1	X	X	X	X	X	X	X											
7.2				X	X	X	X	X	X									
7.3									X	X	X	X	X	X	X	X	X	X
8.1	X	X	X	X	X	X	X	X	X	X	X	X						
8.2										X	X	X	X	X	X	X	X	X
9.1	X	X	X	X	X	X	X	X										
9.2								X	X	X	X	X	X	X				
10	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Table 1: Scheduling chart for the project

3 Success level achieved in the project

In this section we summarize the objectives have been currently achieved. Note that the project finish by December 2007 so, some objective that have not been achieved can be achieved before the end of the project. We indicate for each objective the results obtained and the publications derived.

1. Middleware architecture for the administration of distributed systems (Subproject 2). The main goal of this research line was the development of a model-driven framework based on the CIM model for the systematic development of integrated system administration applications. The result is the AdCIM middleware (see http://adcim.des.udc.es) [21, 22], which provides an infrastructure for: (1) Efficient representation and extension of the CIM model using a custom XML representation called miniCIM. (2) Grammar-based extraction of configuration and management information as miniCIM instances. (3) Transparent persistence of miniCIM data into a repository. (4) Querying and modifying miniCIM information backed by the repository through a web service. (5) Pregeneration and styling of XForms web user interfaces for the administration applications that can interface directly with the web service.

2. Middleware for Grid (Subprojects 1 and 2)

   (a) Prototype deployment of a Grid infrastructure using the computing resources of both subprojects (Subprojects 1 and 2). A Grid prototype infrastructure was deployed using the framework of REDEGRID (Rede Galega de computación Grid y distribuida). This prototype was built not only with the participation of the two groups of this project, but with the participation of other 5 galician research groups, and CESGA (Galician Supercomputing Center).

   (b) Improvements of the performance models (Subproject 1). We paid attention to irregular codes and to compute-intensive applications on Grids and MPI-based codes. All these models are based on values that characterize the system: the latency and bandwidth of the network, the number of CPU, their computational power, etc.; and currently we are studying the effect of the memory hierarchy and the number of resident processes in each node and their priority. We developed an interactive framewok based on CALL [19, 47] includes drivers to obtain monitoring information from NWS and Ganglia to on-line measure the values that characterize the state of the heterogeneous system, interfaces for the user to establish, tune and refine the parameters of the models and a statistical driver based on R to obtain the final best model and its accuracy.

   (c) Process and data distribution and migration, task scheduling to improve performance on Grid environments (Subprojects 1 and 2). We are currently developing a software infrastructure, based on web services, for supporting the execution of fault-tolerant parallel applications in Grid environments. This infrastructure attends execution requests from users, providing resource discovery and selection, job submission, monitoring and migration (if needed). During the execution, the infrastructure also manage the state files required to restart the failed or migrated executions. On the other hand, a strategy to efficiently distribute data in highly intensive computations on Grids were developed [45] and new strategies to balance the workload on heterogeneous systems are currently being developed [46].

   (d) Libraries design for fault tolerant MPI applications (Subproject 2). A tool named CPPC (Controller/Precompiler for Portable Checkpointing, http://cppc.des.udc.es) has been developed [48]. CPPC is a checkpointing tool focused on the insertion of fault tolerance into long-running message-passing applications. It uses a variable-level, non-coordinate, non-logging, portable checkpointing technique. CPPC consists of a library and a precompiler. The precompiler translates parallel code annotated with user directives into fault tolerant code with CPPC library calls.

   (e) Compilation techniques for automatable checkpointing (Subproject 2). The transparency of our solutions for fault tolerant MPI applications has been improved by automating the instrumentation necessary for the application to take advantage of the fault-tolerance library previously developed [49, 50]. Specifically, a source-to-source compiler built on the Cetus compiler infrastructure has been developed. The compiler automatically identifies the variables whose values are needed upon application restart, as well as the functions with non-portable outcome. Current work is focused towards a full automatization of the analysis process, by devising strategies and algorithms for automatically finding suitable places for inserting the

checkpointing directive, taking into account both the performance of the checkpoint operation and the communication state of processes to be checkpointed.

3. Design of efficient communications in Java for clusters with SCI network (Subproject 2). The result within this project was a Java communication middleware transparent to the user, interoperable with other systems, does not need source code modification and offers widely spread APIs (Java Sockets and Java RMI)[61, 62, 63, 64, 65, 66] , that consists of:

   (a) A high-performance sockets library, JFS (Java Fast Sockets) designed to cope with the requirements of Java parallel applications on clusters with high-speed interconnects (such as SCI). Every Java application can run on top of JFS, and even it can serve as transport layer for another Java communication library, such as Java RMI.

   (b) A high-performance Java RMI library, designed to increase performance through the use of JFS. The performance of RMI was improved focusing on: transport improvements (management of data to reduce buffering and socket delays); serialization overhead reduction; and RMI protocol overhead reduction (mainly stub and skeleton operation).

4. Compiler framework for automatic parallelism detection (Subproject 2). Before the beginning of this project, we developed the XARK compiler framework for automatic detection of parallelism in irregular codes. The framework hinges on a generic and extensible algorithm that recognizes program constructs (from now on, computational kernels) that are frequently used by software developers.

   (a) Extension of the collection of kernels recognized by XARK and its applicability to C, C++ or Java. We have extended the recognition algorithms of the XARK compiler so that, at this moment, more than 77 per cent of the loops that appear in some well-known benchmarks are recognized successfully. In particular, we have analyzed the Sparskit-II library for the manipulation of sparse matrices, the Perfect benchmarks, the collection SPEC FP CPU2000 and the package PLTMG for the solution of elliptic partial differential equations [11]. In order to port XARK to GCC, we have studied the characteristics of the intermediate program representations of the GNU GCC compiler, namely, GIMPLE and RTL. We have selected GIMPLE because it eases the implementation of dependence analysis techniques and parallelizing code transformations. At this moment, we are finishing the identification of the information requirements of the XARK compiler and we are giving the first steps in the implementation of the recognition algorithms.

   (b) Improvement of current optimizing/parallelizing compilers using the XARK compiler. This part of the project is still in a preliminary stage and under study. We estimate that this last year of the project will be enough to obtain some encouraging results. We are working on the development of an efficient strategy that executes dependence tests on-demand using the information provided by the XARK compiler.

5. Compile–time and run–time solutions for irregular codes (Subprojects 1 and 2)

(a) Improvement of the IARD representation. The IARD representation was improved for irregular reduction operations [26] that are the core in many N-body problems. Currently we are applying this technique to the N-body problem.

(b) Usage of the IARD representation and the analytical modeling based on probabilistic miss equations to increase locality in irregular accesses (Subprojects 1 and 2). The PME (Probabilistic Miss Equations) model developed by our groups was extended in a first stage to cover codes with irregular access patterns due to the existence of memory accesses guarded by conditional statements [7]. In a second stage, we extended the model to cover irregular access patterns due to the existence of indirections [5, 6, 8, 9, 10]. This latter extension was automated and integrated in the XARK compiler. Our current implementation covers most typical indirections, and we are now working on the characterization of the indirections in complex codes such as those with loops depending on other loops. Two practical successful applications of our model to increase locality in the codes have been studied in the context of this project too. The first one consisted in choosing the optimal tile sizes for regular codes guided by the predictions of the model [27]. The second one was choosing optimal loop nesting orders for codes with irregular accesses. Both optimizations yielded large speedups in different real architectures. Additionally, a new technique to reduce the number of false sharing occurrences in irregular codes was presented [38].

6. Iterative framework for the development of parallel applications (Subproject 2).

(a) Development of a MATLAB toolbox for an environment based on HTAs [12]. A MATLAB toolbox was written that enabled the programming of parallel applications using the Hierarchical Tiled Array (HTA) class. The toolbox allowed the interactive development of applications, as the user could type commands in a client that would broadcast them to a collection of servers where they would be executed in parallel. Comparisons were performed with other approaches to write parallel applications, such as the traditional message-passing paradigm using well-known libraries such as MPI. We found that MATLAB-based HTA programs were much shorter in number of lines of code and could be developed much faster than those written traditional approaches such as FORTRAN + MPI. HTA programs were also easier to understand and maintain, as they offer a global view of the distributed data and a single-threaded view of the execution. On the other hand, while the HTA programs had a reasonable speedup with respect to their sequential MATLAB version, they were not competitive performancewise with the traditional parallel programs. The broadcast of each command to execute, as well as the nature of MATLAB, an interpreted language, and the limitations imposed by its interface are the most important reasons for this difference in performance.

(b) Analysis and implementation of optimizations for the HTA compiler, and migration to an open source environment [13]. The MATLAB HTA was improved in a number of ways, allowing for example to run in server-only mode, so that there would be no need to broadcast the commands. Instead each server would run the whole application in SPMD in the background, while the vision of the execution by the user would continue to be single-threaded. Still, performance would not be competitive

with that of SPMD message-passing programs in traditional languages, and since one of the objectives of our project was to develop an HTA environment in a completely open source environment, a C++ HTA library was written. We are still making optimization in this latter library, but our current results show already performance comparable with that of traditional approaches, while we continue to get advantages in terms of numbers of line of code and readability of the applications.

7. Application specific hardware for computer graphics (Subprojects 1 and 2)

   (a) Elimination of redundant computations in NURBS tessellation. Bezier and NURBS surfaces are widely used in computer graphics for CAD/CAM, virtual reality and other applications. But efficient rendering of these surfaces involves their tessellation before to be sent to the GPU. We have proposed a method for adaptive tessellation of Bezier surfaces and an architecture to implement it in the GPU [23, 24, 25], that eliminate the redundant computations, by using a test map that avoid the recomputation of vertexes during subdivision. As a result, good quality meshes are generated with less computations.

   (b) Algorithms and architectures for point rendering. We have developed an architecture for point rendering that, by using a novel data distribution, reduce the number of bits needed to represent the scene. This result is included in the PhD. Thesis of Paula Novío Mallón.

   (c) Efficient and fast coding of tetrahedral meshes. We have adapted an algorithm, previously developed in our group, for the triangle meshes coding to the coding of the tetrahedral meshes. This algorithm tries to reduce the information required to code the mesh connectivity by avoiding to send the information of the same vertex several times. Although this algorithm was very efficient for triangle meshes, unfortunately the results for tetrahedral meshes has not been as good as expected, because of the complexity of the possible tetrahedral distribution in the real meshes. This result is included in the PhD. Thesis of Paula Novío Mallón. On the other hand and as result of following the research line of extending the graphics card capabilities, we have developed other extensions to GPU that were not foreseeable at the beginning of the project [4, 14, 15]

8. Video compression processors (Subproject 1)

   (a) Architecture of a basic video compression/decompression system and optimization of functional units. Our group has been working intensively on the design an architecture for the H.264 video compression standard. Our efforts have been focused on to obtain an efficient implementation for the arithmetic coder (CABAC) [33, 35], the DCT computation [18] and the motion estimation [34]. Moreover, we have proposed an implementation of CABAC on a VLIW DSP [20]. Actually, we are working to obtain a complete implementation of the H.264 on a high performance VLIW DSP and on the implementation on FPGAs.

9. Functional unit design (Subproject 1)

   (a) Architecture for the low-latency computation of the floating–point MAF operation. Our group has been working previously on the implementation of floating–point

MAF operation, $A + B \times C$. MAF allows to compute with the same unit the addition $A+B$ and the multiplication, $B \times C$. Following with this research, we have proposed a MAF architecture with a reduced latency for the addition operation with any additional cost [17], improving in this way the computation of the most usual operation, the addition.

(b) Improvement of the vector normalization. The normalization of a vector involves the computation of a square root and of a division or reciprocal. We focused in the improvement of these operation. As a result, we have proposed an table–based algorithm and architecture for the computation of the powering function that, obviously, includes the square root operation [41]. This architecture improves the performance of previous implementations. As another different alternative, proposed before by our group, uses a digit–recurrence algorithm suitable for low complexity architectures, although the latency is much larger than in the table–based approach. This algorithm has been improved, by improving the computation of the logarithm function [42]. Additionally, some efforts have been devoted to improve the division operation [30, 43, 44]. The result is an efficient implementation of the normalization operation. However, we are still working on it, mainly in the simplification of the rounding step of the reciprocal, division, square root and inverse square root operations. This will affect to the global performance of the normalization.

10. High performance applications in computer graphics, multimedia processing and engineering simulation (air pollution, microelectronic devices) (Subprojects 1 and 2). We have developed an efficient 3D parallel simulator to study a Si metal-oxide-semiconductor field effect transistor (MOSFET) [2, 3, 28] and high electron mobility transistors (HEMTs) [29, 53, 54, 55, 56, 57, 58, 59, 60]. The simulator is employed to study the impact of the intrinsic parameter fluctuations within a high-k dielectric and polysilicon grains on the threshold voltage and drive current in MOSFETs, and material variations in HEMTs on the on drain current and small signal parameters.

    We have developed different proposals for global illumination based on radiosity methods [36, 37, 51, 52]. We have worked with the main radiosity algorithms: progressive and hierarchical (we are currently working also with the Montecarlo algorithm). We have proposed different strategies for the parallelization of the algorithms in distributed memory systems: efficient methods for the distribution, scheduling of the algorithms and improved visibility determination.

    Image processing was also used for some of our developments. In [39] we show an efficient new strategy to deal with the segmentation problem in parallel systems. Due to the success of this work we were invited to write a chapter in a specialized book [40]

    With regard to the air quality simulation, during last year the parallel implementations of the code has been improved in order to enhance the global performance. Traditional data distribution strategies are not adequate for heterogeneous environments, as the Grid. New dynamic data distributions are proposed and tested in order to improve load balance of this specific application [31]. Finally, the running times of large scale computational science and engineering parallel applications, executed on clusters or Grid platforms, are usually longer than the mean time between failures (MTBF). Hardware failures must be tolerated by the parallel applications to ensure that no all computation done is lost

on machine failures. Two different approaches to endow with fault tolerance the MPI version of the air quality simulation has been developed [32]. A segment level solution has been implemented by means of the extension of a checkpointing library for sequential codes. A variable level solution has been implemented manually in the code. The main differences between both approaches are portability, transparency level and checkpointing overheads. Some additional results are presented in [1, 16, 46].

4 Result indicators

The **publications** derived from the project are listed in the references section. Note that among these references there are a number of publications in high quality journals and symposiums. Moreover five PhD. Thesis have been presented during the project period related with the objectives of the project: Juan Carlos Pichel (2006), Paula Novío (2006), Natalia Seoane (2007), Emilio J. Padrón (2006) and José Carlos Mouriño (2006)

On the other hand, in USC (University of Santiago de Compostela) group there are several **granted researchers and personnel under contract**, currently working on their PhD. in some topics related with the objectives of the project. More specifically, there are three FPU and FPI grants, Manuel Aldegunde (parallel electronic devices simulation), Xulio López Albín (Grid computing) Diego Rodríguez (Grid computing), one *Xunta de Galicia* predoctoral grant, Juan Angel Lorenzo (memory hierarchy) and two personnel under contract, Daniel Piso (functional unit design) and Natalia Seoane (parallel electronic devices simulation) which recently has presented her PhD. Other former PhD students have joined to the Carlos III University, Madrid (J.C. Pichel) and Imagination Technologies, London (Paula Novío)

In the UDC (University of A Coruña) group there are 4 young researchers with a contract, 3 through public research projects and one through a research project with an enterprise. Besides, in a competitive call for research groups of the University of A Coruña to provide teaching assistant positions following criteria of research excellence and potential of education of researchers, the GAC achieved two teaching assistant positions in the (only) calls of 2005 and 2006 (being the only group that got two positions). Three former PhD students (Emilio Padrón, Diego Andrade, and Guillermo L. Taboada) have recently joined the UDC faculty and another one moved to CESGA (Mouriño), so that there is enough availability to host new PhD students through predoctoral FPI grants.

The UDC group participate in the organization of two HPC international conferences: Workshop on Compilers for Parallel Computers (CPC2006, A Coruña, January 2006), and Iberian Grid Infrastructure Conference (IBERGRID2007, Santiago de Compostela, May 2007). Currently, members of the group are Guest Editors of the Journal on Concurrency and Computation: Practice and Experience (Special Issue on Compilers for Parallel Computers) and IEEE Transactions on Education (Special Issue on Grid-based Technologies applied to Education).

Recently, in October 2006 both groups, USC and UDC, involved in this project got grants from the Galician Government (Programa de consolidación y estructuración de unidades de investigación competitivas) to reinforce high-quality Galician research Groups.

In May 2006 the IP of the UDC group was awarded with an I3 grant (the only researcher at UDC that got this grant) in the *Programa de Promoción de la Intensificación de la Actividad Investigadora* from the Galician Government in recognition to his research excellence.

Both groups have strong collaboration with other spanish and international research groups

at universities and companies whose research interests have relationship with some of the topics of the project: University of Illinois at Urbana-Champaign (USA), University of Edinburgh (UK), University of Tubingen (Germany), University of Portsmouth (UK), University of Reading (UK), IBM T.J. Watson (USA), University of California at Irvine (USA), University of Delft (Holland), University of Málaga (Spain), etc.

Besides, the UDC group participates in an project (CYTEDGRID *Tecnología Grid como motor del desarrollo regional*) which has provided to the UDC group new contacts with spanish research groups from the UCM, UAB and UPV, and with other research groups from Portugal and 12 Latin American countries. The USC groups participates in the European Network of Excellence on High-Performance Embedded Architecture and Compilation (HiPEAC). Additionally, both groups participate in the request for funding of a Consolider project (User-Oriented environment for high performance computing application development and management, IP: Emilio Luque, UAB), two Galician research netwoks on High Performance Computing and GIS and currently we are discussing with HP and the CESGA (Centro de Supercomputation de Galicia) how to tackle a new common project, about memory hierarchy organization for supercomputers.

References

[1] J.L. Albín, D.R. Martínez, M. Boullón, J.C. Cabaleiro, T.F. Pena and F.F. Rivera. *Predicción del rendimiento de una aplicación Grid para el control de inundaciones*. XVI Jornadas de Paralelismo. pp. 363-370. Granada. 2005.

[2] M.Aldegunde, J.J.Pombo and A.J.García Loureiro. *Modified octree mesh generation for Manhattan type structures with narrow layers applied to semiconductor devices*. International Journal of Numerical Modelling: Electronic Networks, Devices and Fields. Vol.16, Issue 6, pp. 473-489 (2006)

[3] M.Aldegunde, A.J.García Loureiro, K.Kalna and A.Asenov. *A 3D finite Element Parallel Simulator for Studying Fluctuations in Advanced MOSFETs*. 11th International Workshop on Computational Electronics. Viena (2006)

[4] M. Amor, M. Bóo, E. J. Padrón and D. Bartz. *Hardware Oriented Algorithms for Rendering Order-Independent Transparency*. Computer Journal. 49(2):201-210, 2006.

[5] D. Andrade, B.B. Fraguela, R. Doallo. Modelado Analítico Automático del Comportamiento de la Cache para códigos con indirecciones. Actas XVI Jornadas de Paralelismo, pp. 321-328, September 2005.

[6] D. Andrade, M. Arenaz, B.B. Fraguela, J. Touriño, R. Doallo. *Automated and Accurate Cache Behavior Analysis for Codes with Irregular Access Patterns*. 12th Workshop on Compilers for Parallel Computers, CPC 2006, pp. 179-193, A Coruña (Spain), 9-11 January 2006.

[7] D. Andrade, B.B. Fraguela, R. Doallo. *Analytical Modeling of Codes with Arbitrary Data-Dependent Conditional Structures*. J. of Systems Architecture, 52(7), pp. 394-410. July 2006. (ISSN: 1383-7621).

[8] D. Andrade, B.B. Fraguela, R. Doallo. *Cache Behavior Modelling for Codes Involving Banded Matrices*. Proc. 19th Intl. Workshop on Languages and Compilers for Parallel Computing (LCPC'06). To be published as a volume of Lecture Notes in Computer Science. New Orleans, November 2006.

[9] D. Andrade, M. Arenaz, B.B. Fraguela, J. Touriño, R. Doallo. *Automated and accurate cache behavior analysis for codes with irregular access patterns*. To be published in Concurrency and Computation: Practice and Experience. (ISSN: 1532-0626).

[10] D. Andrade, B. B. Fraguela, R. Doallo. *Precise Automatable Analytical Modeling of the Cache Behavior of Codes with Indirections*. To be published in ACM's Transactions on Architecture and Code Optimization (TACO). (ISSN: 1544-3566).

[11] M Arenaz, J. Touriñ and R. Doallo. *XARK: An E***X***tensible Framework for* **A***utomatic* **R***ecognition of Computational* **K***ernels*. ACM Transactions on Programming Languages and Systems (TOPLAS) journal (submitted).

[12] G. Bikshandi, J. Guo, D. Hoeflinger, G. Alm'asi, B.B. Fraguela, M.J. Garzarán, D. Padua, C. von Praun. *Programming for Parallelism and Locality with Hierarchically Tiled Arrays*. ACM SIGPLAN 2006 Symposium on Principles and Practice of Parallel Programming (PPoPP'06), pp. 48-57. New York (New York, EEUU). March 2006.

[13] G. Bikshandi, J. Guo, C. von Praun, G. Tanase, B.B. Fraguela, M.J. Garzarán, D. Padua, L. Rauchwerger. *Design and Use of htalib – a Library for Hierarchically Tiled Arrays*. Proc. 19th International Workshop on Languages and Compilers for Parallel Computing (LCPC'06). To be published as a volume of Lecture Notes in Computer Science. New Orleans, November 2006.

[14] M. Bó and M. Amor. *High-Performance Architecture for Anisotropic Filtering*. Journal of Systems Architecture. 51(5):297-314, 2005.

[15] M. Bóo, M. Amor and J. Döllner. *Unified Hybrid Terrain Representation Based on Local Convexifications*. Geoinformatica

[16] M. Boullón, J.C. Cabaleiro, R. Doallo P.González, D.R.Martínez, M.Martín, J.C.Mouriño, T.F.Pena and F.F. Rivera. *Modeling execution time of selected computation and communication kernels on Grid*. Lecture notes in computer science (EGC2005). Vol. 3470. pp. 731-740. 2005.

[17] J.D. Bruguera and T. Lang. *Floating–Point Multiply–Add Fused: Latency Reduction for Floating–Point Addition*. Proc. IEEE 17^{th} Symp. Computer Arithmetic (ARITH17). Cape Cod, USA. pp. 42–51. (2005).

[18] J.D. Bruguera and R.R Osorio. *An Unified Architecture for H.264 Multiple Block–Size DCT with Fast and Low Cost Quantization*. Proc. 9th Euromicro Conference on Digital System Design (DSD'2006). Dubrovnik, Croatia. pp. 407–414. (2006).

[19] CALL. A Complexity Analysis Tool. http://nereida.deioc.ull.es/ call.

[20] M.E. Castro, R.R. Osorio and J.D. Bruguera. *Optimizing CABAC for VLIW architectures*. Proc. XXI Conference on Design of Circuits and Integrated Systems (DCIS2006). Barcelona, Spain. (2006)

[21] I. Díaz, J. Touriño, J. Salceda, R. Doallo. *A Framework Focus on Configuration Modeling and Integration with Transparent Persistence*. 1st Workshop on System Management Tools for Large-Scale Parallel Systems (in conjunction with the 19th Annual IEEE International Parallel and Distributed Processing Symposium, IPDPS'2005). IEEE Computer Society, pp. 297 (CD-ROM, 8 pages). Denver, Colorado (USA), April 2005.

[22] I. Díaz, J. Touriño, R. Doallo. *AdCIM: a Model-Driven Framework for Rapid Development of System Administration Applications*. Submitted to the Journal of Systems and Software, 2006.

[23] F.J. Espino, M. Bóo, M. Amor and J.D. Bruguera. *Adaptive Tessellation of Bezier Surfaces Based On Displacement Maps*. Proc. 13-th Int. Conf. in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG'2005). Plzen, Czech Republic. pp. 29–32. (2005).

[24] F.J. Espino, M. Bóo, M. Amor and J.D. Bruguera. *Hardware Support for Adaptive Tessellation of Béier Surfaces with Combined Tests*. Proc. XX Conference on Design of Circuits and Integrated Systems (DCIS2005). Lisboa, Portugal. (2005)

[25] J. Espino, M. Bóo and J.D. Bruguera. *Hardware Support for Adaptive Tessellation of Bezier Surfaces Based on Local Tests*. Journal of Systems Architecture. To appear.

[26] D.E. Singh, M.J. Martín and F.F. Rivera. *Runtime characterization of irregular accesses applied to parallelisation of irregular reductions*. Int. Journal of Computational Science and Engineering. Vol. 1 n. 1. pp. 1?14. 2005.

[27] B.B. Fraguela, M.G. Carmueja, D. Andrade. *Optimal Tile Size Selection Guided by Analytical Models*. Proc. Parallel Computing 2005 (ParCo 2005). pp. 565-572. Málaga, September 2005. Published as volume 33 of the Publication Series of the John von Neumann Institute for Computing (NIC), ISBN 3-00-017352-8.

[28] A.J.García Loureiro, K. Kalna y A.Asenov. *Intrinsic fluctuations induced by a high-k gate dielectric in sub-100 nm Si MOSFET*. International Conf. on Noise and Fluctuations. Salamanca, Septiembre (2005)

[29] A.J.García Loureiro, K.Kalna and A.Asenov. *Efficient three-dimensional parallel simulations of PHEMTs*. International Journal of Numerical Modelling: Electronics networks, devices and field. Vol. 18, pp. 327-340 (2005)

[30] V. Holimath and J. D. Bruguera. *A Linear Convergent Functional Iterative Division Without Look-Up Table*. Proc. 9th Euromicro Conference on Digital System Design (DSD'2006). Dubrovnik, Croatia. pp. 236–239. (2006).

[31] J.C. Mouriño, M.J. Martín, P. González, R. Doallo. *Dynamic Load Balancing for the STEM-II Air Quality Model*. Int. Conf. On Computacional Science and its Applications, ICCSA2006, LNCS Vol. 3980, pp. 701-710, 2006.

[32] J.C. Mouriño, M.J. Martín, P.González, R. Doallo. *Fault-tolerant Solutions for a MPI Compute Intensive Application*. 15th Euromicro Conference on Parallel, Distributed and Network-based Processing, PDP 2007.

[33] R.R. Osorio and J.D. Bruguera. *A New Architecture for Fast Arithmetic Coding in H.264 Advanced Video Coder*. Proc. 8th Euromicro Conference on Digital System Design (DSD'2005). Porto, Portugal. pp. 298–305. (2005)

[34] R.R.Osorio and J.D. Bruguera. *A Combined memory Compression and Hierarchical Motion Estimation Architecture for Video Encoding in Embedded Systems*. Proc. 9th Euromicro Conference on Digital System Design (DSD'2006). Dubrovnik, Croatia. pp. 269–272. (2006).

[35] R.R. Osorio and J.D. Bruguera. *High-Throughput Architecture for H.264/AVC CABAC Compression System*. IEEE Transaction on Circuits and Sytems for Video Technology. Vol. 16. No.11. pp. 1376-1384. (2006).

[36] E.J. Padrón, M. Amor, R. Doallo. *Improving the Performance of Visibility Determination in Global Illumination Methods*. Proc. of CISST'05, CSREA Press, H.R. Arabnia (Ed.), pp. 265-270, 2005.

[37] E. J. Padrón, M. Amor, M. Bóo and R. Doallo. *A Hierarchical Radiosity Method with Scene Distribution. 15-th Euromicro Conference on Parallel, Distributed and Network Based Processing, PDP 2007*, 2007. (in press)

[38] J.C.Pichel, D.B.Heras, J.C.Cabaleiro and F.F.Rivera. *A New Technique to Reduce False Sharing in Irregular Codes Based on Distance Functions*. 8th International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN). pp. 306-311. Las Vegas (USA). 2005.

[39] J.C. Pichel, D.E. Singh and F.F. Rivera. *Image segmentation based on merging of sub-optimal segmentations*. Pattern recognition letters. Vol. 27 n. 10. pp. 1105-1116. 2006.

[40] J.C. Pichel, D.E. Singh and F.F. Rivera. *Image segmentation*. Vision Systems. Edited by Goro Obinata and Ashish Dutta. Advanced Robotic Systems Eds. (2007)

[41] J.A. Piñeiro, S. Oberman, J.–M Muller and J.D. Bruguera. *High-Speed Function Approximation using a Minimax Quadratic Interpolator*. IEEE Transactions on Computers. Vol. 54, No. 3, pp. 304–318. (2005).

[42] J.A. Piñeiro, M.D. Ercegovac and J.D. Bruguera. *High–Radix Logarithm with Selection by Rounding: Algorithm and Implementation*. Journal of VLSI Signal Processing. Vol.40. No. 1. pp. 109-123. (2005).

[43] D. Piso and J.D. Bruguera. *Error Analysis of Goldschmidt Algorithm*. Proc. Advanced Computer Architecture and Compilation for Embedded Systems (ACACES 2005). L'Aquila, Italy. pp. 295–298. (2005).

[44] D. Piso and J.D. Bruguera. *Optimizing the Multiplier Desing for Goldschmidts Division and Reciprocal Units*. Proc. XXI Conference on Design of Circuits and Integrated Systems (DCIS2006). Barcelona, Spain. (2006).

[45] D.R.Martínez, J.L.Albín, J.C.Cabaleiro, T.F.Pena, F.F.Rivera. *A Load Balance Methodology for Highly Compute-Intensive Applications on Grids Based on Computational Modeling*. Lecture notes in computer science. GADA3005. Vol. 3762. pp. 327-336. 2005.

[46] D.R. Martínez, V. Blanco, J.C. Cabaleiro, T.F. Pena and F.F. Rivera. *Modelización del rendimiento de aplicaciones MPI en entornos Grid*. XVII Jornadas de Paralelismo. pp. 301-306. Albacete. 2006

[47] D.R. Martínez, V. Blanco, M. Boullón, J.C. Cabaleiro, C. Rodríguez and F.F. Rivera. *Software Tools for Performance Modelling of Parallel Programs*. PMEO-PDS'.2007. Long Beach (USA) (Accepted).

[48] G. Rodríguez, M.J. Martín, P. González, J. Touriño, R. Doallo. *On Designing Portable Checkpointing Tools for Large-Scale Parallel Applications*. Proc. of ICCSE'2005, pp.191-202, 2005.

[49] G. Rodríguez, M.J. Martín, P. González, J. Touriño, R. Doallo. P*ortable Checkpointing of MPI Applications*. 12th Workshop on Compilers for Parallel Computers, CPC'2006, pp. 396-410, A Coruña (Spain), 9-11. 2006.

[50] G. Rodríguez, M.J. Martín, P. González, J. Touriño *Controller/Precompiler for Portable Checkpointing*. IEICE Transactions on Information and Systems, Special Issue on Parallel/Distributed Computing and Networking, Vol. E89-D, No. 2, pp. 408-417. 2006.

[51] J. R. Sanjurjo, M. Amor, M. Bóo and R. Doallo. *Parallel Global Illumination Method Based on a Non-Uniform Partitioning of the Scene*. 13-th Euromicro Conference on Parallel, Distributed and Network Based Processing, PDP 2005, Lugano (Suiza), 9-11 Febrero, pag. 251-257, 2005.

[52] J. R. Sanjurjo, M. Amor, M. Bóo and R. Doallo. *Improving Locality for Progressive Radiosity Algorithm: A Study based on the Blocking Transformation of the Scene. the 15–th International Conference in Central Europe on Computer Graphics and Visualization'2007 (WSCG'2007)*

[53] N.Seoane, A.J.García Loureiro, K. Kalna y A.Asenov. *Indium content fluctuations in the channel of a 120 nm PHEMT*. Proceedings of 7th International Conference on New Phenomena in Mesoscopic Sytems/5th International Conference on Surfaces and Interfaces in Mesoscopic Devices. pp. 135-136. Maui (Hawai) (2005)

[54] N.Seoane, A.J.García Loureiro, K. Kalna y A.Asenov. *A high-performance parallel device simulator for high electron mobility transistors*. Parallel Computing (ParCo2005). Malaga, Septiembre (2005)

[55] N.Seoane, A.J.García Loureiro, k.Kalna and A.Asenov. *Optimisation of the parallel performance of a 3D device simulator for HEMTs*. LNCS. Vol. 4330, pp. 859-868 (2006)

[56] N.Seoane, A.J.García Loureiro, k.Kalna and A.Asenov. *Atomistic effect of delta doping layer in a 50 nm Inp HEMT*. Journal of Computational Electronics. Vol. 5, pp. 131-135 (2006)

[57] N.Seoane, A.J.García Loureiro, k.Kalna and A.Asenov. *Current Variations in pHEMTs introduced by channel composition fluctuations*. Journal of Physics: Conference Series. Vol. 38, pp. 212-215 (2006)

[58] N.Seoane and A.J.García Loureiro. *Study of parallel numerical methods for semiconductor device simulation*. International Journal of Numerical Modelling: Electronic Networks, Devices and Fields. Vol. 19, Issue 1, pp. 15-32 (2006)

[59] N.Seoane, A.J.García Loureiro, K.Kalna and A.Asenov. *A 3D parallel simulation of the effect of interface charge fluctuations in HEMTs*. 11th International Workshop on Computational Electronics. pp. 81-82. Viena (2006)

[60] N.Seoane, A.J.García Loureiro, K.Kalna and A.Asenov. *Impact of intrinsic parameter fluctuations on the performance of HEMTs studied with a 3D parallel drift-difussion simulator*, Solid-State Electronics (aceptado)

[61] G.L. Taboada, J. Touriño, R. Doallo. *Designing Efficient Java Communications on Clusters*. 7th International Workshop on Java for Parallel and Distributed Computing (in conjunction with the 19th Annual IEEE International Parallel and Distributed Processing Symposium, IPDPS'2005). IEEE Computer Society, pp. 182 (CD-ROM, 7 pages). Denver, Colorado (USA), April 2005.

[62] G.L. Taboada, J. Touriño, R. Doallo. *Non-blocking Java Communications Support on Clusters*. 13th European PVM/MPI Users' Group Meeting, EuroPVM/MPI'06, Bonn (Alemania). In Recent Advances in Parallel Virtual Machine and Message Passing Interface. Lecture Notes in Computer Science no. 4192, pp.256-265, Springer-Verlag, September 2006.

[63] G.L. Taboada, J. Touriño, R. Doallo. *Efficient Java Communication Protocols on High-Speed Cluster Interconnects*. 31st IEEE Conference on Local Computer Networks, LCN 2006, pp. 264-271. Tampa, Florida (USA), November 2006.

[64] G.L. Taboada , J. Touriño, R. Doallo. *Performance Modeling of Message-Passing Libraries on System Area Network Clusters*. Conferencia Latinoamericana de Informática, CLEI2006. (CD proceedings), 2006.

[65] G.L. Taboada, J. Touriño, R. Doallo. *Bibliotecas de Comunicación Eficiente en Clusters para Códigos Java*. Actas XVII Jornadas de Paralelismo, pp.407-412, 2006.

[66] G.L. Taboada, J. Touriño, R. Doallo. *High Performance Java Sockets for Parallel Computing on Clusters*. 9th International Workshop on Java and Components for Parallelism, Distribution and Concurrency (in conjunction with the 21st IEEE International Parallel and Distributed Processing Symposium, IPDPS'2007). Long Beach, CA (USA), March 2007 (accepted for publication).

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Razonamiento automático y optimización: Lenguajes, Modelos, Algoritmos, Aplicaciones TIN2004-07854

IP: Álvaro del Val Latorre
Universidad Autónoma de Madrid

Abstract

El proyecto se encuadra dentro del paradigma general de la inteligencia artificial basada en representaciones declarativas del conocimiento (modelos) expresadas en algún lenguaje formal de representación (lógicas), que son manipuladas mediante algoritmos genéricos de razonamiento y/o búsqueda (razonamiento automático). La experiencia reciente sugiere que todos los aspectos de esta ecuación están altamanente imbricados y que no pueden ser estudiados por separado. Por eso proponemos un enfoque múltiple centrado en los siguientes aspectos: Lenguajes: proposicional (SAT), satisfacción de restricciones (CSPs), lógicas multivaluadas, lenguajes híbridos que incorporen las restricciones de la programación lineal. Comparación de los lenguajes desde el punto de vista de modelización e identificación de clases tratables. Modelos: desarrollo de modelos para dominios concretas (quasigrupos, planificación, optimización) para explorar de forma sistemática el impacto del modelo sobre la eficiencia de resolución de problemas (ej. modelos redundantes, modelos híbridos, modelos de permutación, eficiencia o redundancia de la propagación según el modelo). Algoritmos: Desarrollo de librerías y algoritmos para solución de problemas SAT y CSP, con especial atención a la eficiencia y a la escalabilidad. Algoritmos sistemáticos y estocásticos para la optimización. Se pretende desarrollar estos objetivos en lo posible en conexión con aplicaciones, como se detalla en la memoria del pryecto.

Keywords: RAZONAMIENTO AUTOMATICO, RAZONAMIENTO PROPOSICIONAL, SAT, SATISFACCION DE RESTRICCIONES, CSP, PROBLEMAS DE OPTIMIZACION, AUTOMATED REASONING, PROPOSITIONAL REASONING, SAT, CONSTRAINT SATISFACTION, CSP, OPTIMIZATION PROBLEMS

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Semantic characterization, indexing and universal customized access to audiovisual content under controlled environments TIN2004-07860-C02

Jesús Bescós[*]
Universidad Autónoma de Madrid

Fernando Jaureguizar[**]
Universidad Politécnica de Madrid

Abstract

The main objective of this project is to research in the design of algorithms and the development of tools both oriented to automatic feature extraction in image sequences of controlled environments, additionally the further indexing, storage and customized access to the previously extracted and properly coded information out of those sequences. The project consists of to main stages: Information acquisition and Management of that information to offer services to the final user. The information acquisition out of images from several cameras performs the automatic extraction of features of people present in those images, segmenting in real time the individual persons out of the background, while building a 3D model of the scene thanks to self-calibration techniques. Finally, this information is described according to the MPEG-7 standard to allow for efficient access and storage. The second stage is mainly concerned with the management of the obtained information for its use in the development of applications. Its design is based on the user and terminal/network profiles (MPEG-21 DIA) and on the management of alert levels/options for the delivery of information relevant to each user. According to these profiles, the available coders with the proper parameters are run for the purpose of adaptation and customization of the information during each session.

Keywords: Autocalibration, segmentation, video analysis, compressed domain analysis, content adaptation, 3D modeling, natural coding, synthetic and hybrid coding,

1 Project objectives

1.1 Introduction

This project proposes the joint development of series of research fields, independent but complementary, intended to promote technology for automatic indexing and personalized delivery and retrieval of audiovisual content.

In order to establish a ser of services which enable querying, search, retrieval, adapted transmission, distribution, personalization, etc., in an information system, with the aim of

[*] Email: j.bescos@uam.es
[**] Email: fjn@gti.ssr.upm.es

generating added value and maximize its fields of use, the availability the precise description of the available information, which is commonly known as meta-information or meta-data, play a fundamental role.

There are plenty of initiatives intended to automatically perform automatic description of audiovisual content. However, the high complexity of succeeding in the general case, motivates that current research is focused on specific situation or controlled application domains (e.g., news, soccer, etc.). This project targets a specific domain (static cameras, almost controlled illumination, predefined relevant objects, etc.) which is bounded enough to achieve successful results with practical application, and generic enough to export its achievements to a wide range of domains.

The particular research fields that have been considered to reach a common objective include autocalibration techniques, detailed analysis of image sequences, 3D modeling of such sequences, description and indexing of the content (MPEG7), personalized and adapted access to such information according to network, terminal and user profiles (MPEG21-DIA), and final coding and transmission (MPEG4, H.264).

1.2 Specific objectives

This subsection provides a detailed description of the list of planned objectives:

- Enhancement of present autocalibration techniques, mainly focusing on their robustness and precision. Camera calibration is the base for 3D scene modeling and for splitting the background and the mobile objects, but allowing a later integration of these on the background via description of their geometric relationships.
- Integration of segmentation information into the coding strategy: discriminating between background and relevant objects, scalable coding analysis in several coding schemes (hybrid, wavelets) and integration of developed strategies in video coding standards (JPEG-2000, MPEG-4, H.264).
- Efficient and hierarchical coding of synthetic 3D scenes (previously modeled as a static background and several mobile objects), and integration of hybrid synthetic and natural contents.
- Development of algorithms for real-time foreground/background segmentation, based on compressed domain analysis, as a first step for a later refinement of the segmented regions.
- Real-time feature extraction from sequences captured from the cameras in order to obtain the basic information about the sequence contents. These are the starting point for more advanced feature extraction algorithms.
- Description and indexing of the segmented elements vide meta-data coded according to the MPEG-7 standard.
- Development of real-time transmission control algorithms for maximizing the quality of the decoded information in the user terminal.
- Information management, personalization and adaptation: user, network and terminal profiles management, transcoding and transmoding management, translation from MPEG7 descriptions, coding management of the adapted media, handling of alert criteria, and content synthesis from content descriptions.

1.3 Chronogram and resources

The Project is structured into five work packages: the first focused on establishing the working framework, three more specifically targeting each of the general fields of research considered, and a last one aimed to show practical applications of the achievements. The following Table summarizes the list of activities, with an indication of their current status.

Activities/Tasks	Start	End	Partner	Status
P.1: Framework development				
T.1.1: Environment setup	m01	m06	ALL	Ongoing
T.1.2: Information model	m04	m06	ALL	Finished
P.2: Advanced image sequences processing				
T.2.1: Calibration and capturing system setup	m06	m18	UPM	Finished
T.2.2: Real-time segmentation in the transform domain	m08	m20	UAM	Finished
T.2.3: Segmented objects analysis in the transform domain	m20	m32	UAM	Ongoing
T.2.4: Segmentation in the pixel domain	m08	m20	UPM	Finished
T.2.5: Segmented objects analysis in the pixel domain	m08	m22	UPM	Finished
T.2.6: 3D scene modeling and management	m09	m27	UPM	Ongoing
P.3: Semantic indexing, storage, retrieval and personalization				
T.3.1: Segmented objects description (MPEG-7 + contents)	m06	m18	UAM	Ongoing
T.3.2: Efficient segmented objects storing	m18	m24	UAM	Ongoing
T.3.3: Information management, personalization and adaptation	m06	m30	UAM	Ongoing
P.4: Coding and transmission control for network and client adaptation				
T.4.1: Natural coding	m09	m30	UPM	Ongoing
T.4.2: Synthetic and hybrid coding	m09	m30	UPM	Ongoing
T.4.3: Transmission control	m10	m33	UPM	Ongoing
P.5: Application to remote monitoring services				
T.5.1: Design	m09	m27	ALL	Ongoing
T.5.1: Integration and trials	m24	m36	ALL	

The funding of the project has made available to the participant entities the necessary resources to conduct the research, mainly computers and video cameras have been installed in the participant entities, allowing for the developments of the members of the research team as well as of the students working on their master thesis and PhD Thesis integrated in the project.

2 Project achievements

2.1 Environment setup

At the UPM, initial system design of cameras in public indoor and outdoor places was not possible because the request for their installation was rejected due to legal issues about privacy. For the moment, a set of PTZ cameras (IP connected) has been installed in several parts of the ceiling of a students Laboratory (indoor semi controlled environment) in order to carry out the sequences acquisition for the project developments.

At the UAM the installation of the acquisition systems has been scheduled in three phases, corresponding to the three years of the project, in order to benefit from technology advances in the area. The first year installation comprises three high resolution (1024x768 YUV 4:2:2 at 15 fps) color cameras located in two access entrances to the EPS Building; they provide raw video which is delivered via optic fiber to our laboratory. The second year installation comprises three PTZ IP cameras, all located in the EPS Hall.

A main part of the environment consists of a framework for distributed real-time processing of video information, mainly designed to facilitate the testing of interrelated algorithms.

2.2 Advanced image sequences processing

2.2.1 Calibration and capturing system setup

The research conducted in autocalibration has been concentrated in two areas: the Absolute Line Quadric based algorithms and two cameras based algorithms.

In the first area the concept of Absolute Line Quadric has been introduced. This is the geometrical object that represents the set of lines intersecting the absolute conic, and allows for the obtainment of new autocalibration algorithms, in the same way as it was formerly done with the Absolute Quadric. The new algorithms [1][2][4][5] provide the direct solution of the autocalibration problem for ten or more images taken with square-pixel cameras with arbitrarily varying internal and external parameters.

In the second area the case of two square-pixel cameras with constant parameters has been studied, producing new results of practical interest [3], such as a explicit parameterization of the set of solutions that allows for the fast computation of the Euclidean 3D reconstruction using a single piece of knowledge of the scene, such as the parallelism or the orthogonality of two lines.

2.2.2 Segmentation and object analysis in the compressed domain

These activities confront a same objective with two different approaches, both based on compressed domain analysis. The first starts from uncompressed image sequences and performs a block-DCT image-by-image transform prior to applying conveniently modified typical segmentation approaches. The second starts from compressed video sequences (currently MPEG1 and MPEG2, and now including MPEG4); some advantages are the direct availability of estimators for features that are hard to extract at a pixel level (e.g., the motion field), and a dramatic reduction of the dimensionality (e.g., by working over DC images); the main drawback is that these techniques are highly dependent on the coding standard and on the coding parameters.

We currently apply the first approach just to static camera conditions. The method for foreground detection introduces a context-aware background subtraction method as its main innovation. In the current implementation, temporal change detection represents this context information. Assuming a non-complex background, common in many indoor surveillance applications, the method is able to work in real-time while adapting to smooth illumination changes. Results[7] show a good compromise between efficiency, accurate segmentation and robustness, being able to work with high and low frame rates while complying with real-time requirements (over 300 fps QCIF).

The second approach focuses more directly on a practical situation, as most commercial security cameras trend to output MPEG4 video. Part of the situation is similar to the previous approach, as DCT coefficients can be exactly obtained for any intra-coded macroblock. If not intra-coded, obtaining the exact value of a coefficient requires in general four block IDCTs and one block DCT; however, this can be avoided if we assume to work with an estimation of the coefficients which can be efficiently obtained. However, in the MPEG domain also a sparse motion field can be estimated from the motion estimation vectors, which are highly coding dependent. Our current effort[8][9] is directed to use this motion information to estimate the camera motion (from PTZ cameras) in order to compensate it and then apply segmentation techniques for static camera situations.

2.2.3 Segmentation and object analysis in the pixel domain

The work carried out addressed mainly the development of a set of tools which implement different approaches to the following problems: extraction of static backgrounds (BG), identifying moving objects, determination of foreground (FG) static elements of the scene, and tracking of moving objects for occlusion management and moving objects parts characterization. Researched technical solutions followed a complexity incremental approach, starting from the simplest ones to the more sophisticated –and accurate/robust- (although their real-time implementation may likely require the use of specific hardware processing elements).

Several strategies have been considered, implemented and compared for quasi-static BG/FG characterization and moving objects detection. Discrimination between non static and static (including some non-strictly static) elements has been based on a combination of several Gaussians for each pixel to model the distribution of its gray level value along the sequence carried out on sub-sampled versions of the filtered inputs. It offers a good trade-off between computational costs, adaptability to slow illumination changes, and robustness with respect to the noise. A new segmentation strategy [10] overcomes the problem of automatic detection of static elements in FG (moving object partially occluded by them). Tracking information through a Kalman filter is used to refine the different areas segmentation. Segmentation of different elements within the scene has been used to drive efficient coding strategies, based on JPEG2000 ROIs and on H.264 FMD [11][12].

Tracking is incorporated to the moving objects segmentation and static elements (BG and FG) identification to achieve: occlusion management at moving object level, and independent sub-region tracking for divided objects. Simplified implementations based on graphs theory were very efficient and could cope with simple partial occlusions when moving objects bounding boxes are tracked. Bounding box size evolution is also used to indicate possible separations or joins of independent moving objects which go together. Very efficient implementations have been carried out based on look-up tables. Kalman filtering have been incorporated to the developed techniques, but demonstrated that the simplification of using bounding boxes or centroids of the moving objects as tracking features, is only operative in those situations where the moving object size is small with respect to the observed area, and objects follow quasi-linear motion evolution. Other developed works incorporate relevant object information to the tracking: motion information (motion vectors), object gray level (color is not considered for the computational cost), and some texture indicators. Moving object regions are characterized by those parameters, which constitute representative feature vectors which are projected on the following image using Kalman filters. This complex strategy is only required when the graphs based approach do not produce an

adequate trajectory filtering, and introduces a sub-segmentation of the moving objects into joined regions, with independent motions, that need to be tracked. Therefore, the proposed final strategy is a strategy combining three subsystems: one based on graphs theory, another based on regions (including sub-segmentation of moving objects), and a third one which deals with partial observations and trajectory analysis.

2.2.4 3D scene modeling and management

MPEG-4 and, in particular, its Part 16, AFX (Animation Framework eXtension) for synthetic content, were chosen for 3D scene representation. They provide tools for the description of hierarchically structured, hybrid 3D scenes. Visual objects of an MPEG-4 scene may be grouped to form pyramidal structures to ease their hierarchical coding, transmission and rendering; and may have different nature and characteristics, as is the case of traditional, rectangular-shaped video coming from a natural source vs. synthetic 3D graphics, but also of passive/static BG vs. active/changing FG. These two characteristics, together with the scalable coding features of the MPEG-4 and AFX tools, enable flexible 3D scene modeling and management. More specifically, they provide the required handles for automatic content adaptation based on terminal, network and user profiles. Our work in the field of automatic content adaptation has led to several reports [13] [14].

Besides, researching on synthetic content generation from high-level descriptions or by means of procedural techniques has been conducted[15][16]. The expected outcome is a very efficient method for describing "fractalish" objects (textures, mountains, trees, etc.) in the scene BG, as well as for creating or modifying some active objects, such as virtual humanoids.

2.3 Semantic indexation, storage, retrieval and personalization

This task has two main outputs: generation and creation of different representations of a video surveillance scene with different levels of detail, and a framework for content adaptation. On one side, the application of technologies of generation and transmission of descriptions, allow to reduce transmission and storage requirements as well as to provide descriptions capable of helping the automatic interpretation and triggering of alarms. On the other side, the proposed content adaptation framework provides a generic and extensible adaptation engine working with MPEG-21 and MPEG-7 descriptions for both usage context description and content description.

2.3.1 Segmented objects description (MPEG-7 + content)

This subtask focuses on the generation of descriptions (following the MPEG-7 standard) in order to provide on-line adaptive transmission of video sequence allowing to send the most relevant information and conditions of use (in exchange for greater processing resources both in the sender and receiver terminals, for analysis and synthesis, respectively).

The input to this subtask is a segmentation mask coming from the analysis modules. This segmentation masks indicate the moving regions detected. The original image is also need for further description extraction.

The descriptions of moving objects can be generated at different levels of detail, ranging from the trajectory of the centroid of the moving region to the description of trajectory of the exact object shape, and including bounding boxes trajectory and associated textures, among others. Depending

on the level of quality/detail (closely related with the available bandwidth), textures and background can be transmitted either when there are significant changes, periodically, or on demand. The moving object's texture does not imply to overcome privacy issues, as if the application requires it, the texture can be not shown in regular use, being only available on demand in special authorized situations.

From the master description the system generates the different levels of detail. There are currently two parameters for the levels of detail: granularity of the description, that is, if the moving regions shape and trajectory are more or less detailed; and textures associated to each moving object in the description and the background image. Preliminary results of this approach can be found in [17].

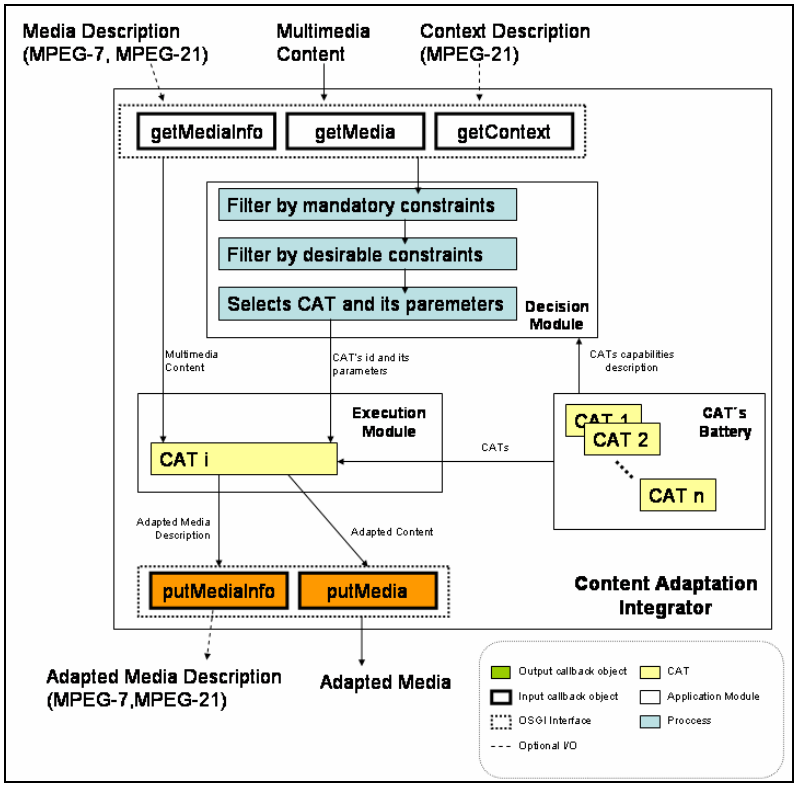


Fig. 1. Content adaptation framework: the CAIN Architecture

2.3.2 Information management, personalization and adaptation

The work in this subtask has focused on the creation of a generic and extensible content adaptation engine named CAIN (Content Adaptation INtegrator)[18][19]. CAIN constitutes a complete system for multimedia content adaptation. CAIN Works with MPEG-7 descriptions for content description and MPEG-21 DIA for usage environment description. Fig.1 depicts the CAIN architecture that, besides the input/output facilities and additional support modules, can be divided in three main functional blocks.

- The Decision Module (DM)[21] is on charge of the making the necessary decisions for performing the adaptations. It returns the Content Adaptation Tool (CAT) to be executed and the input parameters to use in the call. For getting these results the DM will take as inputs the different CATs' capabilities, the network and terminal restrictions, and the user´s preferences. This module could be considered as the one which contribution is the intelligence of the system.
- The Execution Module (EM) is a pretty simple module which aim is to execute the method responsible of the adaptation, in the indicated CAT and with the received input parameters. The decision has already been taken by the DM. In future version this module may be more complex, when dynamic adaptation (based on changing usage environment conditions) is introduced.
- CATs "battery", that is, the set of CATs made available to CAIN thanks to the definition of an extensibility mechanism[20].

Currently we have integrated CATs belonging to the four categories considered in the design of CAIN: Transcoder CATs, Scalable Conten CATs, Real-time Content Driven CATS and Transmoding CATS.

2.4 Coding and transmission control for network and client adaptation

2.4.1 Natural coding

Improvement of coding efficiency to achieve real-time coding strategies based on the development of Fast Mode Decision (FMD) strategies in the framework of H.264 is successful in the project. Two different research approaches have been followed to drive FMD: incorporation of the segmentation process results, and incorporation of thresholded distortion measures regardless of the contents of the video sequences.

The former is useful in our case because an analysis in terms of object segmentation is already available. Its target is to exploit the knowledge on the scene to reduce motion estimation computation to achieve real time coding for transmission and efficient storage. The knowledge of the moving object areas and BG areas allows to reduce the huge amount of modes permitted by H.264. A sub-set of modes is selected to obtain a more accurately motion estimation for active objects and less intensive for quasi-static BG, reducing the number of comparisons needed to perform inter prediction by 60%-70%, while keeping the same image quality and bit-rate. This work was successfully published in [11]. If BG/moving object segmentation is not available, a fast segmentation strategy aiming to detect activity regions in the images was used to drive the mode selection process. This research work [12] complements the previous one, while still exploiting some knowledge about the motion and activity present in the sequences.

Other developed alternatives have as main target to keep rate-distortion measures very close to those obtained with the full exploration of all the possible modes considered in H.264, while reducing dramatically the number of modes to be computed. These works explored predictions in terms of distortion and rate-distortion values computing adaptive thresholds to stop mode selection when quality improvement vs. computational cost is considered perceptually non relevant. Developed work [22][23] successfully improve available FMD strategies in H.264.

2.4.2 Synthetic and hybrid coding

MPEG-4 and AFX provide tools for the efficient hierarchical coding of 3D scenes, in that there are perfectly fit for streaming applications (unlike, e.g., VRML97). A clear example of this is the WSS (Wavelet Subdivision Surface) tool of AFX, which permits to model, code and transmit a 3D shape in a progressive manner: a coarse base mesh is extracted from a given fine mesh with arbitrary connectivity, representing the target surface, thanks to an automatic simplification algorithm; this base mesh is taken as the lowest LOD (Level Of Detail) of a collection of hierarchically nested meshes, obtained by recursively applying a subdivision algorithm, and approximating at increasing LOD the target surface; finally, the 3D details necessary to go from one LOD to the next are efficiently coded thanks to wavelet techniques, to yield a fully embedded bit-stream.

Bit-streams of MPEG-4's WSS tool is only SNR-scalable, as opposed to also spatially scalable: it cannot be easily parsed according to a given maximum resolution (i.e., number of subdivisions or LODs) imposed by the decoder. This is not desirable in environments with mobile platforms (PDAs, cell phones), since they have much less resources (computational power and memory) than fixed ones (PCs). We have devised a novel wavelet-based method for the hierarchical coding of 3D surfaces which features both SNR and spatial scalability, and which has much better compression efficiency than MPEG-4's WSS tool. The work has led to a publication [24] and several proposals in MPEG, which will likely adopt our PLTW (Progressive Lower Trees of Wavelet coefficients) tool in a future edition of AFX. Another limitation of MPEG-4's WSS tool is that there are several restrictions for the attributes (normal vectors, colors, texture coordinates, etc.) that can be attached to the vertices or facets of a 3D mesh. We have proposed as an alternative the IWS (Indexed Wedge Set) format, which has also led to MPEG contributions.

2.4.3 Transmission control

The framework proposed to compute optimized rate control strategies is based on the analysis of the characteristics of the main elements involved in the control problem: the source of coded information and the transmission channel. Special attention has been given to wireless channels, since its inherent stochastic behavior poses difficulty to the transmission of real-time content.

On the one hand, different stochastic packet-loss models have been developed for UMTS and Wi-Fi channels. More specifically, for UMTS Gilbert-Elliot and Fritchman models have been analyzed and a new model based on Hidden Markov chains have been extended to capture the behavior of UMTS channels. Regarding Wi-Fi, Gilbert-Elliot model has been analyzed and new models based on both Markov chains and Hidden Markov chains have been designed and implemented.

With regard to source models, our previous work on R-D models (based on an H.263 codec and for I and P frames) has been extended to an MPEG-4 codec and for I, P and B frames. With these models, from a single rate measurement (at a particular quantizer) on a given frame, it is possible to estimate any other rate and distortion value (at any other quantizer).

Finally, a basic rate control strategy has been developed for the transmission of video content in a Wi-Fi network. This preliminary approach has been focused on increasing the robustness of video transmission in a multicast scenario.

3 Relevance indicators

3.1 Educational Training

3.1.1 Master thesis

- UPM: Ainhoa Mendizábal (Jan. 2005), Laure-Hélène Caron and David Fougerousse (Feb. 2005), Alberto Castillo (Jul. 2005), Cristina Calleja (Sep. 2005), Sergio Sanz (Sep. 2005), José Mª Cobo (Sep. 2005), Sebastián Amengual (Oct. 2005), Laurent Isenegger (Nov. 2005), Marcos Nieto (Nov. 2005), Carlos Cuevas (Apr. 2006), Raquel Ribagorda (Jul. 2006).
- UAM: J. Carlos San Miguel (Sep. 2006).

3.1.2 PhD. Students

- With the Advanced Studies Diploma (DEA): Luis Herranz (UAM, Jul. 2005), Fernando López (UAM, Jul. 2006), Víctor Fernámdez-Carvajal (UAM, Jul. 2006), Enrique Estalayo (UPM, Sep. 2006), Víctor Valdés (UAM, Sep. 2006), Fabricio Tiburzi (UAM, Sep. 2006).
- 2nd year finished: José Cubero (UPM), Javier Molina (UAM), Daniel Márquez (UAM).
- 1st year finished: Carlos del Blanco (UPM), Marcos Nieto (UPM), Juan Carlos Plaza (UPM), Pablo García (UPM), Álvaro García (UAM).
- Starting: Carlos Cuevas (UPM), J. Carlos San Miguel (UAM).

3.2 Publications

3.2.1 Journal papers

- On autocalibration: [1][2][3]
- On automatic graphical content adaptation: [14]

3.2.2 Conference papers

- On autocalibration: [4][5][6]
- On segmentation and object analysis in the compressed domain: [7][8][9]
- On segmentation and tracking in the pixel domain: [10][11][12]
- On automatic graphical content adaptation: [13]
- On content generation from high-level descriptions: [15][16]
- On personalization and audiovisual content adaptation: [18][19][20][21]
- On natural coding: [22][23]
- On 3D surface coding with the PLTW technique: [24][25]

3.3 Regular collaboration with companies and EU projects

Regular contacts are maintained with companies in the sector such as Check-Out Control and Infoglobal (this via a long-term established collaboration), both interested in developments mainly related to their application for video-surveillance non-supervised systems.

Members of the research team at UPM have participated in the EU projects VISIRE (in calibration, tracking and 3D reconstruction) and OLGA (3D modeling and synthetic content and

coding); and in national project MOSSLADAR (in segmentation and tracking). As a result of part of the research work, the research team elaborated successful proposals for FP6, 36 month project I-WAY (in segmentation and tracking) and 27 month project SUIT (in video transmission and rate control) both started in February 2006. Additionally, the research team at UAM are currently involved in two FP6 EU projects (aceMedia and MESH), leading activities related to audiovisual content analysis and adaptation.

3.4 Collaboration with other research groups

Contacts have been established with national research groups in image processing/analysis/coding in Spain: GPI at Univ. Polit. Cataluña, CeTVD at Univ. Ramón Llull, GTS at Univ. Polit. Valencia, G2PI at Univ. Carlos III, GCT at Univ. Rey Juan Carlos. Due to the current research in EU projects, fluent relationship has also been established with European research groups in image processing/analysis/coding and transmission: IT at Univ. Aveiro (PT), Dep. of Computer Science at Univ. Ioannina (GR), CCSR at Univ. Surrey (UK), IBBT at Univ. Ghent (BE), ETRO at Vrije Universiteit Brussel (BE), DESICS Dept. at IMEC (BE), ARTEMIS Unit at INT (FR), MVLab at Queen Mary Univ. of London (UK), IVML at NT Uni. Athens (GR).

4 References

[1] A. Valdés and J.I. Ronda, "Camera autocalibration and the calibration pencil", *Journal of Mathematical Imaging and Vision*, vol. 23, no. 2, pp. 167-174, Sep. 2005

[2] A. Valdés, J.I. Ronda, and G. Gallego, "The absolute line quadric and camera autocalibration", *International Journal of Computer Vision*, vol. 66, no. 3, pp. 283-303, Mar. 2006.

[3] J.I. Ronda and A. Valdés, "Conic geometry and autocalibration from two images", accepted conditionally to a minor revision by the *Journal of Mathematical Imaging and Vision*, 2007.

[4] J.I. Ronda, A. Valdés, and G. Gallego, "Camera autocalibration using Plücker coordinates", *Proc. IEEE Int. Conf. on Image Processing ICIP 2005*, vol. 3, pp. 800-803, Genoa, Italy, Sept. 2005.

[5] G. Gallego, J.I. Ronda, and A. Valdés, "Recursive camera autocalibration with the Kalman filter", submitted to *IEEE Int. Conf. on Image Processing ICIP 2007*.

[6] P. Pérez and N. García, "Robust and accurate registration of images with unknown relative orientation and exposure", *Proc. IEEE Int. Conf. on Image Processing ICIP 2005*, vol. 3, pp. 1104-1107, Genoa, Italy, Sept. 2005.

[7] A. García, J. Bescós, "Real-time video foreground extraction based on content-aware background subtraction", submitted to *IEEE Int. Conf. on Image Processing ICIP 2007*.

[8] L. Herranz, J. Bescós, "Reliability based optical flow estimation from mpeg compressed data", *Proc. International Workshop on Very Low Bit-rate Video*, 4 pags. (CD-ROM), Sardinia (Italia), Sept. 2005.

[9] F. Tiburci, J. Bescós, "Camera Motion Analysis in On-line MPEG Sequences", submitted to the *Int. Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS 2007*.

[10] L. Isenegger, L. Salgado, and N. García, "Moving objects segmentation based on automatic foreground/background identification of static elements", *Proc. Advanced Concepts for Intelligent Vision Systems ACIVS 2005*, vol. LNCS 3708, pp. 491-498, 2005.

[11] M. Nieto, L. Salgado, and J. Cabrera, "Fast Mode Decision and motion estimation with object segmentation in H.264/AVC encoding", *Proc. Advanced Concepts for Intelligent Vision Systems ACIVS 2005*, vol. LNCS 3708, pp. 571-578, 2005.

[12] M. Nieto, L. Salgado, and N. García, "Fast mode decision based on activity segmentation in H. 264/AVC encoding", *VLBV 2005*, vol. LNCS 3893, pp. 146-152, 2006.
[13] L. Salgado, J. Bescós, F. Morán, J. Cabrera, E. Estalayo, and J. Cubero, "Efficient Architecture for Low-Cost Adaptive 3D Graphics-Based E-Learning Applications", *Proc. EuroGraphics EG 2005, Computer Graphics Forum*, vol. 24, no. 3, pp. 41-50, Dublin, IE, Aug. 2005.
[14] F. Morán, M. Preda, G. Lafruit, P. Villegas, and R.-P. Berretty, "3D Game Content Distributed Adaptation in Heterogeneous Environments", accepted conditionally to a minor revision by the *EURASIP Journal on Applied Signal Processing* (special issue on Video Adaptation in Heterogeneous Environments).
[15] V. Fernández-Carbajales, J.M. Martínez, and F. Morán, "Description-Driven Generation of 3D Humanoids Within Authoring 744", *Proc. IEE European Workshop on the Integration of knowledge, semantics and digital Media Technologies EWIMT 2005*, pp. 227-231, London, UK, Nov. 2005.
[16] V. Fernández-Carbajales, J.M. Martínez, and F. Morán, "High-Level Description Tools for Humanoids", *Proc. intl. workshop on Multimedia content Representation, Classification and Security MRCS 2006,* vol. 4105, pp. 387-394, Istanbul, TR, Sep. 2006.
[17] J.C. San Miguel, J. M. Martínez, "On-line adaptive video sequence transmission based on generation and transmission of descriptors", submitted to *IEEE Int. Conf. on Image Processing ICIP 2007*.
[18] V. Valdés, J.M. Martínez, "Content Adaptation Tools in the CAIN framework", *Visual Content Processing and Representation*, Lecture Notes in Computer Science, Vol. 3893, 2006.
[19] J.M. Martínez, V. Valdés, J. Bescós, L. Herranz, "Introducing CAIN: a Metadata-driven Content Adaptation Manager Integrating Heterogeneous Content Adaptation Tools", *Proc. Int. Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS'2005*, Apr. 2005
[20] V. Valdés, J.M. Martínez, "Content Adaptation Capabilities Description Tool for Supporting Extensibility in the CAIN Framework", *Multimedia Content Representation, Classification and Security-MCRS2006*, Lecture Notes in Computer Science, 2006.
[21] F. López, J.M. Martínez, V. Valdés, "Multimedia Content Adaptation within the CAIN framework via Constraints Satisfaction and Optimization", *Adaptative Multimedia Retrieval-AMR06*, Lecture Notes in Computer Science, 2006.
[22] L. Salgado, and M. Nieto, "Sequence independent very fast mode decision algorithm on H.264/AVC baseline profile", *Proc. IEEE Int. Conf. on Image Processing ICIP 2006*, vol.1, pp. 41-44, Oct. 2006.
[23] M. Nieto, L. Salgado, and J. Cabrera, "Fast Mode Decision on H.264/AVC main profile encoding based on PSNR predictions", *Proc. IEEE Int. Conf on Image Processing ICIP 2006*, pp. 49-52, Oct. 2006.
[24] M. Avilés, F. Morán and N. García, "Progressive Lower Trees of Wavelet Coefficients: Efficient Spatial and SNR Scalable Coding of 3D Models", *Proc. Pacific-rim Conf. on Multimedia PCM 2005,* vol. LNCS 3767, pp. 61-72, Jeju, KR, Nov. 2005.
[25] M. Avilés and F. Morán, "Spatial and SNR Scalable Coding of Wavelet Subdivision Surfaces", *MPEG contribution* M12329, Poznań, PL, Jul. 2005.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

GRAMMARS: GRAph-based Methods for the Modelling, Analysis and Realization of large-scale Systems TIN2004-07925-C03

Fernando Orejas *
Departamento de Lenguajes y Sistemas Informáticos
Universidad Politécnica de Cataluña

Abstract

As technology provides more storage, communication and computational resources, the problems faced by Computer Scientists increase in size and complexity. This trend leads to the non-scalability or non-applicability of methods that were traditionally considered adequate. This project aims at understanding, studying and solving problems that are becoming bottlenecks in the new technological scenarios. In particular, the project will focus on the modelling, verification and realization of complex large-scale systems. Different problems will be tackled such as system modelling, verification and visualization using graph transformations, layout synthesis for VLSI circuits using combinatorial optimization. Even though the project will tackle problems of different nature, all of them have a common underlying structure (graphs), and many of them use related techniques for solving similar subproblems. The project is expected to produce fundamental contributions in the aforementioned areas and provide new algorithms and tools with a practical impact in academia and industry.

1 Goals of the project

The overall goal of the project is to contribute with foundations, techniques and methods to the design, verification and realization of large and complex systems, using different kinds of techniques, in particular graph-based methods together with algorithmic and learning techniques. More specifically, the concrete goals of GRAMMARS were originally:

1. Development of a new framework for the specification and modelling of component-based systems, including an operational semantics based on distributed computation that could be the base for prototyping.
2. Development of logic-based and model checking-based verification techniques for this kind of systems.

*Email: orejas@lsi.upc.edu

3. Development of algorithms and algorithmic techniques, especially in connection with the previous goals.

4. Development of algorithmic learning techniques in connection with graph problems.

However, currently (as it was said in our last yearly report), we have given up one of the above research lines and started two new ones. In particular, although some work was done, we have abandoned, to a large extent, the work on verification (especially the work based on model checking). The main reason is that this work was done mainly in cooperation with Intel Research Labs, which were not anymore interested. Instead, we have started a new research line, together with Intel, on the design of elastic networks, which are synchronous systems that have an asynchronous behaviour, i.e. they allow any variation on the latency of its components. In addition, Gabriel Valiente has started also a new research line using graph-based methods, including methods based on graph transformation, in computational biology.

2 Project progress and achievements

In this section we present the main results obtained in GRAMMARS. The section is structured in 5 parts corresponding to the main research lines of the project.

2.1 Modelling of component-based systems

In cooperation with the group at TU Berlin, we have developed a transformation-based generic framework for the architectural modelling of component-based systems using the notion of architectural connector from Allen and Garlan. This framework has been instantiated to the case of Petri Nets and to some UML diagrams (class, state and sequence diagrams). Preliminary results have been presented in [66, 44]. A detailed description of the approach has been submitted for journal publication. The instantiation of the framework to the case of UML class and sequence diagrams is described in detail in the PhD thesis of Sonia Pérez, including the implementation of a plugin for the UML tool Visual Paradigm.

Also in cooperation with the group at TU Berlin, we have been working on defining an operational semantics for the approach based on some form of distributed computation. In particular, we are studying the use of distributed graph transformation for these purposes. Preliminary work including a categorical characterization of distributed graph transformation has been presented in [39].

2.2 Verification

As explained in the section above, most of the work in this area has been given up. Anyhow, it is worth to mention some results that, nevertheless, have been obtained.

In cooperation with the group at TU Berlin, we are studying how to detect conflicts that may yield to inconsistencies for modelling approaches based on graph transformation. The results obtained up to now have been presented in [41, 65].

We have defined a new abstract domain based on octahedrons, which has been shown to be very effective for the analysis of concurrent systems using abstract interpretation. The

approach is fully described in the PhD thesis of Robert Clarisó and the main ideas have been published in [9, 34, 57].

We are studying the use of temporal logic for reasoning about graph transformation. In this context, we have developed a new proof technique for propositional temporal logic which has been shown to be complete but that it is cut-free, which means that can be easily automated. Actually, the technique has been implemented in Isabelle. A paper describing these results has been submitted for publication.

To end this subsection, we should also mention some work in a completely new and exciting area. In particular, we have done some work on exploring the use of model checking techniques in the context of molecular computation. The work, in a sense, is quite preliminary, but as the whole area of molecular computing is. This work has been presented in [56].

2.3 Design of concurrent systems

As explained in the section above, we have started a new research line on the design of concurrent systems with the introduction of the notion of elastic networks. This notion, defined using the project, is a very powerful concept that can be used for the design of asynchronous systems and for synchronous systems that have some asynchronous behaviour. As described elsewhere in this report, this notion has attracted the interest of Intel and Esterel. Several publications show the relevance of the concept both for the design of synchronous and asynchronous systems [4, 10, 55, 58, 59, 62, 63].

2.4 Graph Algorithms

In GRAMMARS we have studied a number of algorithmic problems related with the topics of the project. In particular, special attention has been paid to the problem of graph layout, directly related to the problem of placement in VLSI [25, 26]. This placement problem, which has resulted more difficult than expected, has been studied in a special case in [54]. Another problem that has been studied in some detail, obtaining good results is the graph homomorphism problem, formulated as the graph restrictive H-coloring problem, of interest for the implementation of graph transformation tools. These results have been presented in [14, 15].

An algorithmic technique that has been explored in some detail is the design of parameterized algorithms. Relevant results obtained for different kinds of problems have been presented in [12, 13, 23, 37, 43, 48].

As discussed in the previous section, within GRAMMARS we have also started a new research area, not foreseen when the project was defined, on the application of graph-based techniques in computational biology. This new line has been very productive and has been especially centered around the analysis of biochemical networks and the reconstruction of phylogenies. The results obtained have been presented in [6, 7, 8, 17, 18, 20, 21, 27, 46, 47, 60].

Finally, some work has also been done in an area not directly, but only indirectly, related to the design of algorithms. In particular, to design and, especially, analyze algorithms we must have, beforehand, the basic computation model that would be used to describe the behaviour of the algorithm. In the case of traditional algorithms this basic model is usually implicit. However, in the case of systems distributed over an open network, as, e.g. Internet, the

adequate model to use is not obvious. In this sense, within GRAMMARS we have dealt with this problem obtaining results presented in [1, 16, 35].

2.5 Learning

The main research problem in this area was the development of algorithmic learning techniques for graphs. First, we dealt with a simplified problem, Sagiv formulas. Then we concentrated the work on Horn formulas which allowed us to obtain preliminary results for non-directed labelled graphs and for unlabelled trees. The main results can be found in the PhD thesis of Gemma Casas and in the forthcoming thesis of Jaume Baixeries, and in [2, 3, 30, 31, 33, 40, 61].

In the project proposal, we also considered (as a side problem) the application of learning techniques to natural language processing. In particular, in this topic, the main contribution has been a phrase recognition architecture, which we name Filtering-Ranking, and a global learning strategy for it based on Perceptron. Implementations of our approach participated in the CoLL competition and on Clause Identification, our system was substantially better than any other system, and for Named Entity Extraction and Semantic Role Labeling, our systems are among the top-performing ones. In the case of Shallow Syntactic Parsing, the task with more evaluated systems, our system obtained results that were very close to the top-performing system. The main results can be found in [5, 24].

3 Quantification of Results

In this section we are going to describe quantitatively the results of the project. More precisely, in the first subsection we will describe the impact of the project in PhD training. Then, we will provide an overall description of the publications produced by GRAMMARS. The concrete publications (the most relevant ones) can be found below, in the References section. In the following subsection, we describe the technology transfer activities related to GRAMMARS. Then, we enumerate the UE projects in which we participate. Finally, in the last subsection we describe our cooperation with other groups.

3.1 PhD training

14 PhD students participate in GRAMMARS. In the first two years, 4 thesis have been presented:

- Robert Clarisó: Abstract interpretation techniques for the verification of timed systems. Supervisor: Jordi Cortadella. Departamento de Lenguajes y Sistemas Informáticos. Universidad Politécnica de Cataluña. 2005
- Gemma Casas: Formal methods for mining structured objects. Supervisor: José Luis Balcázar. Departamento de Lenguajes y Sistemas Informáticos. Universidad Politécnica de Cataluña. May 2006.
- X. Carreras: Learning and Inference in Phrase Recognition: A Filtering-Ranking Architecture using Perceptron. Supervisor: Lluis Márquez. Departamento de Lenguajes y Sistemas Informáticos. Universidad Politécnica de Cataluña. October 2005

- Sonia Pérez: Uso y extensión de UML para la especificación y chequeo de consistencia en el empleo de la arquitectura de componentes genéricos. Supervisors: Fernando Orejas and Exiquio Leyva. Centro de Estudios de Ingeniería y Sistemas. Instituto Superior Politécnico José Antonio Echeverría - La Habana, January 2007

Moreover, currently two more thesis are almost ready to be presented since they have passed or are passing the refereeing process of the department. This is the case of the thesis of Edelmira Pasarella and Jaume Baixeries.

3.2 Publications

The list of the most relevant publications of the project can be found below, in the references. These publications are structured as follows. References 1-27 are journal papers (not considering the Springer Lecture Notes Series as a Journal). In the references, when the corresponding journal is indexed by the Journal Citation Reports or by Citeseer, we include its impact factor. References 28-47 are papers in volumes in the Springer Lecture Notes Series, which typically (but not always) are conference proceedings. When the corresponding conference is indexed by Citeseer, we also include its impact factor. References 48-49 are book chapters (not including conference proceesings nor books from the Springer Lecture Notes Series) and 50-53 are books edited by members of GRAMMARS. References 54-63 are papers at conferences that, for some reason, we consider relevant. Again, when the corresponding conference is indexed by Citeseer, we also include its impact factor. Finally, references 64-66 are papers published in an Electronic Journal which are revised versions of papers presented at workshops.

Looking at these references one may see that the participants in GRAMMARS have published 22 papers in journals that are indexed by the Journal Citation Reports (not including the Lecture Notes series), 15 of which have been published in journals with an impact rate greater or equal to 1. Similarly, the participants in GRAMMARS have presented 20 papers in conferences that are indexed by Citeseer, 15 of which have been published in conferences with an impact rate in the first 33,33% of the list (impact rate over 0.77).

3.3 Technology transfer

GRAMMARS has had strong cooperation with two companies. On one hand, at the Spanish level, have cooperated within a PROFIT project (Sistema Inteligente de Gestión Estratégica de la Informacióon de Seguridad),with the company TISSAT S.A. on the application of learning techniques for information security.

On the international level, there is a long cooperation of members of GRAMMARS with Intel Co in the area of design, synthesis and verification of concurrent systems. In particular, the group holds a grant from Intel since 1999. The main topic of collaboration, as explained above, has been the synthesis and verification of elastic networks. On the other hand, the group has also a log standing collaboration with Cadence Design Systems, also in the area of design, synthesis and verification of concurrent systems. Finally, Esterel Technologies has also shown some interest on our work on elastic networks and we have started to collaborate.

The work with Cadence has produced the following patent:

AUTHORS: Jordi Cortadella, Alex Kondratyev, Luciano Lavagno

NAME: Skew insensitive clocking method and apparatus
REGISTRY: 7037532001
OWNER: Cadence Design Systems
YEAR: 2005
COUNTRIES: USA

3.4 European projects

In parallel with GRAMMARS, the members of the group have also participated in the following European projects, some of which were ongoing projects when GRAMMARS started.

- Segravis (RTN2-2001-00346) 2002-2006. **Spanish coordinator**: Fernando Orejas
  Partners: Univ. Padeborn, Univ. Antwerpen, Univ. Tec. Berlín, Univ. Bremen, Univ Kent, Univ. Leiden, Univ. College Londres, Univ. degli Studi Milan, Univ. Fuerzas Armadas Munich, Univ. Pisa, Univ. Roma 1, Univ. Polit. Catalunya.

- COMBSTRU (HPRN-CT2002-00278) 2002-2006. **Spanish coordinator**: Josep Diaz
  Partners: Barcelona UPC, Berlin Freie Univ., Humbodlt Univ. and Tech. Univ. Helmut Alt Bielefeld University, Bordeaux LABRI, Budapest Renyi Institute (Hungarian Academy of Sciences), Oxford Oxford University, Patras CTI, Pisa CNR-IMC, Prague DIMATIA.

- DELIS (FP6-FET pro-active Integrated Project 001907) 2004-2008. **Spanish coordinator**: Josep Diaz
  Partners: University of Paderborn (D), Technical University Munchen (D), ETH Zurich (CH), Universitat Politécnica de Catalunya (E), Ecole Polytechnique Federal de Lausanne (CH), Computer Technology Institute (EL), University of Cyprus (CY), University of Karlsruhe (D), DIMATIA, Charles University (CZ), Polytechnic University of Wroclaw (PL), Universita di Bolonia (I), Telenor Communication (N), Max-Plank Institute (D), Universitat Pompeu Fabra (E), Telecom Italia Learning Services S.p.A (I), University of Cambridge (UK), Universita degli Studi di Roma "La Sapienza" (I)

- AEOLUS (FET pro-active Integrated Project 15964) 2005-2009. **Spanish coordinator**: María José Serna
  Partners: University of Paderborn (D), Technical University Munchen (D), ETH Zurich (CH), Universitat Politécnica de Catalunya (E), Ecole Polytechnique Federal de Lausanne (CH), Computer Technology Institute (EL), University of Cyprus (CY), University of Karlsruhe (D), DIMATIA, Charles University (CZ), Polytechnic University of Wroclaw (PL), Universita di Bolonia (I), Telenor Communication (N), Max-Plank Institute (D), Universitat Pompeu Fabra (E), Telecom Italia Learning Services S.p.A (I), University of Cambridge (UK), Universita degli Studi di Roma "La Sapienza" (I)

- PASCAL (IST Excellence Network). **Spanish coordinator**: José Luis Balcázar.
  The network includes 55 sites.

3.5 Cooperation with other research groups

As one may see from the publications below, GRAMMARS participants have collaborated with quite a number of people. However, there are some groups with which there are stronger links that we may especially mention: At the national level, there is a very strong collaboration between Gabriel Valiente and the group of Francesc Rosselló in Mallorca. At the international level, the group of Fernando Orejas and the group of Hartmut Ehrig in Berlin have been working together for almost 20 years. Finally, Jordi Cortadella and his group have also been cooperating with the group of Mike Kishinevski at Intel for more than ten years.

References

[1] Carme Àlvarez, Maria J. Blesa, Josep Díaz, Maria J. Serna, Antonio Fernández: Adversarial models for priority-based networks. *Networks* 45(1): 23-35 (2005). Impact JCR: 0.742

[2] Jose L. Balcazar, Yang Dai, Junichi Tanaka, and Osamu Watanabe. Provably Fast Training Algorithms for Support Vector Machines, to appear in *Theory of Computing Systems*. Impact JCR: 0.818. Impact Citeseer: 1.02

[3] J.L.Balcazar, G.C.Garriga. On Horn Axiomatizations for Sequential Data. To appear in *Theoretical Computer Science*. Impact JCR: 0.743, Impact Citeseer: 0.91

[4] Josep Carmona, José M. Colom, Jordi Cortadella, and Fernando García-Vallés. Synthesis of asynchronous controllers using integer linear programming. *IEEE Transactions on Computer-Aided Design*, 25(9):1637-1651, September 2006. Impact JCR: 1.000

[5] Xavier Carreras, Lluís Màrquez and Jorge Castro. Filtering-Ranking Perceptron Learning for Partial Parsing. *Machine Learning Journal*, Special Issue on Learning in Speech and Language Technologies, volume 60(1-3):41-71, Sept. 2005. Impact JCR: 3.10, Impact Citeseer: 2.20

[6] J. C. Clemente, K. Satou, G. Valiente. Reconstruction of Phylogenetic Relationships from Metabolic Pathways based on the Enzyme Hierarchy and the Gene Ontology. *Genome Informatics* 16(2):45-55 (2005).

[7] J. C. Clemente, K. Satou, G. Valiente. Finding Conserved and Non-Conserved Regions Using a Metabolic Pathway Alignment Algorithm. *Genome Informatics* 17(2):46-56 (2006).

[8] J. C. Clemente, K. Satou, G. Valiente. Phylogenetic Reconstruction from Non-Genomic Data. *Bioinformatics* 23(2):e110-e115 (2007). Impact JCR: 6.02

[9] Robert Clarisó and Jordi Cortadella. The octahedron abstract domain. *Science of Computer Programming*, 64(1):115-139, January 2007. Impact JCR: 0.73, Impact Citeseer: 1.22

[10] Jordi Cortadella, Alex Kondratyev, Luciano Lavagno, and Christos Sotiriou. Desynchronization: Synthesis of asynchronous circuits from synchronous specifications. *IEEE Transactions on Computer-Aided Design*, 25(10):1904-1921, October 2006. Impact JCR: 1.000

[11] Jordi Cortadella, Alex Kondratyev, Luciano Lavagno, Claudio Passerone, and Yosinori Watanabe. Quasi-static scheduling of independent tasks for reactive systems. *IEEE Transactions on Computer-Aided Design*, 24(10):1492-1514, October 2005. Impact JCR: 1.000

[12] Erik D. Demaine, Mohammad Taghi Hajiaghayi, Dimitrios M. Thilikos: Exponential Speedup of Fixed-Parameter Algorithms for Classes of Graphs Excluding Single-Crossing Graphs as Minors. *Algorithmica* 41(4): 245-267 (2005). Impact JCR: 1.017, Impact Citeseer: 1.31

[13] Erik D. Demaine, Fedor V. Fomin, Mohammad Taghi Hajiaghayi, Dimitrios M. Thilikos: Subexponential parameterized algorithms on bounded-genus graphs and H-minor-free graphs. *Journal of the ACM* 52(6): 866-893 (2005). Impact JCR: 2.19, Impact Citeseer: 1.37

[14] J. Diaz, M. Serna, and D. M. Thilikós. Complexity issues on bounded restrictive H-coloring. *Discrete Mathematics*, 2006. To appear. Impact JCR: 0.346.

[15] J. Diaz, M. Serna, and D. M. Thilikós. The restrictive H-coloring problem. *Discrete Applied Mathematics*, 145:297-305, 2005. Impact JCR: 0.585, Impact Citeseer: 0.50

[16] J. Diaz, M. Serna, and N. C. Wormald. Bounds on the bisection width for random d-regular graphs. *Theoretical Computer Science*, 2006. To appear. Impact JCR: 0.743, Impact Citeseer: 0.91

[17] L. Félix, G. Valiente. Validation of Metabolic Pathway Databases based on Chemical Substructure Search. To appear in *Biomolecular Engineering* (2007). Impact JCR: 1.43

[18] M. Llabrés, J. Rocha, F. Rosselló, G. Valiente. On the Ancestral Compatibility of Two Phylogenetic Trees with Nested Taxa. *Journal of Mathematical Biology* 53(3):340-364 (2006). Impact JCR: 1.61

[19] F. Orejas: Thirty years of algebraic specification. *J. Log. Algebr. Program.* 67(1-2): 1-2 (2006) Impact JCR: 2.125

[20] R. D. M. Page, G. Valiente. An Edit Script for Taxonomic Classifications. *BMC Bioinformatics* 6:208 (2005). Impact JCR: 4.96

[21] T. H. Pham, D. H. Tran, T. B. Ho, K. Satou, G. Valiente. Qualitatively Predicting Acetylation and Methylation Areas in DNA Sequences. *Genome Informatics* 16(2):3-11 (2005).

[22] F. Rosselló, G. Valiente. An Algebraic View of the Relation between Largest Common Subtrees and Smallest Common Supertrees. *Theoretical Computer Science* 362(1-3):33-53 (2006). Impact JCR: 0.743, Impact Citeseer: 0.91

[23] M. Serna and D. M. Thilikos. Parameterized complexity for graph layout problems. *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS*, (86):41-65, 2005.

[24] Mihai Surdeanu, Lluís Màrquez, Xavier Carreras,and Pere R. Comas. Combination Strategies for Semantic Role Labeling. *Journal of Artificial Intelligence Research*. Forthcoming. Impact JCR: 2.25, Impact Citeseer: 2.45

[25] D. M. Thilikos, M. J. Serna, and H. Bodlaender. Cutwidth I: A constructive linear time algorithm for cutwidth. *Journal of Algorithms*, 56(1):1-24, 2005. Impact JCR: 1.138, Impact Citeseer: 1.42

[26] Dimitrios M. Thilikos, Maria Serna, and Hans L. Bodlaender. Cutwidth. II. Algorithms for partial w-trees of bounded degree. *Journal of Algorithms*, 56(1):25-49, 2005. Impact JCR: 1.138, Impact Citeseer: 1.42

[27] G. Valiente. Assessing Clustering Results with Reference Taxonomies. *Genome Informatics* 17(2):131-140 (2006).

PUBLICATIONS IN THE SPRINGER LECTURE NOTES SERIES

[28] J. Alvez and P. Lucio: An Algorithm for Local Variable Elimination in Normal Logic Programs, *Logic Based Program Synthesis and Transformation, 15th International Symposium, LOPSTR 2005. Revised Selected Papers*. Springer Lecture Notes in Computer Science 3901 (2006) pp: 61-79. Impact of the conference in Citeseer: 0.85

[29] Javier Álvez, Paqui Lucio: Equational Constraint Solving Via a Restricted Form of Universal Quantification. *Foundations of Information and Knowledge Systems, 4th International Symposium, FoIKS 2006*, Springer Lecture Notes in Computer Science 3861 (2006) pp: 2-21. Impact of the conference in Citeseer: 0.28

[30] Jaume Baixeries, José L. Balcázar: Characterization and Armstrong Relations for Degenerate Multivalued Dependencies Using Formal Concept Analysis. *Formal Concept Analysis, Third International Conference, ICFCA 2005*. Lecture Notes in Computer Science 3403 Springer 2005: 162-175

[31] José L. Balcázar, Gemma Casas-Garriga: On Horn Axiomatizations for Sequential Data, en *Database Theory - ICDT 2005*. Lecture Notes in Computer Science 3363 Springer 2005, pages 215-229. Impact of the conference in en Citeseer: 1.41

[32] Hans L. Bodlaender, Fedor V. Fomin, Arie M. C. A. Koster, Dieter Kratsch, Dimitrios M. Thilikos: On Exact Algorithms for Treewidth. *Algorithms - ESA 2006, 14th Annual European Symposium*, Springer Lecture Notes in Computer Science 4168 (2006) pp: 672-683. Impact of the conference in Citeseer: 0.99

[33] Jorge Castro: How Many Query Superpositions Are Needed to Learn? *Algorithmic Learning Theory, 17th International Conference, ALT 2006*, Springer Lecture Notes in Computer Science 4264 (2006) pp: 78-92. Impact of the conference in Citeseer: 0.63

[34] R. Clarisó, E. Rodríguez-Carbonell, and J. Cortadella. Derivation of non-structural invariants of petri nets using abstract interpretation. In *Applications and Theory of Petri Nets 2005, 26th International Conference, ICATPN 2005*. Lecture Notes in Computer Science 3536 Springer 2005: 188-207. Impact of the conference in Citeseer: 0.74

[35] Josep Díaz, Xavier Pérez, Maria J. Serna, and Nicholas C. Wormald. Connectivity for wireless agents moving on a cycle or grid. In *STACS 2005, 22nd Annual Symposium on Theoretical Aspects of Computer Science* volume 3404 of Lecture Notes in Comput. Sci., pages 353-364. Springer, Berlin, 2005. Impact of the conference in Citeseer: 0.80

[36] Josep Díaz, Jordi Petit, Dimitrios M. Thilikos: Kernels for the Vertex Cover Problem on the Preferred Attachment Model. *Experimental Algorithms, 5th International Workshop, WEA 2006*, Springer Lecture Notes in Computer Science 4007 (2006) pp. 231-240.

[37] Josep Díaz, Dimitrios M. Thilikos: Fast FPT-Algorithms for Cleaning Grids. En Bruno Durand, Wolfgang Thomas (Eds.): *STACS 2006, 23rd Annual Symposium on Theoretical Aspects of Computer Science*, Marseille, France, February 23-25, 2006, Proceedings. Springer Lecture Notes in Computer Science 3884 (2006) pp. 361-371. Impact of the conference in Citeseer: 0.80

[38] Frederic Dorn, Fedor V. Fomin, Dimitrios M. Thilikos: Fast Subexponential Algorithm for Non-local Problems on Graphs of Bounded Genus. *Algorithm Theory - SWAT 2006, 10th Scandinavian Workshop on Algorithm Theory*, Springer Lecture Notes in Computer Science 4059 (2006) pp: 172-183. Impact of the conference in Citeseer: 0.80

[39] H. Ehrig, F. Orejas, U. Prange: Categorical foundations of Distributed graph transformation, in *Graph Transformation, Third International Conference, ICGT 2006*, Springer Lecture Notes in Computer Science 4178 (2006), pp. 215 - 230. Impact of the conference in Citeseer: 0.93

[40] J. Gaintzarain, M. Hermo and M. Navarro: Goals in the Propositional Horn$^{\supset}$ Language are Monotone Boolean Circuits, *Mathematical Foundations of Computer Science 2005, 30th International Symposium, MFCS 2005* Lecture Notes in Computer Science. Proceedings of MFCS'05. Volume 3618, Springer-Verlag 2005. Impact of the conference in Citeseer: 0.48

[41] L. Lambers, H. Ehrig, F. Orejas: Conflict detection for graph transformation with negative application conditions, in *Graph Transformation, Third International Conference, ICGT 2006*, Springer Lecture Notes in Computer Science 4178 (2006), pp. 61 - 77. Impact of the conference in Citeseer: 0.93

[42] Paqui Lucio, Fernando Orejas, Edelmira Pasarella, Elvira Pino: A Functorial Framework for Constraint Normal Logic Programming. En *Algebra, Meaning, and Computation, Essays Dedicated to Joseph A. Goguen on the Occasion of His 65th Birthday.* Lecture Notes in Computer Science 4060 Springer 2006: 555-577.

[43] Naomi Nishimura, Prabhakar Ragde, Dimitrios M. Thilikos: Parameterized Counting Algorithms for General Graph Covering Problems. En Frank K. H. A. Dehne, Alejandro López-Ortiz, Jörg-Rüdiger Sack (Eds.): *Algorithms and Data Structures*, 9th International Workshop, WADS 2005, Waterloo, Canada, August 15-17, 2005, Proceedings. Springer Lecture Notes in Computer Science 3608 (2005) pp. 99-109. Impact of the conference in Citeseer: 0.82

[44] Fernando Orejas, Sonia Pérez: Towards Architectural Connectors for UML. *Formal Methods in Software and Systems Modeling, Essays Dedicated to Hartmut Ehrig, on the Occasion of His 60th Birthday.* Lecture Notes in Computer Science 3393 Springer 2005: 352-369.

[45] Edelmira Pasarella, Fernando Orejas, Elvira Pino, Marisa Navarro: A Transformational Semantics of Static Embedded Implications of Normal Logic Programs. *Logic Based Program Synthesis and Transformation, 15th International Symposium, LOPSTR 2005. Revised Selected Papers*. Springer Lecture Notes in Computer Science 3901 (2006) pp: 133-146. Impact of the conference in Citeseer: 0.85

[46] F. Rosselló, G. Valiente. Graph Transformation in Molecular Biology. In H.-J. Kreowski, U. Montanari, F. Orejas, G. Rozenberg, G. Taentzer (Ed.), *Formal Methods in Software and System Modeling*, Lecture Notes in Computer Science 3393 (2005), pp. 116-133.

[47] G. Valiente. A Fast Algorithmic Technique for Comparing Large Phylogenetic Trees. In *Proc. 12th Int. Symp. String Processing and Information Retrieval*, Lecture Notes in Computer Science 3772 (2005), pp. 370-375.

BOOK CHAPTERS

[48] Josep Díaz, Maria Serna, and Dimitrios M. Thilikos. Efficient algorithms for parameterized H-colorings. In *Topics in discrete mathematics*, volume 26 of Algorithms Combin., pages 373-405. Springer, Berlin, 2006.

[49] G. Valiente. Efficient Algorithms on Trees and Graphs with Unique Node Labels. To appear in H. Bunke, A. Kendel, M. Last (Ed.), *Applied Graph Theory in Computer Vision and Pattern Recognition*, Springer-Verlag, 2006.

BOOKS

[50] J. L. Balcázar, P. M. Long, F. Stephan (Eds.): *Algorithmic Learning Theory, 17th International Conference, ALT 2006*, Barcelona, Spain, October 7-10, 2006, Proceedings. Lecture Notes in Computer Science 4264 Springer 2006.

[51] José Luiz Fiadeiro, Peter D. Mosses, Fernando Orejas (Eds.): *Recent Trends in Algebraic Development Techniques*, Revised selected papers from the 17th International Workshop, WADT 2004. Lecture Notes in Computer Science 3423 Springer 2005.

[52] Hans-Jörg Kreowski, Ugo Montanari, Fernando Orejas, Grzegorz Rozenberg, Gabriele Taentzer (Eds.): *Formal Methods in Software and Systems Modeling, Essays Dedicated to Hartmut Ehrig, on the Occasion of His 60th Birthday*. Lecture Notes in Computer Science 3393 Springer 2005.

[53] M. Lewenstein, G. Valiente (Eds.), *Combinatorial Pattern Matching, 17th Annual Symposium, CPM 2006*, Lecture Notes in Computer Science 4009 (2006).

CONFERENCES

[54] D. Bañeres, J. Cortadella, and M. Kishinevsky. Layout-aware gate duplication and buffer insertion. In *Proc. Design, Automation and Test in Europe (DATE)*, April 2007. (To appear). Impact of the conference in Citeseer: 0,76

[55] J. Carmona and J. Cortadella. State encoding of large asynchronous controllers. In Proc. *ACM/IEEE Design Automation Conference*, pages 939-944, July 2006. Impact of the conference in Citeseer: 1.16

[56] J. Carmona, J. Cortadella, Y. Takada, and F. Peper. From molecular interactions to gates: a systematic approach. In *Proc. International Conf. Computer-Aided Design (ICCAD)*, November 2006. Impact of the conference in Citeseer: 1.09

[57] R. Clarisó and J. Cortadella. Verification of concurrent systems with parametric delays using octahedra. In *Int. Conf. on Application of Concurrency to System Design*, pages 122-131, June 2005.

[58] J. Cortadella, K. Gorgônio, F. Xia, and A.Yakovlev. Automatic synthesis of asynchronous coomunication mechanisms. In *Int. Conf. on Application of Concurrency to System Design*, pages 166-175, June 2005.

[59] J. Cortadella, M. Kishinevsky, and B. Grundmann. Synthesis of synchronous elastic architectures. In *Proc. ACM/IEEE Design Automation Conference*, pages 657-662, July 2006. Impact of the conference in Citeseer: 1.16

[60] L. Félix, G. Valiente. Efficient Validation of Metabolic Pathway Databases. In *Proc. 6th Int. Symp. Computational Biology and Genome Informatics*, Salt Lake City, Utah, USA (2005), pp. 1209-1212.

[61] J. Gaintzarain, M. Hermo and M. Navarro: On Learning Conjunctions of Horn$^{\supset}$ Clauses, CiE 2005. New Computational Paradigms. Amsterdam, 2005.

[62] J. Júlvez, J. Cortadella, and M. Kishinevsky. Performance analysis of concurrent systems with early evaluation. In *Proc. International Conf. Computer-Aided Design (ICCAD)*, November 2006. (To appear). Impact of the conference in Citeseer: 1.09

[63] S. Krstic, J. Cortadella, M. Kishinevsky, and J. O'Leary. Synchronous elastic networks. In *International Conference on Formal Methods in Computer-Aided Design (FMCAD)*, November 2006. (To appear). Impact of the conference in Citeseer: 1.25

ELECTRONIC JOURNALS

[64] J. Alvez and P. Lucio: Elimination of Local Variables from Definite Logic Programs, J. Alvez and P. Lucio, *Electr. Notes Theor. Comput. Sci.* 137: 5-24 (2005)

[65] Leen Lambers, Hartmut Ehrig, Fernando Orejas: Efficient Detection of Conflicts in Graph-based Model Transformation. *Electr. Notes Theor. Comput. Sci.* 152: 97-109 (2006).

[66] Hartmut Ehrig, Benjamin Braatz, Markus Klein, Fernando Orejas, Sonia Pérez, Elvira Pino: Object-Oriented Connector-Component Architectures. *Electr. Notes Theor. Comput. Sci.* 141(3): 123-151 (2005).

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

INEVAI3D: Integración de Escenarios Virtuales con Agentes Inteligentes 3D

(Virtual Scenarios Integration using 3D Intelligent Agents)

TIN2004-07926

Dr. Francisco José Perales López
Departament de Matematiques e Informatica de la UIB
EPS, A. Turmeda, C/Valldemossa km 7.5
07122, Palma de Mallorca, Spain
email: paco.perales@uib.es
www.inevai3d.com

Abstract

The present project defines a global and unified framework with intelligent 3D agents for the actual systems. Nowadays electronic communication among persons includes from basic chats and GSM services to virtual immersive scenarios with great realism. The differences are obvious, and virtual immersive scenarios provide mechanisms for interacting virtual elements (realistic and intelligent avatars, information, and passive objects) with the scenarios that participate virtually in a universe.

This is a very wide and ambitious field and it's necessary to define clearly which the issues to develop in this project are. The four important issues to develop can be summarized in: a) Unified Integration of Virtual Scenarios; b) Intelligent Humanoid simulation (Agents); c) Human Motion Analysis using force sensors; d) New Multimodal Interfaces and their application in domotic environments with 3D Agents.

Virtual environments own characteristics that allow the user to select their own point of view and to be represented with an avatar. Although a realistic representation of the virtual environment elements is important we consider avatars behaviour simulation as a very interesting field to develop. Intelligent agents will allow the person to interact in the immersive environment through geometric and behaviour models of the virtual elements of the defined world. Computer interaction is through physical and logical devices; therefore a person human's motion analysis can modify the agents or the elements' behaviour of the virtual scene. Due to this, it is important to consider the inclusion of an avatar analysis and synthesis section using no-invasive techniques.

The aim of this project is to advance one more step forward and to use the previous systems including kinetic models and haptic devices for enhancing the studied techniques. The necessity of a visual and physical immersion has to be a key point in the virtual collaborative environments. In particular, a domotic application on realistic facial modelling & natural interaction would be a key issue to consider. User's expression captures and advanced facial representation of 3D agents that understand those expressions is a new interaction setting in virtual environments. The final objective is a virtual mayordomo 3D agent with realistic appearance and knowledge capabilities.

Keywords: Virtual Reality, Agents, Human Motion Analysis & Synthesis, Facial & Gesture recognition, Perceptual User Interfaces

1 Project Objetives

The INEVAI3D project ("Integración de Escenarios Virtuales con Agentes Inteligentes 3D") defines a global and unified framework with intelligent tri-dimensional agents for actual systems and future virtual environments. The results obtained are applied to a domotic environment. In particular, the project aims to promote a friendly and natural environment at home for the elderly or disabled, by providing them with ICT tools that will allow them to live independently. The end-users are allowed to interact with the assistive environment through the natural senses such as voice, gesture and touch. Combining these basic senses, more sophisticated ways of interactions can be reached to develop the home support system. The INEVAI3D Project were grouped around main areas corresponding to research fields in the topics of computer graphics, computer vision, artificial intelligence, perceptual user interfaces and telecommunications.

We can classify the main objectives proposed in five main points:

1) Develop an unified and integrated framework of virtual reality applications
2) Define a simulation environment for intelligent 3D avatars
3) Study new approaches for human motion analysis and synthesis
4) Define new multimodal interfaces paradigms in domotic environments
5) Develop some prototypes to validate the algorithms designed in real applications

The complexity of the tasks proposed above are very high and some practical restriction must be done to reach the final objective of a real prototype at the end of the project. So the main efforts are concentrating in the points related with multimodal interfaces and human to human communication in a natural way, and improve the intelligence of synthetic avatars in the virtual world. The requirements in support networking (distributed virtual systems) and some human motion kinetics solutions are developed in order to guarantee the final successful of the project.

At the moment, the main objectives reached are related with points 2), 3), 4) and are explained in the next sections.

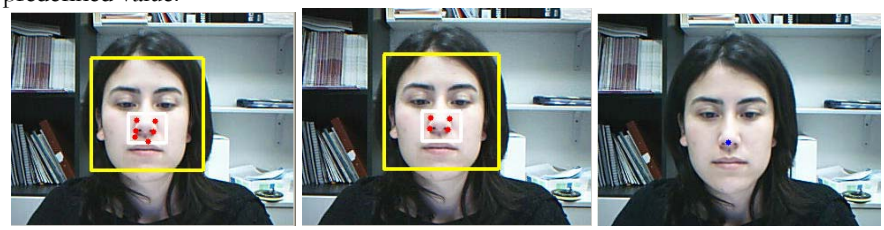
1.1) Real-time Facial Feature Tracking (task 4)

The computer vision algorithm is divided in two steps: initialization and tracking. The initialization step is responsible of learning the user's facial characteristics such as its skin color, its dimensions and their best facial features to track. This process is totally automatic and it can also be used for system's recovering when a severe error occurs, adding the necessary robustness for using it in one human-computer interface.

First, the algorithm detects automatically the user's face by means of a real-time face detection Viola & Jones algorithm. The face will not be considered as found until the user sits steady for a few frames and the face is detected in the image within those frames. A good detection of the features is very important for an effective performance of the whole system and the user must start the process with the so called neutral face: the mouth is closed and the gaze is directed perpendicular to the screen plane, the eyes are open and the eyelids are tangent to the iris. Then, it is possible to define the initial user's face region to start the search of the user's facial features. Based on anthropometrical measurements, the face region can be divided in three sections: eyes and eyebrows, nose, and mouth region. Over the nose region, we look for those points that can be easily tracked, that is, those whose derivative energy perpendicular to the prominent direction is above a threshold. This algorithm theoretically selects the nose corners or the nostrils. However, the ambient lighting can cause the selection of points that are not placed over the desired positions; this fact is clearly visible in Fig. 1 (a). Ideally the desired selected points should be at both sides of the nose and with certain symmetrical conditions. Therefore, an enhancement and a re-selection of the found features must be carried out having in account symmetrical constraints. Fig. 1 (b) shows

the selected features that we consider due to their symmetry respect to the vertical axis. This reselection process will achieve the best features to track and it will contribute to the tracking robustness. Fig. 1 (c) illustrates the final point considered, that is, the mean point of all the final selected features that due to the reselection of points will be centered on the face. Finally, for learning the user's skin color and complete the initialization step, the pixels inside the face region are used as color samples for building a learning set for finding the parameters of a Gaussian model in 3D RGB density using standard maximum likelihood methods.

The tracking step starts finding the *face tracking point*, where the selected best features in the initialization process are used. In this case, the spatial intensity gradient information of the images is used for finding the best image registration. As it was before mentioned, for each frame the mean of all features is computed and it is defined as face tracking point for that frame. The tracking algorithm is robust for handling rotation, scaling and shearing, so the user can move in a more unrestricted way. But again lighting or fast movements can cause the lost or displacement of the features to track. As only the features beneath the nose region are in the region of interest, a feature will be discarded when the length between this feature and the face tracking point is greater than a predefined value.



(a) (b) (c)

Fig. 1. (a) Automatic face detection and initial set of features. (b) Best feature selection using symmetrical constraints. (c) Mean of al features: face tracking point.

In theory, it would be possible to use Kalman filters for smoothing the positions. However, Kalman filters are not suited in our case because they don't achieve good results with erratic movements such as the face motion. Therefore, our smoothing algorithm is based in the motion's tendency of the face tracking point positions (head motion). A linear regression method is applied to a number of tracked face positions through consecutive frames. The computed nose points of n consecutive frames are adjusted to a line, and therefore the nose motion can be carried out over that line direction. For avoiding discontinuities the regression line is adjusted with every new point that arrives.

Once calculated the face tracking point and using the color probability distribution computed in the initialization step, the probability of a face pixel being skin is computed for creating a "skin mask" of the user's face. Using this mask and with the constraint that the face is always facing the screen and not tilted or rotated, the system can detect the user's eyebrows, eyes and mouth bounding boxes and due to their position (right or left side of the face and the relation eyebrows-eyes-mouth) the system can label the zones. One problem can appear if the user has got his eyes a little bit sunken, then due to the shadow in the eyelid, most probably the eyebrow and eye will be found as an only blob. We divide this bounding box assuming that the eyebrow has been detected together with the eye. Finally, from the bounding boxes positions, 10 face features are extracted. These 10 feature points of the face that will later allow us to analyze the evolution of the face parameters (distances and angles) that we wish to study for expression recognition. The aims of Computer Graphics & Vision group of UIB (CGiV-UIB) had to satisfy were:

1) Automatic detection of a group of face features that GIGA (UNIZAR) group needed for the emotion recognition with no need of the user's participation and using computer vision techniques

2) Tracking of the detected face features

3) Interface with GIGA group system (UNIZAR, Zaragoza University), by means of UDP packets.

1.2) An engine for the management and visualization of 3D virtual scenarios and characters. (task 2)

The engine developed, written in C++ and based in a set of open source libraries, is script-directed. The engine manages scene graphs that are built in real-time, dynamically creating and manipulating their elements by means of a simple command interface. These commands are executed via script-files when initiating the application, during execution, or can be introduced through the text console at any moment.

The following elements can be represented in the scene graph:

- Images and texts, that can be shown and positioned as rectangles oriented in space. The main graphic formats (bmp, gif, jpeg, pic, png, rgb, tga, tiff) are allowed and the alpha-channel is supported.
- Simple geometric primitives but also complex geometric models. Maxine imports the most popular formats (3DS, flt, lwo, md2, obj, osg), allowing to create complex virtual scenarios with high level of detail.
- Simple lights & 3D and ambient sound.
- Animated characters and actors. They can be generated with a commercial 3D modelling and animation toolkits. Different types of animations are available, including secondary animations to increase the characters' expressivity and realism. The actors are provided with voice synthesis and facial animation with lip-synch, following the VHML standard and the MPEG4 specification for facial animation.

The system has been developed as a general tool with the aim of making it ease the use of virtual characters as new interfaces in a wide range of applications: as learning tutors, tour guides and ,of course, as domotic helpers.

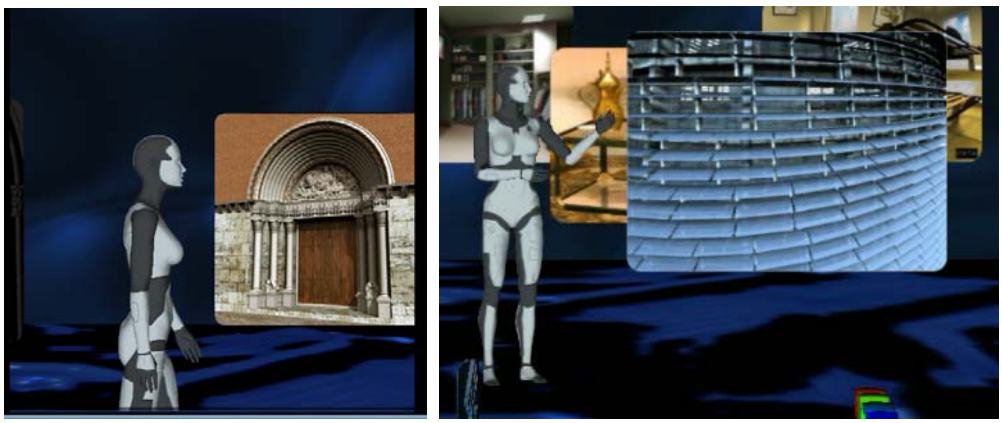


Fig. 2 An example: a virtual character making a presentation in a virtual scenario

Particular emphasis on new emotional and multimodal interaction has been made in the development of the tool:
- It supports interaction with the user through different channels: text, voice, mouse/keyboard, which opens up the use of the applications generated to a wide range of users.
- The discourse is reinforced with referential acts (pointing gestures and/or looking at the objects that are being presented)
- The voice interface enables communication in natural language. The system has been adapted to enable conversations with the character to be held in Spanish, as all the available systems involve conversation in English.
- The system endeavours to encourage understanding between the user and the 3D character: The virtual character's facial expressions, answers to questions and voice (what we refer to as emotional voice) are adapted to the emotional state of the user. In fact, it has its own emotional state, which is subject to modification, depending on its interaction with the user.

Particular effort has been invested in collecting additional information about the user to enhance user/virtual character interaction. In particular, in the image-based estimation of the user's emotional state, so that human interaction becomes more natural, enjoyable and productive: the computer could offer help and assistance to a confused user or try to cheer up a frustrated user, and hence react in more appropriate ways. A simple and effective system for the real-time recognition of facial expressions has been developed. In opposition to other systems that rely on the use of wearable detectors, the system developed in non-invasive and is based on the use of a simple low cost webcam. The automatic features extraction program allows the introduction of dynamic information in the classification system, making it possible the study of the time evolution of the evaluated parameters, and the classification of user's emotions from live video. The feature extraction program has been developed by the CGiV-UIB group (see section 1.1), and captures each facial frame and extracts the 10 feature points which are sent to the emotion classifier. The classifier, developed by the GIGA Zaragoza's group classifies user emotion in the six universal emotional categories (sadness, anger, disgust, joy, fear, surprise) plus the neutral one. When an emotional change is detected an emotion code is sent to the virtual character. In the example shown in the figure, the virtual character's face just mimics the emotional state of the user, accommodating his/her facial animation and speech.

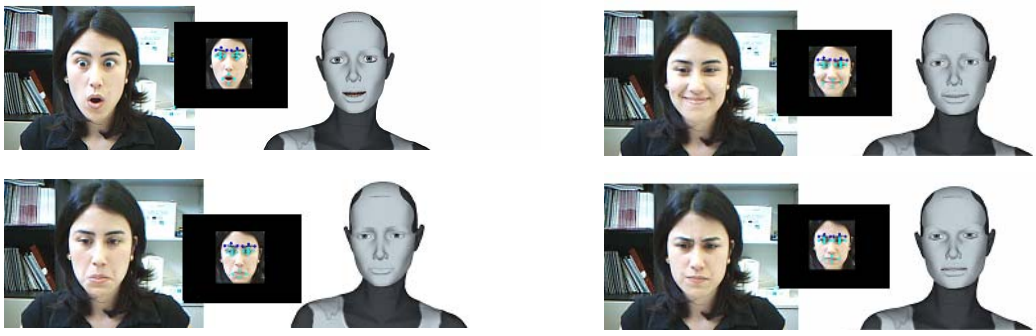


Fig 3. Examples of the integrated real-time application: detection of surprise, joy, sadness, anger.

1.3) Parameterized and Standardization Facial Animation Module (task 3)

Also we have designed an independent facial animation module compliant with actual standard protocols. We started with the analysis of the techniques and algorithms used by other researchers to create parameterized animations. We dedicated special attention to those projects that works with MPEG-4 standard, because it provides an alternative way to model facial expressions and emotions using parameters (feature points FDP, and facial animation parameters FAP). The facial engine selected for this project was Xface, which is a set of open source tools for creation of MPEG-4 and key-frame based 3D talking heads, initiated by Cognitive and Communication Technologies (TCC) division of ITC-irst.

Xface consists of a core library with all the functions needed to manage and animate 3D models, an editor XfaceEd which is used to generate MPEG-4 meshes from 3D models, the XfacePlayer to play the animations and XfaceClient used for control remote over TCP/IP. It is developed in C++ and uses additional libraries to create interface windows, text-to-speech and XML files handling. The animations are generated modifying or creating files that contain the number of frames, the FDP affected and the values in which they are affected. These files are loaded in Xface, read and played using the player or the editor.

So far, we have created FAP files in order to represent basic emotions (happiness, sadness, anger, fear, disgust, and surprise). We have obtained satisfactory results, although a lot of work is involved in the creation of the files. That is why we modified the interface of the editor to be able to change the values of the file interactively. New text boxes have been added with the names and values of the FAPs in each frame, and with a slider these values are modify, therefore the facial expression is modified at the same time.

As we want to develop a realistic face model, we have to manage more expressions representing a wider number of emotions. To achieve this we worked with values of personality traits (openness, conscientiousness, extraversion, sympathy, and neurosis). We are using an emotion and personality engine that provides the mechanism to generate and evaluate emotions. Depending on the values of each trait the resultant emotions will be more or less intense, belonging to one of three groups: high intensity, medium intensity, and low intensity. According to the group to which the emotion considered belongs, the correspondent FAP file will be played.

To represent emotions that are not the basic ones, we are using a technique proposed by Raouzaiou et al., where each emotion has an associated group of FAPs, and each FAP is between some ranges. Each of these groups is call *profile*. Using the profiles of the basic emotions, we are obtaining profiles for new emotions using the calculations and rules proposed by Raouzaiou et al. The results have been expressions close to the final goal, as well as expressions that have needed lots of refinement to be believable.

The use of FAP files avoid the use of facial animation tables (FAT), because they already specify the frames of the animation, and the FAPs affected.

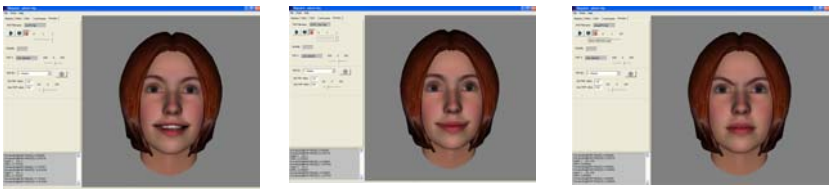


Figure 4: Two Joy expressions, Anger expression using profiles.

In the area of artificial intelligence, we have worked in the conceptual design of a multiagent system, which will be implemented to control the domotics house. The result was a system that divides the house in modules, according to the needs of the user (house accommodation, security, electrodomestic articles, health, activities and preferences, nursing). Each module is controlled by an intelligent agent, who interacts with the butler agent (represented by the 3D face) in order to know the necessities of the user and to inform about the state of the module that is in charged of. This part will be explained in next section

1.4) 3D Intelligent Avatar (Agents) (Task 2)

Our task in the project consists in developing an agent, the butler, capable of an intelligent management of the information. The designed system must be adaptive, with the ability to react in real time, to learn, remember and evolve with the experience and to decide in specific situations in an autonomous way using the external inputs given by the end-user and the home domotic devices. With these requirements in mind, the agent systems represent a good solution to solve the artificial intelligence problems of complex environments and with multiple inputs. The butler receives the corresponding perceptions from the sensors in the domotic house and from the elderly (by means of voice, gestures or commands). Once all this knowledge is collected the butler must act consequently. The butler's behaviour, its decision, depends on large amounts of data: the totality of perceptions together with the evolving model of the environment. All this data can be represented by means of information systems. Our objective in the INEVAI3D project is the design and implementation of an intelligent agent, the butler that cares for the elderly or disabled. The butler's behaviour is reactive, based on the continuous interaction that it maintains with the elderly person and the domotic house. Furthermore, this behaviour must be intelligent enough to be considered as a rational behaviour, allowing the butler to attain the objectives in a reasonable manner.
Currently, we are involved in the development of an intelligent butler prototype in a domotic environment. In order to fulfill this purpose the following steps have been carried out:
The problem's specification (domotic home, butler, elderly …) has been studied and decided.
An appropriate framework of the requirements specification has been considered in order to develop a prototype centered on the intelligent butler.
The computer platform (hardware and software) to implement the prototype has been determined.
At this moment, the domotic visual environment and the butler's behavior are being modeled and implemented.

1.5) Integrated framework of virtual reality applications (task 1)

The three groups of UIB & UNIZAR, UPNA has been working on the issue of using high-level languages and suitable technologies to model and to represent virtual worlds, taking into account the need of sharing them through the network. We have considered the possibility of using a database to store the state of a virtual world, with the intention of sharing the data for different users or different applications. In the same way we have considered the possibility of using different forms to interact with the world. In our case we have been working with a platform of VR based on VRJuggler open source API's. Finally we have considered the possibility of using different devices (PC, PDA, etc….) to use the application (either directly or through the BD).Next we give some details of the different parts and their current state of development.

1.5.1) STEDEL, VRML and X3D as 3D modelling languages for VR scenes

We are using VR technology for a) to design, represent & visualize the current state of a domotic house and b) to control and interact with the domotic goods within the house through INTERNET. VRML is a standard language that allows to create 3D scenes and to interact with them in an intuitive way similar to the reality. Anyway in order to design a scene with this language

it is necessary to acquire complex knowledge that everybody may do not know (for example an end user that needs to design an application based on VRML). This is the reason for which other methods of creation of three-dimensional scenes have been investigated; STEDEL is a spatio-temporal language that allows designing a three-dimensional scene by means of basic slight knowledge. With the increasing of the new technologies as the semantic Web or the huge use of data bases in any application and the lack of tools in VRML to manage such technologies, Web3D has developed a new standard X3D. Although X3D is based on VRML, it is ready to be used with such new technologies. We are designing a virtual domotic house in VRML-X3D where the user will be able to control different real domotic components. The objective of this virtual domotic house is that people with some disability could interact with them in some way. We are making special emphasis in the design of icons as well as friendly and intuitive interfaces that helps the interaction with the virtual world. This virtual world will be connected to a Data Base.

1.5.2) Virtual worlds in JAVA 3D: design & interaction

Java3D allows the interaction with the created Virtual World by means of the standards inputs of the computer, the keyboard and the mouse. A 3D world in JAVA has been created that represents again a domotic house. The application is supported by a MySQL Data Base hosted on the same server where the program "ServerRMI" is (responsible for listening requests from the clients in order to pass to them necessary stuff to execute the Virtual World). This Data Base will be the one in charge "to remember" the different changes happened on the objects from the Virtual World.

1.5.3) VRJuggler

We carried out the installation, configuration and starting of a low cost virtual reality platform. The installation of VRJUGGLER has been made for both operating systems, Windows and Linux; in both cases we have a description of the platform and how it has been implemented. In order to solve all the problems happened within this work; we got in touch with user developers of such technologies on the net. We have checked how the different peripherals are working; it was OK for the "flocks" (6 DOF) and for the glasses (with stereo vision). But gloves are still not working properly.

1.5.4) Consulting and updating a Data Base of virtual world from different devices

The application shows the map of a house, and the position of its domotic components. The user can open to the menu with the information of each component doing click on it. He can change the information of the component, according to the characteristics of each one. He cannot change the position of the component nor add a component (one assumes that this only can be done by the staff that installs the components in the house). The scene is stored on the database and its information is updated every 10 minutes. We are working on a similar application to consult the same Data Base using a PDA (at the moment we have no results).

1.5.5) JESP (Joint Editing Service Platform)

Finally we are exploring a more sophisticated solution based on JESP. The INEVAI3D is layered on top of an IP compatible collaborative support platform (JESP) that has adopted a distributed architecture following the peer-to-peer network communication model. This module offers group communication and session management to the INEVAI3D. It provides a set of communication services, which hide the point-to-multipoint configuration from the applications, ensuring source order delivery of multicast messages over TCP/IP. The session control mechanisms provided by the platform free the applications from specific functions necessary for their inclusion into co-operative environments, such as, communication services multiplexing, support of quality of service, consistency control, admission of new members into a running session, managing early members leave, invocation of new distributed applications, handling of

exception events or failures and definition of roles within the group. A special protocol, Mu3D (Multi-user 3D Protocol), was developed to ensure consistency and mutual exclusion of access to the shared scene across all replicas, using an innovative technique, that ensures total global order of exchanged messages between distributed peers in cooperation. The Mu3D fields are designed to allow specifying the user events in small messages. The current message size, counting JESP and TCP headers varies between 150 and 220 bytes.

1.5.6) Network testing of Virtual World Applications

The Centre of Information Technologies of the UIB, inside the project INEVAI3D, is the responsible in the study of the best solution for the creation of a telecommunications net which must respond to the requirements and necessities of the project. The project started by establishing a first scenario and the needed requirements for transmitting the messages through the net among the different nodes that are participating in the project. It is also presented the different technological options available nowadays that could be used for creating other communication scenarios. For the elaboration of this work measures from different parameters of capability and performance have been carried out, among them we highlight the following ones: band width, delay, delay variation and availability. Net tests have been done only with the University of Zaragoza (UNIZAR). Among the obtained results, the most important ones are:

- The band width of the existent path is sufficient for the transmission of the messages between the INEVAI3d nodes
- The measured delays comply with the project's requirements
- The delay's variation it not excessive and it can be considered as acceptable
- The net has had a very high availability. The scenario has behaved in a very stable mode during the measurement period.
- It is still necessary to study the behaviour using a secure transmission mechanism by means of paths in virtual private nets.

1.6) Others techniques and modules developed (task 3 & 5)

Finally we would like two remarks that specific modules are developed in order to guarantee the whole project. So the UML class specification has been done for the house mayordomo application in collaboration with all the members of the project. Specific use cases are described to define exhaustively the functionality of the final prototype application.

Also a facial aging engine independent module (called VISAGE) has been developed and tested to reach more realism in the 3D avatars. Also we have started strong contact with the enterprises (EPTRON and TAGRV) involved in the project in order to promote the real transference of the prototype mayordomo application in a real domotic environment at the end of the project. The task of dissemination is done by the periodic internationals publications and website actualizations with video demos and trial versions (www.inevai3d.com).

2 Level of Success

In order to assure the main objectives of the project, a full strong collaboration between partners has been done. So the main initial tasks were the UML specification of the application prototype (called virtual mayordomo) and after that we have included the functionality needed in this prototype in coherence with the original projects objectives defined at beginning of the paper.

We have organized periodical working sessions with all the partners in order to solve the main problems and establish the standards protocols communication between shared modules and applications.

The project is organized in several subtasks and periods sections. In the next table we can summarize the evolution of the project.

Activity	Partner	Year 1	Year2	Year3	Actual State
T0: Web Site Maintenance and Project Coordination	UIB	full	full	full	On time
T1: Virtual Unified Scenarios Integration	UIB	full	full		Some delay (6 month)
T2: Intelligent 3D Avatars	UIB, UNIZAR	2º half	full	full	On time
T3: Human Motion Analysis	UIB, ULBRA	2º half	full	1º half	On time
T4: Multimodal Interfaces	UIB, UNIZAR		full	1º half	On time
T5: UML Design	UIB	full	full	full	On time
T6: Modular Integration & Prototype	UIB,EPTRON, TAGRV			full	On time

From the previous sections and the time table presented, we can conclude that we are on time in all the main tasks proposed as objectives in the project. In the task 0 & 5 the project is actively the three years, so continuous contributions are added in order to satisfy the requirements of applications. The task T2 & T4 are reaching the basics objectives proposed with some trial versions of an intelligent agent and some multimodal VBI systems, of course the final integration is not yet done. The facial computer animations modules are developed. At the moment we are increasing the contacts with enterprises to define carefully the final prototype commercialization if is possible (task 6). To be honest some delay is detected in section T1, we have defined the communications protocols between the distributed systems, high-level languages and suitable technologies to model and to represent virtual worlds have been studied and tested, and some low level QoS (quality of Service) network testing has been done, but not yet is completely define the conceptual model and virtual object portability between systems specially in portable devices. We need to do more efforts in this section in the next months.
Also in the T3 task we have some restrictions about haptics devices, because financial funding for the original proposed haptic system was denied, so we are trying to reorganized and search a low cost haptic system or equivalent platform devices. In any case, the 3D facial animation subtask is on time.

3 Results

During the last two years, several national and international publications have been presented in important newsletters by the partners of the project. In the reference section, we present the most relevant contributions and only the published work at the moment (many under revision papers are at this time). Also the UIB has been organized an International workshop (AMDO2006) to promote the topics of the projects and this event is today an international worldwide reference conference in the areas of the project. The big numbers are: 27 papers distributed in 9 Journals and 18 Conferences Proceedings. I would like to remark the international price reached by the work done in cooperation with Dr. Ronan Boulic of EPFL with title: Real-Time IK Body Movement Recovery from Partial Vision Input**",** in Proceedings of the 2nd International Conference on Enactive Interfaces (ENACTIVE05), as the *1st ENACTIVE Best Collaborative*

Paper Award. In relation of this work the Dr. Ronan Boulic propose to UIB-Computer Graphics & Vision group to be invited member of NoE ENACTIVE. In this time also several Ph. D. Thesis have been developed by the member of the project in particular Dr. Jose Maria Buades Rubio from UIB and Dr. Alex Cruz Berg from ULBRA (Brazil). Actually are four theses under process and many students' projects.

In the last year, we have developed and extension proposal of INEVAI3D project as a European Project named AALIVE (Last Call of VI UE program), where the coordinator was the UIB. The proposal was not accepted, but the score and general feedback was very positive, so this year will be presented again after some improvements to the VII European program. In the local environment (Balearic Islands), the partial results of the projects (facial mouse VBI) has been translate to UIB students and scholar activities in fairs and in near future to all computer rooms at undergraduate school for disable o handicapped students in our community in Balearic Islands.

4 References

[1] D. Arellano Távara, A. Meier. Agents with Personality for Videogames. IV Conference of Articulated Motion and Deformable Objects (AMDO 2006), pp.484-493, Palma de Mallorca, Spain, 2006.

[2] F. Perales et al., E. Cerezo et al, A. Pina et al. El proyecto INEVAI 3D: Agentes Autónomos 3D, Escenarios virtuales e Interfaces Inteligentes para aplicaciones de Domótica y de Realidad VirtualCEDI 2005 (UCAMI), Granada, Septiembre 2005.

[3] I. Azkue, A. Pina, M. Vazirgiannis. Towards an Integrated Technological Framework for Modelling Shared Virtual Spaces: Languages and Domotic Applications. 474-483. IV Conference of Articulated Motion and Deformable Objects, 4th International Conference, AMDO 2006, Port d'Andratx, Mallorca, Spain, July 11-14, 2006, Proceedings. Lecture Notes in Computer Science, Springer 2006

[4] R. Boulic, J. Varona, L. Unzueta, M. Peinado, A. Suescun, F.J. Perales. "Evaluation of on-line analytic and numeric inverse kinematics approaches driven by partial vision input". Virtual Reality 10(1): 48-61, ISSN: 1359-4338, 2006.

[5] Sistema MOBiL. Deformaciones musculares en la locomoción humana. Congreso Español Informática Gráfica 2005, Granada, España, 13-16 September 2005, Proc CEIG'05, J. Regincós, D. Martín (eds.), ISBN: 84-9732-431-5, pp.159-168

[6] "Emotional Facial Expression Classification For Multimodal Interfaces", IV Conference On Articulated Motion And Deformable Objects (Amdo 2006) Andratx (Mallorca, España)12-15 July 2006 Lecture Notes in Computer Science, vol 4069, ISBN-10 3-540-36031-X , pp. 405-413. (Springer)

[7] "Individualizing The New Interfaces: Extraction Of User's Emotions From Facial Data" Ibero-American Symposium on Computer Graphics (SIACG 2006), Santiago de Compostela (España), 5-7 July 2006, Proc. SIACG 2006, edits. P. Brunet, R. Correia, G. Baranoski. , ISBN: 3-905673-60-6, pp.179-185.

[8] "Maxine PPT: Using 3D Virtual Characters for Natural Interaction", 2nd International Workshop on Ubiquitous Computing and Ambient Intelligence, Puertollano (Ciudad Real, España), 15 November 2006, ISBN: 84-6901744-6, pp.241-250.

[9] Miró-Julià, M. Degenerate Arrays Cast and Tools for Robotics, Vehicular and Communication Systems. Alexis Quesada-Arencibia, R. Moreno-Díaz jr. J. Rodríguez (Eds.). ISBN:84-689-0432-5 pp. 7-11. (2005)

[10] Miró-Julià, M.Degenerate Arrays: a Framework for Uncertain Data Tables. Lecture Notes in Computer Science. R. Moreno Díaz and F. Pichler (Eds.) Computer Aided Systems Theory - EUROCAST 2005. Springer-Verlag.Volumen: 3643, pp. 21-26 Berlin Heidelberg

[11] Miró-Julià, M.Declarative Expressions: a New Perspective for Handling Information Systems. Artificial Intelligence and Soft Computing, ISBN: 0-88986-612-0, pp79-84, 2006.

[12] Miró-Julià, M.; Fiol Roig, G.Information Systems and Intelligent Agents. Current Research in Information Sciences and Technologies. Multidinary approaches to global information systems ISBN: 84-611-3105-3 Depósito legal: BA-564-06, V2, pp. 243-247, 2006

[13] A. Jaume-i-Capó, J. Varona, F.J. Perales. "Interactive applications driven by human gestures", in SIACG 2006 Ibero-American Symposium on Computer Graphics, Santiago de Compostela, Spain. July 2006.

[14] A. Jaume-i-Capó, J. Varona, M. González-Hidalgo, R. Mas, F.J. Perales. "Automatic Human Body Modeling for Vision-Based Motion Capture", in Short Communication Proceedings of the 14th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG'06), Plzen, Czech Republic. February 2006.

[15] C. Muñoz, D. Arellano, F. J. Perales, G. Fontanet. "Perceptual and Intelligent Domotic Systems for Disabled People", in IASTED International Conference on Visualization, Imaging and Image Processing(VIIP 2006), Palma de Mallorca, Spain. August 2006.

[16] C. Manresa-Yee, J. Varona, F.J. Perales. "Towards Hands-Free Interfaces Based on Real-Time Robust Facial Gesture Recognition". Lecture Notes in Computer Science 4069: 504-513, ISSN: 0302-9743, 2006.

[17] C. Manresa-Yee, J. Varona, T. Ribot, F.J. Perales. "Non-verbal communication by means of head tracking", in SIACG 2006 Ibero-American Symposium on Computer Graphics, Santiago de Compostela, Spain. July 2006.

[18] C. Manresa-Yee, J. Varona, F.J. Perales. "Face-Based Perceptual Interface for Computer-Human interaction", 14th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG'06), Plzen, Czech Republic. February 2006.

[19] J.M. Buades, F.J. Perales, J. Varona. "Matching for Perceptual User Interface", in Poster Proceedings of the 14th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG'06), Plzen, Czech Republic. February 2006.

[20] J.M. Buades, F.J. Perales, M. González-Hidalgo, J. Varona. "Upper Body Tracking for Interactive Applications". Lecture Notes in Computer Science 4069: 514-523, ISSN: 0302-9743, 2006.

[21]A. Cruz Berg, F.J. Perales, M. Gonzalez, "A facial Aging Simulation Method using flaccidity deformation criteria. in 10th International Conference on Information Visualisation, London, England. 14-16 July 2006. pp. 791-796, ISBN-ISSN 1550-6037, 0-7695-2602.

[22] C. Manresa-Yee, J. Varona, R. Mas, F.J. Perales. "Hand Tracking and Gesture Recognition for Human-Computer Interaction". Electronic Letters on Computer Vision and Image Analysis 5(3):96-104, ISSN: 1577-5097, 2005.

[23] J. Varona, J.M. Buades, F.J. Perales. "Hands and face tracking for VR applications". Computers & Graphics 29(2):179–187, ISSN: 0097-8493, 2005.

[24] R. Boulic, J. Varona, B. Herbelin, L. Unzueta, A. Suescun, A. Jaume-i-Capó, F. Perales, D. Thalmann. "Vision-Based Comparative Study of IK Techniques for Arm Movements" in First International Enactive Workshop, Pisa, Italy. March 2005.

[25] R. Boulic, J. Varona, L. Unzueta, M. Peinado, A. Suescun, F.J. Perales. "Real-Time IK Body Movement Recovery from Partial Vision Input", in Proceedings of the 2nd International Conference on Enactive Interfaces (ENACTIVE05), Genova, Italy. November 2005. *1st ENACTIVE Best Collaborative Paper Award.*

[26] Paola Britos, María José Abásolo, Ricardo García Martínez, Francisco Perales. Identification of MPEG-4 FDP Patterns in Human Faces using Data-Mining Techniques. WSCG'2005 - The 13-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, UNION Agency Pr, ISBN: 80-903100-8-7. Plzen (REPÚBLICA CHECA), 2005

[27] Low Cost Avatars Animation System from Real Images Compliant MPEG4, B. Mir, A. Salas, A. Jaume. M. Bez, F. Perales, The 13-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, UNION Agency Pr, ISBN: 80-903100-8-7. Plzen (REPÚBLICA CHECA), 2005

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

jspTIN-2007: Multi-valued Logic: Foundations and application to management of vagueness and imprecision TIN2004-07933-C03

Francesc Esteva *
Artificial Intelligence Research Institute (IIIA)
Spanish national Research Council (CSIC)

Ventura Verdú †
Dept Prob. Logic and Statitistics
University of Barcelona (UB)

Cèsar Fernández ‡
Computer Science Department
University of Lleida (UdL)

Abstract

The project had three basic objectives. The first one was the logical and algebraic study of t-norm based many-valued logics in the frame of both residuated and substructural logics. The second one was the formalization, within the framework of the above logics, of several deductive soft computing mechanisms, based on fuzzy logic, to deal with vagueness, uncertainty and imprecision. The last one was to continue theoretical and experimental research on many-valued SAT and Max-SAT algorithms and their application to solve computationally difficult problems. We have deeply worked on the three goals and obtained a significant number of results, which we summarize in this report and which have been published in several conferences and journal papers.
Keywords: Multivalued Logics, t-norm based Logics, Substructural Logics, Multivalued SAT and Max-SAT, Soft Constraints

1 Project goals

The objective of the project is the study of multi-valued logics and their computational applications. Mainly we have proposed to study both infinitely-valued ones (fuzzy logics valued over [0,1]) and their algebraic semantics and finitely-valued ones typically used in SAT problems. The basic assumption of the first goals of the proposal was that residuated many-valued (fuzzy) logics, (what was known as t-norm based -fuzzy- logics) are a good tool to formalize vagueness and imprecision. From this assumption, the project goals described in the proposal are the following ones:

*Email: esteva@iiia.csic.es
†Email: v.verdu@ub.edu
‡Email: cesar@diei.udl.es

1 .- *Logical and algebraic study of t-norm based logics.* MULOG project proposes the study of quasi-varieties of BL and varieties of MTL and IMTL.

2 .- *Study of t-norm based logics from the substructural point of view.* Being t-norm based logics nice and interesting examples of logics without contraction, MULOG project proposes to study these logics from the point of view of substructural logics, mainly starting from the monoidal structure of the corresponding algebras. In the proposal we say that perhaps their study would need more than Hilbert and Gentzen formalisms (axiomatic and Gentzen systems) and we would need to use Avron's formalism (mainly hypersequents).

3 .- *Free algebras and functional representation of substructural and t-norm based logics.* The goal was to study free algebras in order to describe the structure of the algebras of propositional formulas. Moreover we are interested in functional completeness theorems like Mc Naughton for the infinite valued Lukasiewicz logic.

4 .- *Formalization of some Soft Computing techniques in the framework of t-norm based logics.* Here the proposed goals were:
i) To study the logic obtained by adding truth constants to the logic of a t-norm and its residuum (specially in the case of a continuous t-norm), with special attention to evaluated formulas, the kind formulas used in many fuzzy knowledge-based systems or in fuzzy logic programming.
ii) Similarity based reasoning. Special attention is paid to both coherence problems and interpolation in fuzzy rule based systems.
iii) To model Soft-CSP problems using the language of the extended possibilistic logic with fuzzy constants.

The basic assumption of the last two goals is that many-valued clausal forms provide a suitable formalism for representing and solving decision problems via their reduction to SAT and optimization problems via their reduction to Max-SAT:

5 .- *Design, implementation and evaluation of SAT algorithms for multivalued clausal forms and their application to solve computationally difficult problems.* The goal was to incorporate in our algorithms both the recent techniques introduced in Boolean SAT, as well as new multivalued branching rules and variable selection heuristics that take into account the structural information provided by the variable domains.

6 .- *Design, implementation and evaluation of Max-SAT algorithms for multivalued clausal forms and their application to solve problem with soft constraints.* The project proposes to work on this topic, which was unexplored when we wrote the proposal of the project.

Eventhough there is a "crono-gram" in the proposal we think that the work in these general goals is a continuous task during the three years of the project and thus we will present in the next section the results we have achieved at the end of the second year (December 2006).

2 Project results

We next describe the results obtained in the development of the project for each one of the goals described in the previous section.

1 .- *Logical and algebraic study of t-norm based logics.* The results obtained in this goal are mainly referred to varieties of MTL and IMTL and only someones related to varieties (not yet quasi-varieties) of BL. We have published the two papers [EGM04] and [DEGM05] about subvarieties of BL. The first proves that the logic of a continuous t-norm and its residuum is finitely axiomatizable and gives an algorithm to find a finite system of axioms for each continuous t-norm, result that is used posteriorly in the study of the logic obtaining by adding truth constants (see results of goal 4). The second is a extended survey of known results (and some news ones) about subvarieties of BL in a special issue of a journal edited by two of the members of the IIIA team in the MULOG project. In the MTL and IMTL cases we have published the four journal papers [GNE05, NEG05b, NEG05a, Gis03, GT05, MNH06]. This papers describe a large amount of subvarieties of MTL and IMTL with their axiomatizations and standard completeness results. Moreover, they present different methods to try to obtain a complete description of MTL and IMTL subvarieties and even thou this goal is not completely obtained the results are relevant. Moreover we have also presented several communications to conferences (among them we cite [nNE04, NEJ05, NEJ06]) and have been invited to give general talks in the EUSFLAT'06 conference (the conference of the European Society for Fuzzy Logic and Technology) [Est05a] and in the Seminar on Fuzzy Set Theory in Linz [Est05b].

2 - *Study of t-norm based logics from the substructural point of view.* The results concerning this section are contained in the papers [VGCB04a, VGCB04b, VGCB06a, VGCB06b, TVFG06]. T-norm based logics are truth-preserving logics and they can also be seen as substructural logics because they are extensions of the Intuitionistic logic without contraction (or monoidal logic) ML. The study of ML is very useful to obtain general properties of t-norm based logics. We have proved in [VGCB06a] that the fragment without implication and without negation is exactly the corresponding fragment of classical logic. As a consequence the same situation happens for every t-norm based logic, in other words when we do not deal with the connectives of implication and negation the behaviour of t-norm based logics is exactly the same as classical logic. In particular for every t-norm based logic, its fusion-fragment and its conjunction-fragment are equal. This kind of pathological behaviour will be analyzed in future papers and is mainly due to the fact that these logics are truth-preserving. In [VGCB04b, VGCB06b] we have studied the fragment without implication and with negation of the substructural logic ML. This fragment is strictly included in the corresponding fragment of Intuitionistic logic (and so in the corresponding fragment of classical logic). In [TVFG06] we have started the study of t-norm based logics preserving degrees of truth. Preserving degrees of truth means that the degree of truth of a consequence of a finite set of formulas is greater or equal than the degree of truth of every one of these formulas. We obtain a Gentzen-style and a Tarski-style axiomatization for the infinite-valued Łukasiewicz that preserves degrees of truth. We hope that we will obtain a Hilbert-style axiomatization for this logic in a near future.

3 .- *Free algebras and functional representation of substructural and t-norm based logics.* The results obtained in this goal are descriptions of free algebras of varieties of algebraic structures related with MTL-logic. In [TC06] we give a complete description of free algebras in varieties of some special cases of BP_0-algebras, concretely the MTL-algebras

with Glivenko property and boolean retract. In[TV] we show that, in general, the variety of all Łukasiewicz implication algebras, corresponding with the implicative fragment of the logic given by the Łukasiewicz t-norm, has its free algebras directly indecomposable. In [GT07] we obtain that the free algebra of bounded BCK-algebras, corresponding with the implicational and negational fragment of Monoidal logic, is semi-simple; from this result we also show (with some difficulty) that the variety generated by bounded BCK-algebras is generated by its simple finite members. To obtain the description of the free algebras in varieties of MTL-algebras we use boolean representation, method that cannot be used in Łukasiewicz implication algebras, which need a more direct description. To analyze free bounded BCK-algebras we refer to some results on residuated lattices.

4 .- *Formalization of some Soft Computing techniques in the framework of t-norm based logics.* Most relevant results in this goal are devoted to item i), the study of the logics obtained by adding truth constants to the logic of a t-norm and its residuum. We have published three journal papers [SCE$^+$06, EGN06, EGGN] and a chapter of a book [EGN]. The two first papers contain full completeness results for Product and Gödel logics with any set of truth constants associated to a countable subalgebra of the corresponding standard algebra. The last journal paper contains general completeness results for any logic of a continuous t-norm and its residuum by adding any set of truth constants associated to a countable subalgebra of the standard algebra. In both papers special attention is paid to the case of evaluated formulas, formulas of type $\overline{r} \to \varphi$ where $\overline{r}$ is a constant an φ is a formula without constants. Finally the book paper contains a survey of all results plus new completeness results for the logic of the nilpotent minimum t-norm (which is not continuous) and its residuum with truth constants. Moreover we have presented several communications to conferences and workshops, among them we cite [EGN04, CEGN04]) and we note that members of the IIIA team have been invited to give plenary talks in the conferences "The logic of Soft Computing and Workshop of the ERCIM working group on Soft Computing" in 2005 (Ostrava, Czech Republic) [Est05c] and in 2006 (Málaga, Spain) [God06].

With respect to item ii) we have presented the communications [ERBMD04, BMEG$^+$05] and the chapter book [AEGT04]. In the communications we have explored new formal tools for interpolation based on similarity and in the book chapter we have given a method to obtain similarities relations in fuzzy case-based reasoning. Moreover we have to say that we have not worked in Soft-CSP mainly because we have taken advantage of the fact that Prof. Chesnevar (from Universidad Nacional del Sur, Bahia Blanca, Argentina) obtained a 5-year contract in the Ramon y Cajal program in Lleida University. Since he is a reputed researcher in argumentation, we have worked in the extension of argumentation systems to deal with possibilistic uncertainty and in this subject we have published several joint papers [CSAG04, CSAG05, CnSG05, CSGA05, ACG$^+$06].

We have also published some papers on fuzzy modal formalizations of probability logics [GM06, FG], possibilistic decision theory [GZ05, GZ06] and some on graded BDI architectures for agents [CGS05, CGS06]. These papers and the papers on argumentation have in common the management of uncertainty in the studied systems extending the "classical systems" using mainly classical logic.

Finally we have collaborated in a book chapter [DEGP05] where a general view about

what is vagueness and how to manage it in different scenarios from the point of view of information systems is given. We want to stress the interest of this paper since it has been published in a philosophical oriented book where vagueness is studied from a different point of view and with different formalizations. The discussion there has been very interesting and fruitful.

5 .- *SAT algorithms for multivalued clausal forms and their application to solve computationally difficult problems.* The results obtained about the SAT problem of many-valued clausal forms have been published in [AM04, AM05a, ABCM04, AdVD$^+$04, BMC$^+$07]. In [AM04, AM05a] we have defined mappings that translate many-valued clausal forms into Boolean clausal forms, which can be used to solve many-valued SAT instances with Boolean SAT solver as well as to define more efficient Boolean SAT encodings of combinatorial problems with many-valued variables. In [ABCM04] we have given new complexity results for some classes of many-valued SAT problems and analysed experimentally the frontier between P and NP many-valued SAT problems. In [AdVD$^+$04] we have defined a variable selection heuristic for solving quasigroup completion problems with many-valued and CSP solvers that outperforms all the existing heuristics. In [BMC$^+$07] we have described both complete and local search solvers for the subclass of many-valued clausal forms known as Regular CNF formulas, and proved that encoding combinatorial problems as instances of Regular SAT and solving them with a Regular SAT solver can accelerate the search for a solution. We have also defined a generator of computationally difficult sudokus encoded as SAT instances [ABF$^+$06].

6 .- *Max-SAT algorithms for multivalued clausal forms and their application to solve problem with soft constraints..* The research on many-valued Max-SAT has also produced relevant results on Boolean Max-SAT and Weighted CSP. A complete inference system for Many-Valued Max-SAT and its relationship with the soft local consistencies defined for Weighted CSP is described in [ABLM07], and the first solver we have developed for Many-Valued Max-SAT is detailed in [ADL$^+$06]. We have also designed and implemented two solvers for Boolean Max-SAT: MaxSatz and Lazy [AMP04, AMP05]. MaxSatz won the First Max-SAT Evaluation, a co-located event of SAT-2006. It is a branch and bound solver that incorporates the lower bound computation methods defined in [LMP05, LMP06] and refinements of the complete inference resolution rule for Max-SAT defined in [BLM06]. Finally, we have defined a clausal form formalism [AM05b, AM06] and the corresponding Max-SAT solvers for dealing with problems with hard and soft constraints.

3 Quality indicators of the project achievements

In the framework of the project six new researchers have been formed: C. Noguera has obtained his Ph.D. in the University of Barcelona in June 2006 (supervisors: F. Esteva(IIIA) and J. Gispert (UB)), E. Marchioni has obtained his Ph.D. in the University of Salamanca in November 2006 (supervisors: M. Manzano (US) and L. Godo (IIIA)), C. Ansótegui has obtained his Ph.D. in the University of Lleida in July 2004 (supervisor: F. Manyà (UdL)), A. Garcia-Cerdaña will defend his Ph.D. at the beginning of 2007 (supervisors: V. Verdú and F. Esteva), J. Planes will defend his Ph.D. at the beginning of 2007 (supervisors: F. Manyà

(UdL) and Chu-Min Li (Picardie)), and J. Argelich plans to defend his Ph.D. at the end of 2008 (supervisor: F. Manyà (UdL)). Moreover C.Noguera has obtained the students Galois award of the Catalan Mathematical Society in 2005.

A second indicator is the coauthors of papers that is a deep indicator of the relationships that the members of the team have with well recognized foreign researchers. In fact the MULOG researchers have consolidated relations (many of them supported by bilateral projects[1]) with Prof. Hàjek, Novak, Cintula and Horcick (Czecz Republik (bp)), the Italian researchers Prof. Montagna (Siena) and Prof. Di Nola (Salerno), the Argentinian researchers Prof. Cignoli and R.Rodriguez (Buenos Aires (bp)), A. Casali (Rosario (bp)), G. Simari (Bahia Blanca) and A.Zapico (Cordoba), the French researchers Prof B.Bouchon and M. Rifqi (Paris(bp)), Prof. D.Dubois and Prof. H. Prade (Toulouse(bp)) and Prof. Chu-Min Li (Picardie), Prof. Gottwald (Germany), Prof. M. Baaz and A. Ciabattoni (Austria(bp)), Prof. C. Gomes (USA), as most important.

The third indicator is the invited talks in conference and workshops and the awards both for papers and software already cited in the corresponding sections.

References

[ABCM04] Carlos Ansótegui, Ramón Béjar, Alba Cabiscol, and Felip Manyà. The interface between P and NP in signed CNF formulas. In *Proceedings, 34th International Symposium on Multiple-Valued Logics (ISMVL), Toronto, Canada*, pages 251–256. IEEE CS Press, Los Alamitos, 2004.

[ABF+06] Carlos Ansótegui, Ramón Béjar, Cèsar Fernàndez, Carla Gomes, and Carles Mateu. The impact of balancing on problem hardness in a highly structured domain. In *Proceedings of the 21st National Conference on Artificial Intelligence, AAAI-2006, Boston/MA, USA*, 2006.

[ABLM07] Carlos Ansótegui, María Bonet, Jordi Levy, and Felip Manyà. The logic behind weighted CSP. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence, IJCAI-2007, Hyderabad, India*, 2007.

[ACG+06] Teresa Alsinet, C. Chesñevar, Lluís Godo, Sandra Sandri, and G. Simari. On the computation of warranted arguments within a possibilistic logic framework with fuzzy unification. In Jurgen Dix and Anthony Hunter, editors, *Proc. of 11th Workshop on Nonmonotonic Reasoning*, pages 227–235, 2006. Technical report IfI-06-04.

[ADL+06] Josep Argelich, Xavier Domingo, Chu Min Li, Felip Manyà, and Jordi Planes. Towards solving many-valued maxsat. In *Proceedings, 36th International Symposium on Multiple-Valued Logics (ISMVL), Singapore*, pages –. IEEE CS Press, Los Alamitos, 2006.

[AdVD+04] Carlos Ansótegui, Alvaro del Val, Iván Dotú, Cèsar Fernàndez, and Felip Manyà. Modeling choices in quasigroup completion: Sat vs. csp. In *Proceedings of AAAI-2004*, pages 137–142, 2004.

[1] We will remark it by (bp) in the corresponding cases

[AEGT04] E. Armengol, F. Esteva, L. Godo, and V. Torra. On learning similarity relations in fuzzy case-based reasoning. In J.F. Peters and A et al Skowron, editors, *Transactions on Rough Sets II: Rough Sets and Fuzzy Set*, volume 3135 of *Lecture Notes in Computer Sciences*, pages 14–32. Springer Verlag, 2004.

[AM04] Carlos Ansótegui and Felip Manyà. Mapping problems with finite-domain variables into problems with boolean variables. In *Proceedings of the 7th International Conference on Theory and Applications of Satisfiability Testing (Revised Selected Papers), SAT-2004, Vancouver, Canada*, pages 1–15. Springer LNCS 3542, 2004.

[AM05a] Carlos Ansótegui and Felip Manyà. Mapping many-valued CNF formulas to boolean CNF formulas. In *Proceedings, 35th International Symposium on Multiple-Valued Logics (ISMVL), Calgary, Canada*, pages 290–295. IEEE CS Press, Los Alamitos, 2005.

[AM05b] Josep Argelich and Felip Manyà. Solving over-constrained problems with SAT technology. In *Proceedings of the 8th International Conference on Theory and Applications of Satisfiability Testing, SAT-2005, St. Andrews, Scotland*, pages 1–15. Springer LNCS 3569, 2005.

[AM06] Josep Argelich and Felip Manyà. Exact Max-SAT solvers for over-constrained problems. *Journal of Heuristics*, 12(4–5):375–392, 2006.

[AMP04] Teresa Alsinet, Felip Manyà, and Jordi Planes. A Max-SAT solver with lazy data structures. In *Proceedings of the 9th Ibero-American Conference on Artificial Intelligence, IBERAMIA 2004, Puebla, México*, pages 334–342. Springer LNCS 3315, 2004.

[AMP05] Teresa Alsinet, Felip Manyà, and Jordi Planes. Improved exact solver for weighted Max-SAT. In *Proceedings of the 8th International Conference on Theory and Applications of Satisfiability Testing, SAT-2005, St. Andrews, Scotland*, pages 371–377. Springer LNCS 3569, 2005.

[BLM06] María Bonet, Jordi Levy, and Felip Manyà. A complete calculus for Max-SAT. In *Proceedings of the 9th International Conference on Theory and Applications of Satisfiability Testing, SAT-2006, Seattle, USA*. Springer LNCS, 2006.

[BMC+07] Ramón Béjar, Felip Manyà, Alba Cabiscol, Cèsar Fernández, and Carla Gomes. Regular-SAT: A many-valued approach to solving combinatorial problems. *Discrete Applied Mathematics*, 2007.

[BMEG+05] B. Bouchon-Meunier, F. Esteva, L. Godo, M. Rifqi, and S. Sandri. A principled approach to fuzzy rule based interpolation using similarity relations. In *Proc. EUSFLAT-LFA 2005*, pages 757–763, Barcelona, Spain, September 2005. Technical University of Barcelona.

[CEGN04] R. Cignoli, F. Esteva, L. Godo, and C. Noguera. On product fuzzy logic with truth constants. In *Actes IFSA05, International Fuzzy System Association*, Beiging, China, August 2004. University of Beiging.

[CGS05] A. Casali, L. Godo, and C. Sierra. Graded bdi models for agent architectures. In J. Leite and P. Torroni, editors, *Computational Logic in Multi-Agent Systems: 5th International Workshop, CLIMA V, Revised Selected and Invited Papers*, volume 3487 of *Lecture Notes in Artificial Intelligence*, pages 126 – 143. Springer, 2005.

[CGS06] A. Casali, L. Godo, and C. Sierra. A methodology to engineer graded bdi agents. In *Proc. of the XII Congreso Argentino de Ciencias de la Computación, Worskshop WASI*, pages 758–767, Octubre 17-21 2006.

[CnSG05] C. Chesñevar, G. Simari, and Lluís Godo. Computing dialectical trees efficiently in possibilistic defeasible logic programming. In C. Baral et. al, editor, *Proc. of the 8th Intl. Conferenceon Logic Programming and Nonmonotonic Reasoning LPNMR 2005*, volume 3662 of *Lecture notes in artificial intelligence*, pages 158–171. Springer, 2005.

[CSAG04] C. Chesñevar, G. Simari, T. Alsinet, and L. Godo. A logic programming framework for possibilistic argumentation with vague knowledge. In M. Chickering and J. Halpern, editors, *Procs. of the Uncertainty in Artificial Intelligence Conference (UAI-2004)*, pages 76–84, Banff (Canada), July 2004. AUAI Press.

[CSAG05] C. Chesñevar, G. Simari, T. Alsinet, and L. Godo. Argument-based expansion operators in possibilistic defeasible logic programming: characterization and logical properties. In L. Godo, editor, *Symbolic and Quantitative Approaches to Reasoning under Uncertainty ECSQARU 2005*, volume 3571 of *Lecture Notes on Artificial Intelligence*, pages 353–365. 2005.

[CSGA05] C. Chesñevar, G. Simari, L. Godo, and T. Alsinet. On warranted inference in possibilistic defeasible logic programming. In B. Lopez et al, editor, *In Artificial Research and Development*, volume 131 of *Frontiers in Artificial Inteligence and Applications*, pages 353–365. IOS press, 265–272, 2005.

[DEGM05] A. Di Nola, F. Esteva, L. Godo, and F. Montagna. Varieties of bl-algebras. *Soft Computing Journal*, 9(12):875–888, December 2005. Special issue devoted to BL Logic and variety (Esteva & Godo ed.).

[DEGP05] D. Dubois, F. Esteva, L. Godo, and H. Prade. An information-based discussion of vagueness. In Claire Lefebvre Henri Cohen, editor, *HANDBOOK OF CATEGORIZATION IN COGNITIVE SCIENCE*, pages 892–913. 2005. Part 8.- Machine Category learning.

[EGGN] F. Esteva, J. Gispert, L. Godo, and C. Noguera. Adding truth constants to logics of a continuous t-norm: axiomatization and completeness results. To appear in Fuzzy Sets and Systems.

[EGM04] F. Esteva, L. Godo, and F. Montagna. Equational characterization of the subvarieties of bl generated by t-norm algebras. *Studia Logica*, 76(2):161–200, 2004.

[EGN] F. Esteva, L. Godo, and C. Noguera. Adding truth constants to t-norm based logics. To appear in the book Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms (E.P.Klement, S.Gottwald, U.Hohle editors).

[EGN04] F. Esteva, L. Godo, and C. Noguera. On rational nilpotent minimum logics. In *Actes ESTYLF04*, pages 413–419, Jaen, Spain, September 2004. University of Jaen.

[EGN06] F. Esteva, L. Godo, and C. Noguera. On rational weak nilpotent minimum logics. *Journal of Multi-valued Logic & Soft Computing*, 12:9–32, February 2006.

[ERBMD04] F. Esteva, M. Rifqi, B. Bouchon-Meunier, and M. Detyniecki. Similarity-based fuzzy interpolation method. In *Proceedings IPMU04. Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 1443–1449, Perugia,Italy, July 2004. University of Perugia.

[Est05a] F. Esteva. From t-norms to t-norm based logics. In *EUSFLAT-LFA 2005*, Barcelona, Spain, September 2005. Technical University of Barcelona. Plenary talk.

[Est05b] F. Esteva. Frrom t-norms to t-norm based logics. In *Proc. 26th Seminar on Fuzzy Set Theory*, pages 47–49, Linz, Austria, February 2005. University of Linz. Plenary talk.

[Est05c] F. Esteva. Logics of a continuous t-norm and its residuum with truth constants. In *Abstract of The Logic of Soft Computing IV and 4th workshop of the ERCIM working group on Soft Computing 2005*, pages 89–91, Ostrava, Czec Republic, September 2005. Technical University of Ostrava. Plenary talk.

[FG] T. Flaminio and L. Godo. A logic for reasoning about the probability of fuzzy events. Fuzzy Sets and Systems (2006), doi: 10.1016/j.fss.2006.11.008.

[Gis03] J. Gispert. Axiomatic extensions of the nilpotent minimum logic. *Reports on Mathematical Logic*, 37:113–123, 2003.

[GM06] L. Godo and E. Marchioni. Coherent conditional probability in a fuzzy logic setting. *Logic Journal of the IGPL*, 14(3):457–481, 2006.

[GNE05] A. García-Cerdaña, C. Noguera, and F. Esteva. On the scope of some formulas defining additive connectives in fuzzy logics. *Fuzzy Sets and Systems*, 154(1):56–75, 2005.

[God06] L. Godo. On argumentation models in a possibilistic logic programming framework with fuzzy unification. In *Abstract of The Logic of Soft Computing IV and 4th workshop of the ERCIM working group on Soft Computing 2006*, pages 13–15, Málaga, Spain, September 2006. Universidad de Málaga. Plenary talk.

[GT05] J. Gispert and A. Torrens. Axiomatic extensions of imt3 logic. *Studia Logica*, 81:311–324, 2005.

[GT07] J. Gispert and A. Torrens. Bounded bck-algebras and their generated variety. *Math. Log. Quart*, 2007. Submitted.

[GZ05] L. Godo and A. Zapico. Lexicographic refinements in the context of possibilistic decision theory. In P. Sobrevilla and E. Montseny, editors, *Proc. of Joint 4th EUSFLAT & 11th LFA Conference*, pages 559 – 564, Barcelona, Spain, September 2005.

[GZ06] L. Godo and A. Zapico. Lexicographic combinations of preference relations in the context of possibilistic decision theory. *Mathware and Soft Computing*, 13(3):157–171, 2006.

[LMP05] Chu Min Li, Felip Manyà, and Jordi Planes. Exploiting unit propagation to compute lower bounds in branch and bound Max-SAT solvers. In *Proceedings of the 11th International Conference on Principles and Practice of Constraint Programming, CP-2005, Sitges, Spain*, pages 403–414. Springer LNCS 3709, 2005.

[LMP06] Chu Min Li, Felip Manyà, and Jordi Planes. Detecting disjoint inconsistent subformulas for computing lower bounds for Max-SAT. In *Proceedings of the 21st National Conference on Artificial Intelligence, AAAI-2006, Boston/MA, USA*, pages 86–91, 2006.

[MNH06] F. Montagna, C. Noguera, and R. Horshik. On weakly cancellative fuzzy logics. *Journal of Logic and Computation*, 16:423–450, 2006.

[NEG05a] C. Noguera, F. Esteva, and J. Gispert. On some varieties of mtl-algebras. *Logic journal of the IGPL*, 13:443–466, 2005.

[NEG05b] C. Noguera, F. Esteva, and J. Gispert. Perfect and bipartite imtl-algebras and disconnected rotations of prelinear semihoops. *Archive for Mathematical logic*, 44(7):869–886, October 2005.

[NEJ05] C. Noguera, F. Esteva, and J.Gispert. On the variety of wnm-algebras. In *Proc. EUSFLAT-LFA 2005*, pages 345–352, Barcelona, Spain, September 2005. Technical University of Barcelona.

[NEJ06] C. Noguera, F. Esteva, and J.Gispert. On n-contractive fuzzy logics: first results. In *Proc. IPMU 2006*, pages 1612–1619, Paris, July 2006. University of Paris-CNRS.

[nNE04] Garcia-Cerda na, C. Noguera, and F. Esteva. On definability of additive connectives in fuzzy logic. In *Proceedings IPMU'04. Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 485–492, Perugia,Italy, July 2004. University of Perugia.

[SCE+06] P. Savický, R. Cignoli, F. Esteva, L. Godo, and C. Noguera. On product logic with truth-constants. *Journal of Logic and Computation*, 16:205–225, 2006.

[TC06] A. Torrens and R. Cignoli. Free algebras in varieties of glivenko mtl-algebras satisfying the equation $2(x^2) = (2x)^2$. *Studia Logica*, 83:155–179, 2006.

[TV] A. Torrens and J.P. Díaz Varela. Decomposability of free £ukasiewicz implication algebras. *Archive for mathematical Logic*. Online DOI 10.1007/s00153-006-0023-1.

[TVFG06] A. Torrens, V. Verdú, J.M. Font, and A.J. Gil. On an infinite-valued £ukasiewicz logic that preserves degree of truth. *Archive for Mathematical Logic*, 45:839–868, 2006.

[VGCB04a] V. Verdú, A. García-Cerdaña, and F. Bou. Analysis of two fragments of the logic of residuated lattices. In *ComLog'04. Combination of Logics:Theory and Applications*, pages 175–185, 2004.

[VGCB04b] V. Verdú, A. García-Cerdaña, and F. Bou. On some substructural aspects of t-norm based logics. In *IPMU 2004 Information Processing and Managment of Uncertainty in Knowledge-Based Systems*, pages 545–552, 2004.

[VGCB06a] V. Verdú, A. García-Cerdaña, and F. Bou. On some classical fragments of t-norm based fuzzy logics. In *IPMU 2006 Information Processing and Managment of Uncertainty in Knowledge-Based Systems*, pages 1596–1603, 2006.

[VGCB06b] V. Verdú, A. García-Cerdaña, and F. Bou. On two fragments with negation and implication of the logic of residuated lattices. *Archive for Mathematical Logic*, 45:615–647, 2006.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

SELF: Software Engineering and Lightweight Formalisms TIN2004-07943-C04

Ernesto Pimentel [*] U. Málaga
María Alpuente [†] T. U. Valencia
Pascual Julián [‡] U. Castilla–La Mancha
Ricardo Peña [§] U. Complutense de Madrid

Abstract

One of the current problems in software systems development is the increasing complexity of analysing and guaranteeing the reliable behaviour of these systems. This project is oriented towards the development of the methods, tools and techniques necessary for supporting quality software construction, with emphasis on practical application to the industrial processes of software companies. This proposal is based on the use of lightweight formal methods in Software Engineering, i.e., the partial application of formalisms at different levels: language, modelling, analysis and composition. The basic idea is to subordinate general methods that support the entire development process and to enhance the real application of formal methods at certain phases of the software life cycle. In order to illustrate the feasibility of this approach, most of the project activities are within the field of component-based software development.

The project is a coordinated proposal of four university teams with previous R+D experience and skills in the area of formal methods in Software Engineering and multiparadigm declarative programming.

1 Project overview

Such as it was planned in the project proposal, all the activities to be developed in this context may be included in one of the following fields: a) Semantic description of software components. b) Software verification and debugging, and c) Optimization of multiparadigm programs. Different contributions in each of these areas were proposed, by using an approach based on lightweight formalisms, i.e. considering the application of formal methods but only when its transfer to an industrial context may be considered as feasible.

In the first research area, the ideas was to explore different possibilities provided by specification languages (e.g. MAUDE) or functional languages (e.g. HASKELL) to describe software systems where each element (subsystem or component) may present different complementary views. In order to guarantee the applicability to industrial environments, a number of goals were included among the project's objectives: the development of techniques to analyse the termination of executable specifications, the study of its efficiency, the proposal of automated optimization techniques, etc. Also related with the semantic description of software components, the project exposed the

[*]Email: ernesto@lcc.uma.es
[†]Email: alpuente@dsic.upv.es
[‡]Email: Pascual.Julian@uclm.es
[§]Email: ricardo@ucm.es

need of complement/extend current interface description languages (used in current component-based platforms) to ensure the correctness of systems constructed by combination of components, by means of automated tools.

In the second research line, the main project's goal was the improvement of *model checking* and *proof carrying code* techniques to make the automated validation of software systems feasible. Thus, on the one hand, the experience of two of the research groups on the application of abstract interpretation techniques to model checking could help to the adaptation of notations, algorithms and tools based on model-checking to the analysis of functional and extra-functional properties. On the other hand, an alternative approach to validate software correctness are the *proof carrying code* techniques, and different open issues were included as goals of the project; in particular, the project included several tasks oriented to apply proof carrying code techniques to declarative programming environments.

As it has been already mentioned, many of the project's goals are related with the exploitation of the high expressive power provided by multiparadigm declarative languages (logic, functions, concurrency, constraints, etc.). For this reason, the proposal was also interested in improving the efficiency of these languages, making their performance comparable to that in imperative languages.

Concerning applications, and maintaining the agile (lightweight) approach, the proposal stressed on specific application domains.

2 Project's achievements

In order to describe the achievements of the coordinated project, we will present the current situation of each subproject. The reader must take into account that we still have one year to finish the project. In any case, the achieved goals are already substantial enough to claim that the project will be successful.

2.1 University of Malaga UMA

The UMA subproject was redefined just before starting, because one of his members, who was responsible of one of the work packages, changed his dedication to a different project. In order to compensate this movement, a new member was added to the UMA's group, who already was cooperating in the work package 1.2. These changes were applied for to the *Subdirección General de Proyectos de Investigación* (Ministry of Education and Science), and they were approved.

Module 1.2: Coordination of Software components

The basic goal of this work package was the development of a coordination methodology to synchronize software components, by extending already well-known interface description languages to deal with behavioural issues. Different task were defined in the project, and the corresponding results are shown below.

Modeling and analysis of software components. We have proposed a process algebra based formalism for describing the interactions between software components [42, 44]. With this same purpose we have explored different coordination mechanisms, such as Linda or Reo [26], comparing their expressive power [25], and UML sequence diagrams [129, 130]. Other different technologies have also been used with this goal, as aspect-orientation [131, 132].

All this work is based on the idea of providing an extension of interface description languages, in such a way that behavioural information is also included in a component interface description. However, making useful this additional information is also an objective. Thus, we have also established techniques of analysis for detecting incompatibility and interoperability problems at the behavioral level [27, 47, 34].

Component adaptation. We have developed a theory of component adaptation. In this sense, we have defined a process algebra based notation for specifying the adaptation required [39, 45, 46], and we have defined its semantics [41]. We have implemented a prototype for an adaptation algorithm for components with interfaces described according to a behavioral type system.

Incremental and dynamic adaptation. We have extended our model and algorithm for component adaptation to take into account dynamic and incremental adaptation [51, 43]. Dynamic adaptation refers to situations in which the requited adaptation between the component may change over time, depending on the context. For that purpose, we have extended the aforementioned notation for considering contextual information. Incremental adaptation refers to situations in which software maintenance, in particular when new functionality is added to components may cause interoperability problems. For that purpose we are exploring how reflection [132] and Aspect-Oriented [50, 49] techniques may help, both defining maintenance in an incremental way, and allowing to get information on components functionality automatically.

Application to Web Services. We have applied the models, notation and tools mentioned above to the adaptation of Web Services, in particular for the adaptation of Web Service choreography and orchestrations [48]. We have also applied Web services-based technologies (e.g. subservicing) as an approach to achieve component adaptation [40].

Module 1.3: Analysis of component-based systems

Research in this module has been mainly focussed on two interrelated lines which have produced a number of results as described below.

Analysis of non-functional properties The goal in this task has been the extension of well-known techniques (typically, model checking techniques) to analyze non-functional properties of systems, such as performance and real time properties. Thus, on the one hand, we have studied the integration of abstraction techniques and model checking in the context of tccp (timed concurrent constraint programming language) to improve the applicability of model checking. We have developed a complete methodology for the abstraction of tccp models by source to source transformation, which allows the complete reuse of tccp model checkers [21, 16]. The proposal is interesting because tccp is a synchronous language and it contains sentences to model typical real time behaviors, such as timeouts, that have not a direct abstraction. In this same context, we have extended the tccp semantics to allow standard model checkers, that analyze temporal properties, to verify real time properties [22, 19].

The other approach in this task is related with the seamless integration of tools that analyze different system properties (performance and functional properties). In particular, making use of XML technologies, we have developed the language PiXL which behaves as a kernel language to interchange the input formats of different tools. PiXL has been used to analyze Enterprise Information Systems (EIS) and it is currently playing a key role in the task of analysis of programming languages described below [121, 89, 88, 87].

Analysis of implementation languages In this task, we have dealt with the problem of analyzing final applications described in a typical implementation language as C. We have explored different approaches to manage the complexity of analyzing programming languages. First, we have concentrated on the verification of distributed applications written in C, making use of the API Socket. In this context, we have followed the model extraction approach in order to construct mixed models for the distributed systems to be analyzed. The new models, that may be verified with SPIN [91], contain both instructions in the modelling language Promela giving the behaviour of the API, and instructions in the original programming language C.

We prove the correctness of the transformation by means of an operational semantics of the API. Tool SocketMC implements this proposal [52, 54].

A parallel research in this area is the integration of abstraction methods and classic static analysis in order to drastically improve the size of the state space stored when realizing model checking. To this end, we have developed the so-called influence analysis which may be used to guide the construction of the so-called abstract matching functions which are used to compare states during the execution of on-the-fly model checking algorithms [53, 85].

We have also worked on the extension of the verification toolbox CADP [1] integrating the model extraction technique above described for concurrent and distributed systems that make use of APIs with well defined semantics. In particular, we have designed and implemented several CADP modules to translate C programs into labelled transition systems which may be handled by the tool [86]. Language PiXL mentioned above has been extensively utilized in this transformation.

Other activities

On the other hand, in a different context, some members of the team have been working on the development of implementation techniques of linear logic-based languages [102, 101]. Although we intended to apply this kind of languages to the description of component behavioural interfaces, we have not explored this research line. Another fruitful research area developed by members of the group is related with the integration of constraint systems in declarative languages [80, 79, 78, 77, 81], in particular to functional-logic languages [64, 63, 66]. The application of these ideas to a specific functional-logic language (TOY) has also been made [28, 65, 82, 83, 30, 29].

2.2 Technical University of Valencia UPV

Module 2.1: Analysis, Optimization and Coordination of Software Components.

In this module we have developed new techniques for proving program termination and other related properties of programs. This includes research on basic techniques, which are central in all termination provers (e.g., polynomial interpretations), basic computational mechanisms which can be useful to model computations in sophisticated programming languages (like context-sensitive rewriting), and transformations of termination problems.

With regard to tasks T.2.1.1, T.2.1.4, and T.2.1.7, the following results have been achieved:

1. We have developed a new framework for using polynomial interpretations over the reals in proofs of termination [104]. This provides the basis for the generation of polynomial orderings in MU-TERM, see [105]. In [106], we have shown that the polynomials with *rational* coefficients are more powerful than (the more usual) polynomial interpretations with integer coefficients (this was a 25-years-old open problem in termination); we have also shown that the use of more general coefficients (e.g., irrational numbers) can also improve over the use of rational coefficients.

2. We have developed a number of new techniques for proving termination of context-sensitive rewriting (*csr*). We have implemented them as part of the tool MU-TERM, whose last version is described in [4]: 1) We have extended Arts and Giesl's dependency pairs approach to *csr* [3, 5]; 2) We have investigated the relative power of the different transformations for proving termination of *csr* [107]; 3) We have developed an expert for automatically proving termination of *csr* by combining the different techniques implemented in the tool [4]. In [90] we have investigated the use of proofs of termination of *CSR* for proving *confluence* of term rewriting systems.

3. We have developed software components (via a COM DLL) with the basic capabilities of MU-TERM, and a graphic interface for using MU-TERM within the .NET framework through

this COM DLL [6, 7]. This provides an interesting means to interconnect software tools, compilers, interpreters, etc., in complex software systems [24].

4. We have developed new techniques for proving termination of concurrent programs with fairness assumptions that can be modeled as term rewriting systems e.g. in Maude [110]. We have also introduced a new notion of termination (called *operational termination*) which is useful to reason about termination of programs whose operational behavior is modeled by means of inference systems [109]. This is interesting to deal with sophisticated programming languages with powerful expressive features which do not fit simple operational models like pure term rewriting. On this basis we have developed new techniques for proving (operational) termination of Maude programs. In [56], we describe a correct transformation of Maude programs into *csr* systems, which allows is to prove termination of these systems by using MU-TERM

Concerning tasks T2.1.2 and T2.1.3., we have investigated the completeness of conditional *csr* systems [103]. We have given conditions ensuring that restricted computations are still able to obtain canonical forms like head-normal forms, constructor terms or even normal forms. Finally, regarding tasks T.2.1.5 and T.2.1.6. the theoretical basis for the extension of Maude with functional-logic capabilities has been discussed in an invited talk given at *15th International Workshop on Functional and (Constraint) Logic Programming, WFLP'06* [58], and also in [13]. As a first step, a preliminary mechanism of unification modulo associativity and commutativity in Maude has been formulated. In 2007, we plan to extend this mechanism to consider more sophisticated narrowing strategies such as natural narrowing. We also plan to implement context-sensitive narrowing in Maude. As a new line emerging from the results achieved in this module, we are starting to apply rule-based techniques to the modeling and analysis of web sites and the semantic web [108].

Module 2.2: Declarative techniques for rule extraction

This module investigates new rule-based techniques for knowledge discovery. We have focused on the definition of a new generalisation framework which is based on the notions of pattern and distance. We have also developed new techniques for extracting or adapting comprehensible models, and specifically, for adapting them to a new context (misclassification costs and class distributions).

In Task 2.2.1, the concept of distance-based binary generalisation operators was introduced [68]. A generalization of this framework which allows us to work with n-ary operators was proposed in [72]. We have applied this framework to first-order objects (atoms and Horn clauses) in [73, 76] and to graphs in [71, 75]. Another related problem investigated in Task 2.2.4 is the definition of similarity functions over structured data, which allows us to apply distance-based methods to complex domains. In [67, 74] we described and analised several distance and pseudo-distance functions for structured datatypes: sets, multisets, terms, atoms, lists and graphs. Some of these functions were derived from kernel functions by exploiting the formal relationship between distances and kernels. In Task 2.2.5 we implemented some of these functions in the Distance-Based Decision Tree (DBDT) learning system defined in [74], which combines decision tree learning and distance-based learning. The key ingredient of our learning method is a new splitting criterion which is based on centroids of only one attribute, while other methods compute centroids by using the whole examples. Hence our method is able to induce models, i.e. decision trees, by using only the distance that corresponds to the datatype of the attribute selected by the splitting criterion. Consequently, the algorithm is able to deal with any datatype which is endowed with a metric distance. Since each partition involves just one attribute, the resulting model can be easily expressed by using declarative rules. The DBDT system has been experimentally evaluated over propositional problems and on datasets that contain structured attributes. It has been also applied to the classification of web pages in

[69, 70]. In Task 2.2.5 we have also applied data mining techniques for hospital management ([2]). The obtained results show a satisfactory performance in all the considered scenarios.

One of the approaches for extracting comprehensible models from other less intelligible models is the mimetic technique. Basically, it consists in using an accurate but generally incomprehensible model as an oracle for generating and labeling an invented dataset. Together with the original training set, this dataset can be then used for training a comprehensible model (for instance, a decision tree) which mimics the oracle. In Task 2.2.2 we have investigated the application of the mimetic technique when the original training data are not available (this is a typical situation in many areas such as expert systems, engineering, diagnosis, medicine, manufacturing, business, etc.). In [36], we have defined a theoretical framework based on the MML (Minimum Message Length) principle for estimating the optimal size of the invented dataset. In Task 2.2.2, we have also investigated the application of the mimetic technique to model revision. Our methodology formalized in [37, 38] applies to the scenario in which a change in the problem definition occurs, e.g. when new data belonging to an area initially not covered by the original model are considered. The key idea is to use the old model as an oracle in order to generate a wide and large dataset representing the initial situation. The empirical evaluation of the method shows that, by considering this dataset (where we have many data) together with the new one (where data are less detailed), the mimetic model performs better than the the classical approach.

In Task 2.2.3, we have defined new evaluation measures which use both estimated probabilities and example ordering [84]. In [35], we analyse three different techniques to establish an optimal-cost class threshold when training data are not available: one technique is directly derived from the definition of cost, a second one is derived from a ranking of estimated probabilities, and the third one is based on ROC analysis.

Module 2.3: Software Verification and Debugging

The main goal of this module is to develop methods, tools and techniques for (automatic) software verification and debugging. The module was organized in 5 tasks. Tasks 2.3.1 and 2.3.2 focused on a main problem of automated program verification and analysis: the state explosion problem. We have investigated three different approaches that improve previous techniques for verifiying reactive systems which are specified in ccp. In [15], we developed a symbolic method which is based on an extension of binary decision diagrams for handling constraints symbolically. In [18, 16, 17] we defined a method based on abstract interpretation which uses both over- and under-approximations in order to improve the accuracy of the approximation while correctly models synchronization. In [20, 22], we have defined a refinement of the logic which is used for the specification of properties in t ccp that also improves the precision and quality of the verification. We proposed a new model for timed ccp which includes an explicit notion of time and allows the user to verify more sophisticated properties. in [23] we have integrated a functional engine within the ccp paradigm. This framework has been implemented in the tccp-func prototype which is written in curry and support the specification of reactive systems that need some kind of sophisticated, arithmetic computations, .

The increased complexity of Web sites and the explosive growth of Web-based applications has turned their design and construction into a challenging problem. In Tasks 2.3.3 and 2.3.4, we have developed a framework for the automated verification of Web sites which can be used to specify integrity conditions for a given Web site, and then automatically check whether these conditions are fulfilled [8, 9, 31]. Our methodology is based on a novel rewriting-based technique, called *partial rewriting*, in which the traditional pattern matching mechanism is replaced by tree *simulation*, a suitable technique for recognizing patterns inside semistructured documents. The framework was in Java and endowed with a powerful rewriting engine written in Maude. We have also developed a novel, semi-automatic rewriting-based methodology for repairing faulty Web sites, as well as its application to the filtering of XML documents based on partial rewriting [10, 11, 33, 32]. With regard to program debugging, starting from the previous system Debussy, we have built an abstract

debugger and a program corrector for Maude which are based on a highly compressed semantics that we developed for this language [12].

Concerning program optimization, we have defined a framework for the analysis, detection and removal of redundant arguments from functions. Our removal mechanism preserves the original program semantics. The proposed framework includes analysis techniques which are based on different formalisms including tree automata, and abstract interpretation. We have implemented a tool that includes some of the proposed techniques in [14].

Finally, Task 2.3.5 focused on the analysis and verification of communication protocols. We have reformulated the "NRL Protocol Analyzer" by exploiting Maude capabilities. We formulated a declarative framework based on rewriting logic in which security properties can be specified and analyzed [57, 62, 59, 60, 61].

2.3 University of Castilla–La Mancha UCLM

In general, the activities of the different tasks have progressed according to the workplan. In the following subsections we summarize the subproject's achievements.

Module 3.1: Multi-paradigm declarative languages: implementation advanced techniques

The main goal of this module is to develop implementation techniques that may contribute to the optimization and increasing efficiency of multi-paradigm declarative languages. We think that the optimization of these languages may help to use them as rapid prototype tools, as well as tools for introducing formal techniques in the field of Software Engineering. Also the improvement of multi-paradigm declarative languages may facilitate their introduction in industrial areas.

We investigate different ways of improving these languages: first, introducing optimizations in high level implementations of multi-paradigm declarative languages; second, examining how to adapt program transformation techniques developed in the functional programming area to the context of multi-paradigm declarative languages. Also we want to explore how to introduce fuzzy unification into the core of these languages. The first goal is dealt with in the task 3.1.1: Improvement of high level implementations.

In [55] we define an algorithm, based on an analysis of definitional trees, that transforms a constructor based, weakly orthogonal program into an inductively sequential one with a deterministic behavior (and therefore, more efficient). We describe the criteria that make this transformation effective and we prove its correctness if non-terminating or undefined expressions are not considered. When non-terminating or undefined expressions are permitted, we provide a negative result: it is impossible to construct a complete transformer to inductively sequential programs.

In [99] and [100] we address the problem of adapting the implementation of a Warren Abstracta Machine (WAM) to incorporate similarity-based fuzzy unification. The clue is to transform the Prolog source program into an intermediate code compiling the information provided by a set of similarity equations, which is proved to be semantically equivalent to the original one. As a result, we obtain a Prolog implementation based on similarity relations that we call S-Prolog. To the best of our knowledge this is the first WAM implementation that supports similarity-based SLD resolution.

Module 3.2: Multi-Paradigm Declarative Languages: transformations and extensions based on fuzzy logic

During the last years, we have had conscience of the important role played by fuzzy logic in the development of sophisticated software tools in many different settings, such as expert systems, medicine, industrial control, and so on. In order to help the development of such applications, it has been observed a growing interest for designing fuzzy declarative (in particular, logic) languages.

Among their expressive resources, these languages incorporate the ability for dealing with imprecise information in a natural way. On the other hand, (the so-called) multi-paradigm declarative programming integrates, together with concurrency, the (two) most important declarative programming paradigms: the functional and the logic ones. Having into account these antecedents, in the present module we focus our attention in two ambitious goals (to be developed in two task too) which can be summarized as follows: the complete integration into the multi-paradigm declarative framework of functional, logic, fuzzy and concurrent programming styles (task 3.2.1), and the subsequent design of transformation, optimization and specialization techniques to be applied to programs written with this kind of languages (task 3.2.2).

Concerning Task 3.2.1, we have developed an intensive work. Thus, in [115] and [118] we propose our first approach to the integration of functional–logic and fuzzy–logic programming paradigms, in order to obtain a richer programming framework where mathematical functions cohabit with fuzzy logic features. We have proposed an hybrid dialect where a set of rewriting rules associated to the functional logic dimension of the language, are accompanied with a set of similarity equations between symbols of the same nature and arity, which represents the fuzzy counterpart of the new framework. We have extended the operational mechanism of needed narrowing to deal safely with similarity relations. On the other hand, in [116] and [120] we have proved that the fuzzy variant of needed narrowing verifies the following properties: (crispness) it computes at least the same elements of the crisp case; (fuzziness) all similar terms of a given goal are completely treated by exploiting the similarities collected in a given program as much as possible; (termination) it avoids the risk of infinite loops associated to the intrinsic (reflexive, symmetric and transitive) properties of similarity relations. Finally, in [117] and [119] we complete the previous operational principle by accompanying it with a powerful notion of equality, called strict-similar equality, which is specially well-suited for declarative languages amalgamating functional-fuzzy-logic features. We have showed that, extending the notion of "strict equality" (used in languages such as Curry) with the more flexible of similar equality (used in the language Likelog), similarity relations can be successfully treated while mathematical functions are lazily evaluated in a given program. Our method can be implemented at a very high abstraction level by simply performing a static pre-process at compilation time which only manipulates the program at a syntactic level.

With regard to Task 3.2.2 (Transformation, optimization and specialization of fuzzy declarative programs), in [94] we introduce an unfolding rule for an extension of the fuzzy logic language defined by Vojtas and Paulík (1996). We demonstrate the strong soundness and completeness of the transformation. In order to prove the strong correctness property we need to adapt the classical concept of a *computation rule* to our setting, and we prove the independence of the new computation rule. Also, we show that our transformation rule always produces an improvement in the efficiency of the residual program, by reducing the length of successful fuzzy SLD-derivations. Moreover, such benefits can be reinforced with the combined use of unfolding with the preliminary version of a folding described in [114]. In [92] we adapt the unfolding rule introduced in [94] for the framework of multi–adjoint logic programming —developed by Medina et. alt. (2004)—. Also for this new framework we prove: i) the independence of the computation rule for multi–adjoint admissible computations. ii) the strong correctness and efficiency properties of the fuzzy unfolding rule.

In [93] and [97], following the clean separation between the operational and interpretative phases in a multi-adjoint logic computation, we distinguish between an operational and interpretive phase during the unfolding process. We introduce a novel *interpretive* unfolding transformation rule, which is proved strong sound and complete. In order to formalize the notion of interpretive unfolding we provide a procedural characterization of the interpretive phase of a computation (in terms of a state transition system) what constitute a distinguished feature with regard the approach considered in the original multi-adjoint logic programming framework.

In [96] we have introduced a definition of the concept of Partial Evaluation (PE) for multi-adjoint logic programs and goals. Using PE techniques we are able to define a refinement to the notion of classical reductant, that we call $PE-reductant$. Reductants are crucial to cope with a

problem of incompleteness in the context of multi-adjoint logic programming. In [95] and [98] we provide a concrete algorithm for the construction of PE-reductants which is based on unfolding with a set of dynamic thresholds. This provide an effective and efficient method for computing reductants, based on PE techniques, which constitutes a novel application in the multi-adjoint logic programming framework.

2.4 Universidad Complutense de Madrid UCM

This subproject is placed in the Proof Carrying Code research area. In this framework, a code consumer verifies that the code provided by a non-reliable producer fulfils a set of security rules, which constitutes its *security policy*. This is done be mecanically checking a formal proof which the producer attaches to its code. The security policy which constitutes the main objective of the subproject is to certificate that programs have a *safe* memory management. In our context, safety means two things: (a) absence of dangling pointers; (b) programs have a proved upper bound on its run-time memory needs.

We have approached these elements from the program analysis perspective, analyses based both on abstract interpretation and on types and effects systems. We have decided the source language to be a functional one, building on our previous knowledge on this area. The detailed objectives can be enumerated as follows:

1. Definition of the language facilities from the programmer's point of view.

2. Definition of the different analyses. These should focus on placement of data structures, sharing properties, program termination and memory consumption. Usage of types guarantees certificates compositionality, so that a function that uses an already certificated function may incorporate that certificate.

3. Semantics definition, abstract machine definition and compilation process definition. From the source language we intend to obtain a Java-like byte-code.

4. Generation of certificates about the properties satisfied by programs.

5. Implementation of all the above.

We still have not achieved the objectives related to the termination analysis, the memory cost analyisis, and the certificate generation. A cooperation is running with the subgroup at UPV in order to use their termination tools to approach the first objective. Once this is solved, we hope that it will the basis to solve the second one. The third one —certificate generation— is being approached with an ongoing cooperation with Julio Rubio's group at the *Mathematics and Computation* Department of *Universidad de La Rioja*. This group has expertise in mecanically proving theorems with tools such as ACL2, Isabelle, and Coq. A help from them is possible in order to translate our correctness theorems into a formal proof that these tools could validate

¿From the results achieved in the subproject, we have presented and published the language syntax, examples of language usage and a preliminary version of the type system in [122, 123]. The sharing analysis and the big-step operational semantics have been presented in [124, 127] and will be formally published in [128]. The abstract machine derivation and the code generation phase have been presented in [125, 126] and they have been submitted for formal publication. The safety-types inference algorithm will be shortly presented at The Trends on Functional Programming Workshop TFP07, next April [113]. The correctness theorems have not been published yet. We are currently preparing a journal version of the whole language which will include these proofs.

Below, we give detailed information concerning work-packages such as they were defined in the project proposal.

Module 4.1: Design of the high-level language and the compiler–certifier

This first module was scheduled around the following activities:

Source language definition A first version of the language, called SAFE, has been thoroughly defined. It is eager, first order, polymorhic and with user-defined data constructors. There is a full-SAFE version, in which programmers write programs, and a Core-SAFE version in which the different analyses are defined. Full-SAFE has a Haskell-like syntax and programs scarcely differ from the ones a conventional functional programmer would write. The programmer has a novel facility for *destructive pattern matching* which allows to explicitly dispose the cells occupied by a particular data structure. A higher-order extension of the language is foreseen.

Big-step operational semantics Core-SAFE has explicit *regions* in which data structures are placed. No garbage collector is needed. The big-step operational semantics detailing the evaluation of expressions and programs, as well as region and closure creation/disposal has been defined.

Small-step operational semantics A second small-step operational semantics has also been defined. This is needed in order to prove the correctness of the type system and for the formal derivation of the abstract machine. We have formally proved the equivalence of the big-step and the small-step operational semantics.

Sharing analysis At Core-SAFE level, we have defined an abstract interpretation-like sharing analysis which annotates programs with four relations between the variables in scope at each particular expression. These relations tell which are the sharing properties of the data structures to which the variables refer to. For instance $x\ Sh\ y$ tells that structures pointed to by x and y will share at run-time a common substructure.

Sharing analysis correctness theorem We have defined an *abstraction function* between run-time pointers and compile-time program variables and formally proved that the sharing detected by our analysis is an upper bound of the actual sharing taking place at run-time. In this proof we have used the big-step operational semantics.

Safety type system The above relations between variables are extensively used by our type system ir order to give correct types to programs. The novelty of this type system is that it annotates types with information about the destruction of data structures. In this way, a data structure (e.g. a list) can be *condemned* (this means that it will be destroyed in the future), *in danger* (this means that it shares a condemned piece) or *safe* (it does not share or contain condemned pieces). Based on these annotated types, the type system prevents programs from performing dangerous actions such as reading an already destroyed data structure. The type system also detects in which regions data structures should be created in order that region disposal does not generate dangling pointers.

Safety type system correctness theorem We have formally proved that a well-typed program is guaranteed not to run into dangling pointers when executed. This has been the major achievement of the project up to now. The correctness proof is about ten pages long and uses the small-step semantics of the language, as well as the sharing analysis correctness theorem.

safety type inference algorithm We have defined an inference algorithm to compute SAFE types automatically, without relying on programmer's annotations. The algorithm has a Hindley-Milner phase with some additions in order to give polymorphic types to region identifiers, and a *destruction-propagation* phase which infers the condemned, in-danger or safe types for each of the variables and function arguments. The algorithm is modular in

the sense that the inference of each function can be done by only using the already inferred types of the invoked functions.

Implementation All the above phases, except code generation, have been implemented. We have now a prototype of the certifying compiler to which new analyses and phases can be added. It is written in Haskell and has about 4000 lines up to now.

Module 4.2: Imperative abstract machine design and certify checking

The activities of this module have been made around the following tasks:

Imperative abstract machine From the small-step operational semantics we have formally derived several abstract machines, the last one of which is very close to the Java Virtual Machine. This process has been done by proving the correctness of each step so that it is guaranteed that the abstract machine correctly implements the semantics of the language.

Formal derivation of the code generation Translation schemes from Core-SAFE to abstract machine byte-code have been defined and proved correct. Now, we are in a position to generate imperative code from the full-SAFE language.

3 Results indicators

We include below some relevant indicators which provide information about the activities in which project members have been involved.

3.1 Summary of the most relevant publications:

In Section 2, different works were referenced showing the contributions in each subproject. Thus, the bibliography of this document may be considered a full list of publications derived from the project. To give an idea about the relevance of these contributions, we classify them by scope and type of publication:

International journals	30 (19 indexed in SCI)
Book chapters	2
Electronic journals	18 (ENTCS series)
Int. conferences and works.	48 (Springer LNCS, IEEE Press, ACM Press)
National conferences:	27
Others (PhD Thesis)	2

3.2 Software systems:

αSPIN	`http://gisum.lcc.uma.es/~gisum/fmse/tools`
MU-TERM	`http://www.dsic.upv.es/users/elp/slucas/muterm`
Natur	`http://www.dsic.upv.es/users/elp/natur`
GVERDI	`http://www.dsic.upv.es/users/elp/GVerdi/`
GVerdi-R	`http://www.dsic.upv.es/users/elp/GVerdiR/`
tccp-func	`http://www.dsic.upv.es/~villanue/tccp-func/`
DBDT	`http://www.dsic.upv.es/~flip/dbdt/`

3.3 Training activities:

The following PhD students are currently developing their PhD within the topics of the project:

Module 1.2: Javier Cubo (FPI-MEC), Javier Cámara (FPI-Junta Andalucía), Ana M. Roldán (Assistant Prof, Univ. Huelva), Silvia Amaro (Lecturer, Univ. Comahue, Argentina), José Antonio Martín (SELF contract).

Module 1.3: David Sanán (FPI-MEC).

Module 2.1: Beatriz Alarcón (FPU-MEC), Sonia Flores (ALFA LerNet grant), Raúl Gutiérrez (FPI-SELF grant), and Rafael Navarro (SELF contract; FPI-GV grant, pending).

Module 2.2: Antonio Bella (SELF contract; FPU-MEC grant, pending), Ricardo Blanco (SEIT-ANUIES), and Vicent Estruch (Ayudante).

Module 2.3: Mauricio Alba (ALFA LerNet grant), José Iborra (FPI-UPV grant), Alexey Lescaylle (SELF contract), Pedro Ojeda (MIUR grant, co-tutela), Daniel Romero (ALFA LerNet grant), and Matteo Zanella (Socrates co-tutela).

Module 4.2 Manuel Montenegro

By means of "Juan de la Cierva" contracts, Christoffe Joubert from INRIA-ALPES, who has has been working with Málaga in 2006, joins the UPV group in February 2007. With similar conditions, Gwen Salaün will (also from INRIA-ALPES) joined the UMA group since December 2006. Their activities will be developed basically in modules 2.3 and 1.2, respectively.

As part of the training activity of the project, we could also mention that some people involved in the project have moved to companies: Javier García (ex-FPI grant, with Hewlett-Packard now), Jose Daniel Llopis (ex-SELF contract, with Regional Industry Ministry now).

3.4 New national and international Collaboration:

Strongly related to the SELF objectives, the UMA team has participated in different CENIT (industrial-oriented projects) proposals, still pending of evaluation. UMA node is also actively participating in one of the work packages of the IST-033563 project (funded by the EU), and coordinated by Manuel Díaz (also from the University of Malaga). UMA node is currently maintaining some contacts with Sangiorgi and Zavattaro (Univ. of Bologna, Italy), Honda and Yoshida (Queen Mary College, UK), Pistore (ITC-irst Trento, Italy) and Margaria (Univ. of Dortmund, Germany), to apply for a project in the first call of FP7.

Also related to the SELF project, a new German-Spanish Acción Integrada has been recently approved which starts in January 2007.The bi-national project, lead by Jürgen Giesl (RWTH Aachen, Germany) and Salvador Lucas (UPV) aims to investigate the development of efficient tools for proving program termination. On the other side, a new national research network which focus on the development of the language Maude also starts in January 2007, led by Prof. Narciso Martí-Oliet from the Universidad Complutense de Madrid (UCM) (TIN2005-25854-E). Other partners of this project are UMA, UPV, UEX, UMU, and USA. Other two French-Spanish collaboration actions (Acción Integrada) were also applied for and scientifically well valuated, but they were not approved because of funding limitations.

The UPV node is member of the Coordination Action (CA) for Ubiquitous Knowledge Discovery, KDubiq, funded by the European Union under IST (Information Society Technology), FET Open (Future and Emerging Technologies) in the 6th Framework Programme. The research topics supported by this CA are related to the research activity developed in Module 2.2.

UCM node has started a new contact with researchers working on similar languages, such as the Munchen's group of Martin Hofmann and Hans-Wolfgang Loidl, and the Scottish groups that are currently developing the functional language Hume. UCM group has invited Hans-Wolfgang to visit Madrid next spring in order to interchange their respective experiences.

3.5 Other activities:

1. Participation in Program Committees:

   (a) International conferences: LPAR-12, FASE'05, COORDINATION'05, IJCAI'05, ICML'05, ECML'05/PKDD'05, LOPSTR'05, AISC'06, FMOODS'05, FACS'06, FASE'06, COORDINATION'06, ICLP'06, LPAR-13, ISOLA'06, FASE'06, ECAI'06, ECML'06/PKDD'06, IDEAL'06, RTA'07, PPDP'07, COORDINATION'07, FMICS'07.

   (b) International workshops: RULE'05, AAIP'05, ROCML'05, WCAT'05, IDEAS'05, RISE'05, FMICS'05, IDEAS'05, FOCLASA'05, IDEAS'06, WFLP'06, WRLA'06, WLPE'06, WRS'05, WRS'06, WRS'07, WWV'05, WWV'06, ROCML'06, WTDIA'06, WCAT'06, FOCLASA'06, RISE'06, FMICS'06, IDEAS'07.

   (c) National conferences: JIDI'05, PROLE'05, JISBD'05, TAMIDA'05, CAEPIA'05, PROLE'06, JISBD'06, WISBD'06, JIDI'06, JISBD¡07, PROLE'07, CEDI'07, CAEPIA'07, TTIA'07

2. Research Stays: Some members of the project have made different research stays:

Researcher	Centre	Period
María Alpuente	Birla Science Center (India)	Feb. 2005
María Alpuente	U. San Luis (Argentina)	Oct-Nov. 2006
Carlos Canal	Univ. of Pisa, Italy	Sept. 2005
Carlos Canal	Univ. d'Evry, France	July 2006
Carlos Canal	Univ. Extremadura	June 2006
Francisco Durán	Urbana-Champain (USA)	Aug 2005
Santiago Escobar	University of Illinois at Urbana-Champaign (USA)	Apr. - Aug. 2005 &Apr. - Sept. 2006
Santiago Escobar	U. Udine (Italia)	May 2006
Santiago Escobar	ENS Cachan (France)	March 2006
Santiago Escobar	ETH Zurich	March 2006
Santiago Escobar	SRI, CA (EEUU)	March 2006
Santiago Escobar	Naval Research Lab (USA)	Nov. 2005 & Sept. 2006
Santiago Escobar	Birla Science Center (India)	Feb. 2005
Vicent Estruch	Fraunhofer Inst. Bonn (Germany)	Oct. - Dec. 2006
Javier García-Vivó	University of Siena (Italia)	Sept. - Oct. 2005
Raul Gutierrez	Birla Science Center (India)	Dec. 2006
Raul Gutierrez	Cons. Nal. des Arts et Metiers CNAM	Dec 2006
José Iborra	U. Siena (Italy)	Nov.-Dec. 2006
Pablo López	Carnegie Mellon	Jul-Sept 2005
Salvador Lucas	University of Illinois	May 2005 & May 2006
Salvador Lucas	U. San Luis (Argentina)	Oct. 2005
Salvador Lucas	Tech. Universität, Wien (Austria)	Sept. 2005 & March-Apr. 2006
Pedro Ojeda	U. Siena (Italy)	Oct.-Dec. 2006
Ernesto Pimentel	U. Pisa (Italy)	Feb. 2005
Daniel Romero	U. Udine (Italy)	May 2006
Daniel Romero	Birla Science Center (India)	Sept. 2006
Alicia Villanueva	TU Wien	Nov. 2005 & March 2006
Alicia Villanueva	U. Málaga	Nov. 2006

And we have also enjoyed the visits of a number of researchers in 2005-2006:

- Researchers visiting UMA node: Carlos Areces (INRIA, Nice, France), Antonio Brogi (Univ. Pisa, Italia), Miguel Katrib (Univ. La Habana, Cuba), Tiziana Margaria (Univ. Dortmund, Germany),José Meseguer (U. Illinois in Urbana-Champaign, USA), Pascal Poizat (Univ. d'Èvry, France), Jeff Polakov (Nat. Inst. of Industrial Science and Tech., USA), Gwen Salaün (INRIA-ALPES, France).
- Researchers visiting UPV node: Marco Comini (U. di Udine, Italia), Evelyne Contejean (LRI, U. Paris-Sud), Pierre Courtieu (LRI, U. Paris-Sud), Nachum Dershowitz (Tel Aviv University, Israel), Yaniv Eytani (U. Illinois in Urbana-Champaign, EE.UU.), David Dowe (Monash University, Australia), Bernhard Gramlich (Technische Universität Wien, Austria), Christophe

Joubert (INRIA Grenoble, Francia), José Meseguer (U. Illinois in Urbana-Champaign, USA), Ugo Montanari (U. di Pisa, Italia).

3. Organization of International Conferences/Workshops:

   (a) 18th European Summer School on Language, Logic and Information. ESSLLI. August, 2006. Malaga.

   (b) 5th Workshop on the Foundations of Concurrent Languages and Software Architectures (FOCLASA'06), Bonn (Germany), August 31, 2006, affiliated to CONCUR 2006.

   (c) 2nd Workshop on Coordination and Adaptation Techniques for Software Entities (WCAT'06). Nantes (France), July 4, 2006, associated to ECOOP 2006.

   (d) 4th Workshop on the Foundations of Concurrent Languages and Software Architectures San Francisco (USA), August 27, 2005, associated to CONCUR 2005.

   (e) 1st Workshop on Coordination and Adaptation Techniques for Software Entities (WCAT'05). Glasgow (United Kingdom), July 25, 2005, associated to ECOOP 2005.

   (f) 1sA t Int'l Workshop on Automated Specification and Verification of Web Sites WWV 2005, March 14-15, 2005, Valencia, Spain (Proc. in ENTCS [112], Elsevier)

   (g) ROC Analysis in ML, ROCML'05, held within the 22nd Int'l Conf. on Machine Learning ICML 2005, Bonn, Germany, Aug 7-11, 2005.

   (h) 2nd Int'l Workshop on Automated Specification and Verification of Web Sites, co-located with ISOLA 2006, Paphos, Cyprus, November 19, 2006 (Proc. IEEE Computer Society Press [111])

   (i) Jornadas de Programación y Lenguajes. PROLE 2004. Malaga, November 10-12, 2004.

   (j) Jornadas de Ingeniería del Software y Bases de Datos. JISBD 2004. Malaga, November 10-12, 2004.

   (k) Annual Meeting of the European Consortium for Informatics and Mathamatics (ERCIM). Malaga, November 2-5, 2004.

References

Because of space limitations, references (which are the result of the project's dissemination activity) are not included in this document, but they may be consulted at:
`http://platon.escet.urjc.es/jspTIN2007/ActasPDF/TIN2004-07943.pdf`

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

Semantic Interoperability in Systems that Manage Domain Specifics Data (Coordinated Project) TIN2004-07999-C02

M. Aranzazu Illarramendi *
Main Researcher of Subproject 1

Eduardo Mena †
Main Researcher of Subproject 2

Abstract

.

Keywords: Data Management, Semantic Web, Ontologies, Agent Technology, PDAs

1 Goals of the Project

This project is framed inside the area of the data management. Its main target consists of developing an architecture that allows the systems, in charge of managing domain specifics data, a semantic interoperation. That kind of interoperation can be at an internal level, on the data that the systems handle and the access to the same ones, as well as at an external level, with other systems.

2 Achieved goals in the project

In short, the project has two main purposes. First of all, to advance in the development of "intelligent" systems for the management of data of specific domains, which constitutes undoubtedly a type of basic research work. Secondly, to reflect the results of this theoretical work in different prototypes, which constitutes a type of essentially applied work.

Concerning the first purpose, as we have mentioned previously, the main goal consists of developing an architecture that allows the systems that are in charge of managing data of specific domains to semantically interoperate at internal level (on the information that they handle and the access to the same ones) as well as at external level (with other systems). Each level is considered by one of the two subprojects that constitutes this coordinated project.

*Email: a.illarramendi@ehu.es
†Email: emena@unizar.es

2.1 Semantic Interoperability at External Level

Through the years, different organizations and enterprises have developed several data systems that are vital for their daily operations. A step forward in this scenario consists of offering a real interoperation possibility among those independently developed data systems, by making minor adaptations on them. By real interoperation we mean an interoperation based on the semantics of the communications, which takes the matter far beyond the syntactic functionality provided by developed exchange standards such as the widely spread XML and, more specifically, EDI standards in the area of electronic commerce.

Furthermore, the capability of agent technology for approaching problems showing highly distributed nature that needs flexible and adaptable solutions is broadly recognized, and so there is a tendency to enhance the functionality of data systems by appropriate data agents. Communication among agents is in general based on the interchange of messages. However, different data systems have incorporated different classes of messages as their Agent Communication Language (ACL) to the point that they do not understand each other. In our opinion, a real interoperation across systems will be possible if there exists a wide agreement on the classes of messages used by the agents and the possibility of constructing new kinds of messages by composition or restriction of already known classes.

Emerging Semantic Web technologies can play a relevant role in the development of that new kind of data systems interoperation frameworks and in this sense we have built a formal ontology, that we call CommOnt (COMMunication acts ONTology), which can play a key role in such type of frameworks and, in particular, in the framework that we have developed.

2.1.1 COMMONT Ontology

From a pragmatic viewpoint, ontologies are agreements to accomplish specific objectives. In our case the objective is *agentized data systems interoperation* and so, our goal is to stress the benefits of achieving an ontological commitment about communication acts performed by cooperative data agents. We do not however pretend to present our ontology as a definitive ontology for communication among agents

Ontological commitments are agreements to use the shared vocabulary in a coherent and consistent manner. An agent commits to an ontology if its observable actions are consistent with the definitions in the ontology. The behavior of agents committed to our CommOnt ontology must be observably consistent with the definitions in CommOnt. Thus, our approach is compatible with and complementary to the development of standards for agent communication languages like KQML or FIPA-ACL.

We are conscious of the many possible different approaches to the leading guideline for defining the terms and relationships in a communication acts ontology. The design criteria adopted for the communication acts category of our CommOnt ontology is to follow the *speech acts* theory, a linguistic theory that is recognized as the principal source of inspiration for designing the most familiar standard agent communication languages.

In the speech acts theory, messages are considered actions, specifically communicative actions, as they are intended to perform some action by they being sent. Following that theory, and keeping in mind our context, every communication act is the speaker's expression of an attitude toward some possibly complex proposition; and that perspective has been called the *F(P) framework*. Therefore, a communication act has two main components. First, the atti-

tude of the speaker which is called the *illocutionary force F*, that expresses social interactions such as informing, requesting or promising, among others. And second, the *propositional content P* which is the subject of what the attitude is about. In CommOnt we follow this F(P) framework. Different kinds of illocutionary forces and propositional contents lead to different classes of communication acts.

The designed ontology is divided into three interrelated layers: *upper*, *standards* and *applications* respectively and we highlight three categories: *actors* that interact among them using different kinds of messages; *communication acts* that have different purposes and deal with different kinds of contents, and *contents* that are the sentences included in the messages. The upper layer includes the term `CommunicationAct`. Every message is an individual of the class `CommunicationAct`. Main subclasses of `CommunicationAct` are `Constative`, `Directive` and `Commissive`, among others. There exist some specializations of those classes. For instance, `Inquiry` is a subclass of `Directive` including the messages that ask for some information (i.e. messages asking queries), while `Request` is another subclass of `Directive` including messages demanding the performance of another kind of action. The standards layer is designed to include the classes of messages that different standard agent communication languages have defined (or can define in the future). Finally, the applications layer reflects the terms describing messages used in each single data system. Axioms exist in the ontology which describe the interrelationships among terms of these categories. We claim that the whole communication acts ontology provides interoperability support resulting from the recognition of communication acts steming from one ACL as instances of communication acts in another ACL. Sometimes the comprehension will not be complete, but partial comprehension of the communication may be useful and preferable to the "not understood" answer given nowadays.

2.1.2 Framework

The framework that we propose permits agents belonging to different data systems 1) to send each other suitable messages without requiring the establishment of a common communication pattern in advance; 2) to understand, completely or partially, those messages that are interchanged among themselves; and 3) to invoke web services offered by the data systems at a high level without needing to go into technical details.

Following, we present a simple set theoretic model of the proposed framework which provides an abstract understanding of the relationships among the components, independently of implementation details.

In order to deal with our proposal, the following 7-tuple $\Gamma = (\Omega, \Delta, O, \tau(E), \tau(\Lambda), \pi, \lambda)$ of components should be incorporated at each node where a data system runs.

Let C, P, I, V be the sets of ontology concepts, properties, instances and literal values respectively. We call T to the set of OWL triples in the cartesian product $(C \cup I) \times P \times (C \cup I \cup V)$. Then,

- $\Omega \subseteq T$ is the CommOnt ontology where the classes of messages, contents and actors are designed. CommOnt has three layers; the two first layers can be imported from one repository and the application layer must be designed by the data system administrator;
- $\Delta \subseteq T$ is a *domain ontology* where the terms for referring to objects and properties of the application domain are specified. This ontology can be profitably imported from shared repositories;

- $O \subseteq T$ is an *ontology* describing suitable *operations* or *actions* to be performed by agents. This ontology can also be imported.

Messages are communicative actions expressing the sender's attitude toward some possible complex proposition. Therefore, a message has two main components. First, the attitude of the sender which expresses intentions such as informing, requesting or promising, for example. And second, the propositional content which is the subject of what the attitude is about.

Let $M = E \cup \Lambda$ be the vocabulary for writing messages; where E represents the vocabulary subset for writing the "envelope" of messages and Λ represents the vocabulary for the content language of messages. Then,

- $\tau(E) \subseteq E \times (C \cup P)$ is an *attribute mapping* that relates attributes from the envelope of messages with ontology terms;

- $\tau(\Lambda) \subseteq \Lambda \times (C \cup P)$ is an *attribute mapping* that relates content language features with ontology terms.

A key point of our approach is that every message has an abstract representation as an individual of a shared universal class of messages. Such representation is founded on OWL triples[1]. The Web Ontology Language OWL is a W3C recommendation that facilitates greater machine interpretability.

Let $\Phi \subseteq T \cap (I \times P \times (C \cup I \cup V))$ be the collection of OWL triples which express statements about instances; and let M^* and Λ^* be the sentences for messages and contents, respectively. Then,

- π: $M^* \stackrel{\tau(E),\Omega}{\longrightarrow} \Phi \times \Lambda^*$ is a *platform mapping* that decompose messages into a collection of triples and in one content language expression.

- λ: $\Lambda^* \stackrel{\tau(\Lambda),\Omega,\Delta,O}{\longrightarrow} \Phi$ is a *content language mapping* that applies content language expressions on the corresponding collection of triples;

Following we describe the steps followed during a communication process among agents from different data systems. There is a particular agent, called *CommOnt Manager* (CM) in charge of coordinating the process of dealing with the ontologies (Ω, Δ, O), and it plays the role of intermediary among agents from different data systems. There will be a *CommOnt Manager* associated within each data system. One main reason to include this *CommOnt Manager* agent in the framework is that only one agent needs to be a specialist in the process of dealing with the ontologies. However, this agent could be eliminated and the proposed framework would also be valid in order to allow direct communication among agents of different data systems (in this case each agent should deal with the ontologies). Let us suppose two systems Σ and Θ, and an agent A_Σ from the first system sending a message to an agent A_Θ from the second system. The process consists of the following steps (see figure 1):

1. The agent A_Σ creates the appropriate message $\mu\Sigma$ in the format used in the Σ system. Also creates and sends an $\epsilon\Sigma$ message requiring the CM_Σ agent to send the $\mu\Sigma$ message.

[1]More precisely, they are RDF triples. But, since OWL is an extension of RDF, we highlight them as OWL triples.

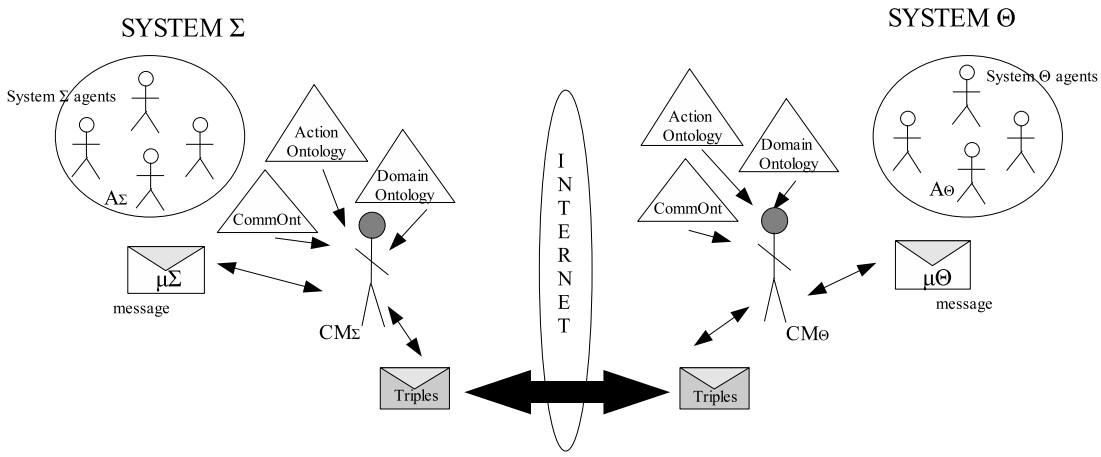


Figure 1: Architecture of the proposal.

2. The CM_Σ agent transforms the message $\mu\Sigma$ into an equivalent collection of triples using $\pi_\Sigma(\mu\Sigma) = (\mu\phi, \mu\kappa)$ and $\lambda_\Sigma(\mu\kappa)$.

3. The CM_Σ agent sends the collection of triples to the CM_Θ agent.

4. The CM_Θ agent, with the help of ontology reasoners derives more triples and calculates the most specific class of the $\mu\Sigma$ message within the CommOnt ontology. Let us call it $msc(\mu\Sigma)$.

5. If CM_Θ recognizes $msc(\mu\Sigma)$ as an understandable message class for A_Θ, then it constructs a $\mu\Theta$ message out of the enriched collection of triples using[2] π_Θ^{-1} and λ_Θ^{-1}. A complete understanding is achieved in this case if the systems Σ and Θ share the domain and operation ontologies Δ and O.

6. In case of CM_Θ does not recognize $msc(\mu\Sigma)$ as an understandable class for A_Θ, it sends a message to CM_Σ advertising which are the subclasses of $msc(\mu\Sigma)$ that A_Θ is able to deal with. In this case, for the moment, only a partial understanding is achieved. However, CM_Σ can search in CommOnt for ontological relationships among those subclasses of $msc(\mu\Sigma)$ and classes of messages that A_Σ understands. Then CM_Σ has the opportunity to inform A_Σ of the capabilities of A_Θ just in terms of Σ messages. In the best case, A_Σ is able to reconstruct its original $\mu\Sigma$ message (or something similar) and renew the communication process.

2.2 Semantic Interoperability at the Internal Level

The World Wide Web is an information resource with virtually unlimited potential. However, this potential is not fully exploited by traditional web search engines, because they do not consider the context of user keywords: the same keywords can be used by different users with

[2] π_Θ^{-1} and λ_Θ^{-1} are the inverse mappings of π_Θ and λ_Θ, respectively.

the purpose of accessing to different information, i.e., keywords can be interpreted differently as they lack explicit semantics. In this context, ontologies (which offer a formal, explicit specification of a shared conceptualization arise as a solution to help to contextualize user keywords, since they make possible to associate concepts and properties around a specific domain. The more ontologies consulted (each one representing the point of view of their creators), the more chances to find the semantics that the user assigned to keywords.

As motivating example, let us suppose that the user wants to find information about the life of famous people and writes the following keywords: *"life of stars"*. Google returns about 269,000,000 hits[3] when we enter those keywords but, unfortunately, the first tens of hits link to astronomy-related web pages. We see how syntactic-based search engines are very influenced by the enormous amount of information about popular issues on the Web. Similar results are obtained if the keyword is "java": Java as programming language eclipses the rest of possible senses (the Indonesian island, a coffee plant, different US cities, etc).

Our approach takes as input a list of plain keywords provided by the user, discovers their semantics in run-time and obtains a list of senses extracted from different ontology pools; it deals with the possible semantic overlapping among senses. We summarized the main steps in the following (see Figure 2).

2.2.1 Extraction of Keyword Senses

In the first step, the user keywords are normalized (e.g., rewriting them in lowercase, removing hyphens, etc.), and then, the *sense discovering* step queries Swoogle, WordNet and other ontology repositories to find ontological terms that match those keywords. A *black list* is managed to avoid known low-quality ontologies, and a *buffer* stores previously parsed ontology terms to avoid accessing the same remote data twice. Second, the *sense extractor* step builds a sense for each URI obtained in the previous step. Each keyword sense is enhanced incrementally with the synonym terms extracted by the *synonym extractor* step (which also searches the ontology pool). The *sense alignment* step integrates the keyword sense with those synonym senses representing the same semantics. After discarding the synonym senses that do not enrich the keyword sense, the result is a list of possible senses for each keyword. Finally, as previous senses were built with terms coming from different ontologies, a final *sense alignment* step integrates senses of the same keyword that are so similar that they represent the same semantics.

Notice that the whole process can be limited in time; obtaining the senses is executed in parallel for each keyword; within that task, the semantic enrichment of each keyword sense with its synonym senses is performed in parallel too.

2.2.2 Alignment of Senses

As we said before, the same iterative process is performed 1) to align synonym senses with its keyword sense, and 2) to align the candidate senses of a keyword to avoid redundancy. This process improves the quality and efficiency of the ontology matching and enables the reuse of new discovered senses. In Figure 2, on the right, we can see this process in detail: each input sense is compared with the rest of stored senses in order to decide if it can be integrated with some sense. For this, the system *computes the synonymy probability between two senses* to

[3]Obtained in December 12, 2006.

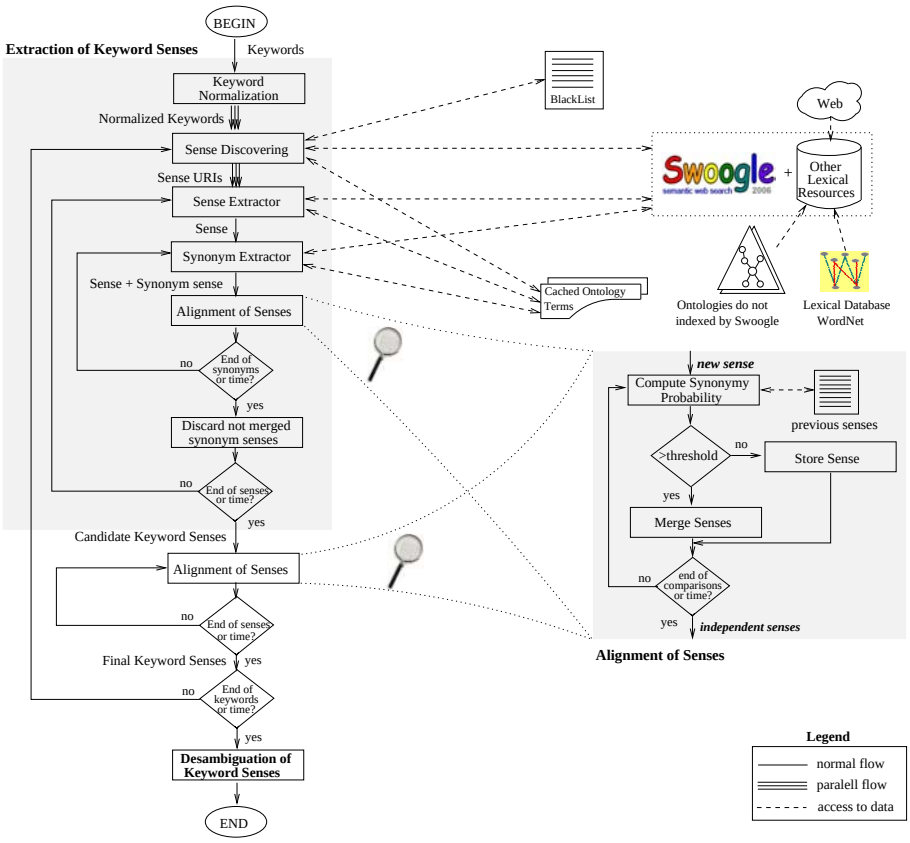


Figure 2: Obtaining the possible senses of user keywords.

conclude whether they represent the same semantics or not); if this happens, the same process is repeated for the integrated sense until all the comparisons are performed or a specified amount of time is spent.

2.2.3 Disambiguation of Keyword Senses

The previous steps provide a set of possible (non-redundant) senses for each user keyword, but what is the intended meaning of each keyword? we guess that the user intention is implicitly contained in the set of keywords that he or she entered. For example, if the keywords are "star astronomy planet", the intended meaning of "star" could be "celestial body"; but in "Hollywood film star" the meaning for "star" is probably "famous actor'. To identify the right meaning in these contexts is an easy task to a human observer, but it is difficult to be performed automatically. Other examples are still ambiguous even for a human, as our motivating example "life of stars".

We propose a method to disambiguate[4] iteratively each user keyword taking into account

[4]Disambiguation is the process of picking up the most suitable sense of a polysemous word according to a

the possible senses of the other keywords. Its output is a single ontological sense for each keyword. Such a method computes a relatedness measure based on Google to compare each possible sense with respect to the senses of the other keywords. Then a relevance degree is calculated for each keyword sense and the list of senses are rearranged according to it, filtering the less relevant ones. The method can operate in two modes: 1) automatic, the system selects the one with the highest relevance degree, so avoiding user intervention but trusting the decision taken by the computer; and 2) semiautomatic, the system proposes to the user the rearranged and filtered list of senses obtained automatically, and he or she selects the most suitable one for each keyword.

In summary, we have developed a semantics-guided approach to discover the possible senses for a set of user keywords, by searching and extracting relevant knowledge from different ontology pools; ontology matching and synonymy estimation techniques are used to merge senses considered similar enough. The main features of our proposal are the following:

1. It uses an iterative approach to retrieve from different knowledge repositories the possible senses of each user keyword, in a parallel manner. A sense is represented basically as the (multi)ontological context of a term, and the system is able to deal with senses corresponding to different kind of ontology terms (classes, properties, and individuals).

2. It considers not only the senses corresponding to ontology terms syntactically matching the user keywords but also the senses of ontology terms matching the synonyms of the user keywords, recursively, in order to semantically enrich the keyword senses retrieved within a certain synonymy threshold.

3. It measures the synonym degree between two senses by considering their linguistic and structural similarity. Statistical techniques like sampling and parallel processing are used to improve the performance of this process.

Some experiments show that our system is able to discover keyword senses not found by traditional Web-based search engines. We believe that this technique to find out the semantic differences between senses (subsets of ontologies) can be applied to many fields. As example, we are currently working on using the retrieved senses to generate queries expressed in a knowledge representation language to retrieve data corresponding to the intended semantics of user keywords.

2.3 Applications

We present two different scenarios in which the ideas developed in the global project have been applied. In the first one a special emphasis has been put on the aspects considered by the subproject 1, and in the second scenario we have insisted more in the aspects developed in the subproject 2.

2.3.1 Health-care Domain

The combination of new advances in sensor technology, PDAs and wireless communications favors the development of a new type of monitoring systems that can provide patients with

given context.

assistance anywhere and at any time. In particular we have developed a monitoring system that has two main components: a user PDA and the Control Center. Each monitored person carries a PDA along with multiple, single or no wireless sensors. The sensors sample physiological data from a person and send them to the PDA. Since the sensors are body-worn and a PDA is a wearable device, *portability* is a key issue for our system. Moreover, another key issue in our approach (and, as far as we know, not yet supported by related works) is *locality* which consists in detecting anomalous situations using the PDA. The other component, the Control Center is in charge of monitoring people. Its main task is to react in response to user's alarms ensuring that appropriate action is taken and to manage the web services which are offered to relatives and physicians. Two are the scenarios in which we have applied the monitoring system.

1. AINGERU. The goal of this system is to provide a tele-assistance service that overcomes the main constraints associated with currently available tele-assistance services using the most recent advances in the fields of Semantic Web, mobile computing and networking.

   Considering those existing tele-assistance services, there are two main aspects that our system improves on. The first issue, that is, the reduced coverage, is overcomed by our system using cellular communications. Technologies such as GSM, GPRS and UMTS give us the possibility to keep the monitored user under surveillance outside his/her home (ubiquity). Referring the second issue, active monitoring, our system incorporates biomedical sensors that send information about vital signs to the PDA. The PDA uses those data to feed a decision support system (that makes use of an ontology), which in turn evaluates the current state of the user. If the user is at risk, the PDA itself notifies the Control Center about the situation and sends data about the user in order to help him to take the proper actions (pro-activity).

2. MOLEC. The goal of this system is to provide an ECG monitoring service that performs local real-time ECG signal classification by detecting possible rhythm irregularities "in situ". Although existing holters present the advantage of allowing patients to live a normal life, they also present a serious drawback: if a person suffers from a serious rhythm irregularity, the holters only record it, i.e. they do not react to it. New commercial monitoring systems and research proposals that react to rhythm irregularities are appearing but they present different restrictions. We have advocated for a system that *performs a local real-time ECG classification* but on a mobile device like a PDA carried by the user because in that way a real anywhere and at any time assistance can be provided.

Both systems, AINGERU and MOLEC, use agents in their implementation and those agents are prepared to deal with semantic communications (techniques developed by Subproject 1)

2.3.2 Software Domain

One of the most frequent tasks of computer users is to obtain new software. Different kinds of users need different kinds of software. Nowadays a common procedure to obtain software is to visit some of the several websites that contain freeware, shareware and demos (such as Tucows, Shareware.com, and Download.com). However, this procedure presents problems for many users and in particular for users of wireless devices because many of them are naive users

(it is growing up the use of Personal Digital Assistants (PDAs) by non computer specialists) and so their knowledge about software is limited. Moreover, the use of wireless media should be minimized during the software selection process due to its high cost and unstability. Tucows-like systems show the same information to all the users and, furthermore, they do not try to minimize the size of data transmitted. We present an alternative procedure: a Software Retrieval Service (SRS), based on the use of an ontology and the agent technology, that allows users to find and retrieve software in an easy, efficient and adaptive way, anywhere and anytime. *Easy*, because with the help of knowledge-driven agents, users can browse locally the ontology that describes semantically the content of a set of data sources containing pieces of software, and so they can select the software from it (the service makes transparent for the users the technical features of their devices and the location and access method of various remote software repositories); *efficient*, because by taking advantage of the capability of mobile agents to deal with disconnections and to move to other computers, agents optimize the use of the wireless media by customizing software catalogs to the user needs; and *adaptive*, because agents take into account the network status (to decide the communication method and the amount of data transferred), past user actions (to anticipate future user requests), and they also learn of their own mistakes (wrong decisions).

The specific contributions associated with this service are the development of the SoftOnt Ontology and the development of the following four kinds of autonomous agents, which react to the context in order to achieve a better performance:

- Alfred, which is able to manage two types of knowledge (independent of the service and related to the service) and probabilistic information of user answers in order to serve the user the best it can. Hence it automatically complements user requests with restrictions concerning the user device and the preferences of such a user.

- The Software Manager agent, which obtains a software catalog customized for the user, according to 1) the constraints specified by the user and Alfred, and 2) the current network speed (the higher speed, the higher level of detail of the software catalog).

- The Browser agent, which adapts to the behavior of the user with the purpose of anticipating to his actions and so helps him to find the wanted software in the presented catalog as soon as possible.

- The Catalog Updater agent, which autonomously selects on the fly the best strategy to update the software catalog presented to the user, by taking into account the current network status.

  The previous agents optimize the use of wireless communications, and deal with network failures autonomously by applying a retry policy.

The user provides Alfred with a list of keywords concerning the kind of software he or she is looking for. Alfred uses the techniques developed in the Subproject 2 (see Section 2.2) to enrich user keywords semantically and then send them to the Software Manager agent who obtains a customized software catalog for that user.

3 Results

Journal Papers

- "Interoperation among Agent Based Information Systems through a Communication Acts Ontology", J. Bermúdez, A. Goñi, A. Illarramendi, M.I. Bagües, Information Systems Journal (en revisión) (2007) (JCR 1.027)
- "Is an ontology based on the Speech Acts Theory adequate for formal agent communication?", J. Bermúdez, A. Goñi, A. Illarramendi, S. Santini, Apllied Ontology Journal (2007) (en revisión)
- "A Software Retrieval Service based on Adaptive Knowledge-Driven Agents for Wireless Environments", E.Mena, A. Illarramendi, J.A. Royo, A. Goñi, ACM Transactions on Autonomous and Adaptive Systems (TAAS) Vol.1, 67-90 (2006) (revista nueva)
- "Location-Dependent Queries in Mobile Contexts: Distributed Processing Using Mobile Agents", S. Illarri, E.Mena, A. Illarramendi, IEEE Transactions on Mobile Computing Vol.5 N. 8, 1029-1043 (2006) (JCR 3.034)
- "Semantic Interoperation among Data Systems at a Communication Level", M.Bagües, J. Bermúdez, A. Illarramendi, A. Tablado, A. Goñi, Journal on Data Semantics Vol. 6 1-24 (2006) (revista nueva)
- "Real-Time Classification of ECGs on a PDA", J. Rodríguez, A. Goñi, A. Illarramendi, IEEE Transactions on Information Technology in Biomedicine Vol. 9 N. 1 23-34 (2005) (JCR 1.376)

International Conference Papers Published in Books

- "3D Monitoring of Distributed Multiagent Systems", S. Ilarri, J.L. Serrano, E. Mena and R. Trillo, International Conference on Web Information Systems and Technologies (WEBIST'07), Barcelona (Spain), (INSTICC Press), March 2007.
- "Performance Analysis of an Adaptive User Interface System Based on Mobile Agents", N. Mitrovic, J.A. Royo and E. Mena, Engineering Interactive Systems. Conference on Design Specification and Verification of Interactive Systems. 14th DSV-IS Conference Salamanca (Spain), (Springer Verlag Lecture Notes in Computer Science LNCS 2844), March 2007.
- "Discovering and Merging Keyword Senses using Ontology Matching", M. Espinoza, R. Trillo, J. Gracia and E. Mena, 1st International Workshop on Ontology Matching (OM-2006) @ 5th International Semantic Web Conference ISWC-2006, Athens, Georgia (USA), (CEUR-WS), November 2006.
- "Querying the Web: A Multiontology Disambiguation Method", J. Gracia, R. Trillo, M. Espinoza and E. Mena, Sixth International Conference on Web Engineering (ICWE'06), Palo Alto, California (USA), ACM , July 2006.
- "SPRINGS: A Scalable Platform for Highly Mobile Agents in Distributed Computing Environments", S. Ilarri, R. Trillo and E. Mena, 4th International WoWMoM 2006 workshop on Mobile Distributed Computing (MDC'06), Buffalo, New York (USA), (IEEE), June 2006.
- "Discovering the Semantics of Keywords: An Ontology-based Approach", M. Espinoza, J. Gracia, R. Trillo and E. Mena, The 2006 International Conference on Semantic Web and Web Services (SWWS'06), Las Vegas, Nevada (USA), (CSREA Press), June 2006.
- "Using decision trees for real-time ischemia detection", L. Dranca, A. Goñi, A. Illarramendi, CBMS 2006. The 19th IEEE International Symposium on Computer-Based Medical Systems. (IEEE) (2006)

- "An innovative system that runs on a PDA for the continuous monitoring of people", M I. Bagüés, J. Bermúdez, A. Burgos, A. Goñi, A. Illarramendi, J. Rodriguez, A. Tablado, CBMS 2006. The 19th IEEE International Symposium on Computer-Based Medical Systems. (IEEE) (2006)
- "Processing of Data Streams with Prediction Functions", S. Illarri, O. Wolfson, E. Mena, A. Illarramendi, N. Rishe, 39th Hawaii International Conference on System Sciences (HICSS-39) (IEEE)(2005)
- "Adaptive User Interfaces Based on Mobile Agents: Monitoring the Behavior of Users in a Wireless Environment", N. Mitrovic, J.A. Royo and E. Mena, I Symposium on Ubiquitous Computing and Ambient Intelligence (UCAmI'2005), Granada, (Spain), (Thomson), September 2005.
- "MOLEC: Anywhere and at Any Time Arrhythmia Classification", J. Rodríguez, A. Goñi, L. Dranca, A. Burgos, A. Illarramendi, Proceedings of the 32nd International Conference "Computers in Cardiology".(IEEE) (2005)
- "An On-Line Ischemia Monitoring Algorithm for Mobile Devices", L Dranca, A Goñi, J Rodríguez, A Burgos, A Illarramendi, Proceedings of the 32nd International Conference "Computers in Cardiology Lyon, France (IEEE)(2005).
- "Searching the Web: From Keywords to Semantic Queries", J A. Royo, E.Mena, J. Bernad, A. Illarramendi, Third International Conference on Information Technology and Applications (ICITA'05), Sydney, 244-29 (IEEE)(2005)
- "An Intelligent System for Assisting Elderly People", A.Tablado, A.Illarramendi, M.I. Bagües, J. Bermúdez, A.Goñi, Foundations of Intelligent Systems. ISMIS 2005 (Springer Verlag LNAI) (2005)
- "A Wireless Application that Monitors ECG Signals", J. Rodriguez, A. Goñi, L., A. Illarramendi, Enterprise Information Systems VI. (Springer Verlag) 2005

International Conference Papers Published in Proceedings

- "A communication acts ontology for agent interoperability", Jesús Bermúdez, Alfredo Goñi, Arantza Illarramendi, Miren I. Bagüés, Alberto Tablado, FOCA. Formal Ontologies for Communicating Agents
- "Aingeru: an application that takes profit of semantic web technologies at a PDA", M.Bagües, J. Bermúdez, A.Goñi, A. Illarramendi, A. Tablado, 2nd European Semantic Web Conference Heraklion. Grecia

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

ADACO: ADAptive and COllaborative web based systems

TIN2004-08000-C03

Pascual González López[1]
University of Castilla-La Mancha

Francisco L. Gutiérrez Vela[2]
University of Granada

Toni Granollers Saltiveri[3]
University of Lerida

Abstract

Collaborative systems have emerged as one of the most relevant future working environments. These systems offer a working way that allow new human cooperation strategies based on new interaction paradigms, which take into account not only individuals, but also groups. Due to the relevance of this new collaborative systems usage, it is important to provide new solutions to the construction of robust and complete collaborative systems. The starting point is the description of the system by using ontologies at different abstraction levels. Parts of this description will be connected with the architectural design of groupware applications. From the user interaction point of view, we propose the inclusion of new interaction paradigms (ubiquitous computing, augmented reality, etc.) within the development process of new usability and accessibility engineering methodologies. In addition, new interaction patterns which support the user interface design process and multi-agent systems that support the user interface adaptativity of the collaborative environment are included among other things.
This document reports the results of the first twenty months of the project life.

Keywords: Collaborative environments; user interaction design; methodologies and models; adaptative user interfaces; new interaction paradigms

1 Project Purposes

This project aims to contribute new solutions to the creation of web-based collaborative environments. The subject will be approached from different viewpoints. On one hand, will be outlined the architecture supporting the system, as well as its patterns, techniques and tools, which would enrich the current development of software engineering. On the other hand, and from users' interaction perspective, including new mechanisms of usability and accessibility engineering, will be identified new interaction paradigms (ubiquitous computing, augmented reality, etc.), interaction standards supporting the process of users' interfaces and multicast systems design, which would maintain the interfaces adaptivity of collaborative environments.

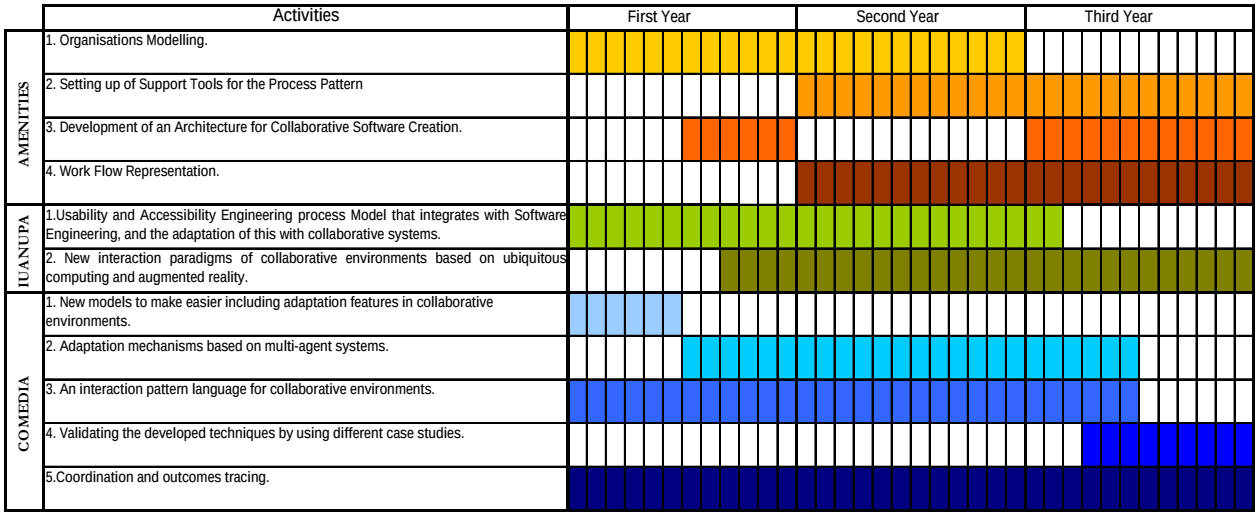
[1] Email: pascual.gonzalez@uclm.es
[2] Email: fgutierr@ugr.es
[3] Email : tonig@griho.net

In order to reach the expected outcome, it is compulsory to coordinate the contributions of the different teams involved in this proposal. The reasons of this organisation are based on taking advantage of each group's experience, which contributes different but complementary views and knowledge.

The specific objectives of each subproject are the following:

- To define an adaptive and evolutionary system architecture whose purpose is to facilitate collaborative systems design.
- To obtain a representation system of the workflow through hypermedia webs.
- To incorporate usability and accessibility engineering in the framework of development methods of web-based collaborative environments, in order to improve the achieved systems quality.
- To incorporate new interaction standards into web-based collaborative environments.
- To portray models, which catch the new systems adaptivity characteristics and to create a multicast system able to offer adaptability in web-based collaborative systems.
- To define a language of interaction patterns, which would enclose the implicit knowledge conveyed by the interface designers of this sort of web-based collaborative systems.
- To set up aid tools for the development of a collaborative software.
- To choose several case studies, with different levels of difficulty, in order to apply and corroborate the previous advancements.

These points have been developed through three subprojects. The suggested chronogram of activities is indicated herein and presented separately by each subproject. Every activity (highlighted in a different colour) corresponds to each sub-objective proposed in each subproject. It is interesting to stand out that, although the responsibility of each subproject falls in one of the research groups participating, in some of their tasks, members of other teams collaborate, and in those activities or tasks strictly related to, joint work teams are created, who hold specific sessions in the meetings planned in the frame of the coordination task. These two annual coordination meetings have been fundamental to the right development of the project. The first two (one in Albacete, in February 2005 and the other in Granada, in September 2005) gave the team the opportunity to know each other and laid the foundations for a teamwork. To the above mentioned meetings were invited well-known researchers (Dr. Fabio Paternò, from the CNR in Pisa, Italy, and Dr. Gustavo Rossi from La Plata University, in Argentina) to contribute their vision of the problem and of the project purposes and tasks. The other two meetings (in Lerida, in April 2006 and in Sitges, in October 2006) were of use to set up the work teams and to outline combined actions about common topics to the different teams, such as the ontology of collaborative pattern, usability and collaboration, adaptation, plasticity and new standards, etc.



2 Project Outcome

Hereinafter are detailed the contributions and the current state of the different activities related to the three subprojects.

Subproject 1 (UGR). Analysis and modelling of collaborative systems (AMENITIES)

- **Organisations Modelling.**

To carry out the activity of ontologic description of a CSCW system, the conceptual frame outlined in AMENITIES [3] has been taken into account in order to know which elements should be considered in the process of description of a collaborative system. This conceptual frame has been studied, analysed and applied in the modelling of concrete systems and has given place to a domain ontology including the main components, which constitute such a system. Starting from this ontology, application ontologies are obtained for concrete collaborative systems. The suggested design multilevel diagram [33] eases the change processes specific to this type of systems and allows the reusing of elements described in other ontologies. As a practical case study, we have chosen the description of cooperation in business processes, to which participate various enterprises/organisations [34].

Different metadata languages have been examined in order to formalise the suggested ontologies. Furthermore, different standard languages are being considered to use them in the classification of restrictions and questions on suggested domain and application ontologies.

A UML profile has been outlined and its purpose is to introduce the development of collaborative applications based on AMENITIES and under MDA standard. The different patterns semantics and the link among them have also been established [32].

The communication, coordination and cooperation processes, usually carried out in a collaborative system, have been analysed with the intention of looking for patterns, which can initially facilitate the system modelling and afterwards help its development. As a result of this study, and due to the UML language constraints, on a first stage we have developed a UML profile for patterns representations, then applied it to the definition of different organisation patterns, activities and

interaction [15] and presently we are working on the patterns introduction to the methodological frame of AMENITIES [14].

- **Setting up of Support Tools for the Process Pattern.**

This activity is greatly based on the outcome obtained from the other activities; for this reason its accomplishment has been envisaged towards the last months of the project. Currently, we have begun to create some designs whose final aim is to implement the setting up and management tools of the patterns previously suggested in the frame of methodological processes.

- **Development of an Architecture for Collaborative Software Creation.**

The works related to the architecture have been dealt with from different research tracks, trying to cover several but complementary issues.

On one hand, the systematic development of groupware applications has been approached with a concrete architectural proposal, which meets certain quality standards (extensibility, transferability, reusability,...) [5], as well as its implementation by way of concrete middelware [4]. A proposal based on an ontology has been suggested in order to estimate the quality of the different architectures recommended for the setting up of collaborative systems. All these contributions aim is to be introduced to the proposed methodological frame.

On the other hand, an architecture based on the technology of Web services has been suggested [1]. In it the main elements in charge of communication, coordination and collaboration are implemented by means of services. We are currently working on the introduction of adaptation (and/or evolution) to suggested architectures.

Lastly, we are presently working in the inclusion of security aspects, since the first stages of the system development. Thus, we have developed a part of the architecture in charge of the administration of the access control to resources and activities by the functional elements of the architecture [11]. We have also carried out an extension to the pattern of the RBAC access control, in order to incorporate the dynamism necessary to a complex organisation in which collaborative activities are performed [12].

- **Work Flow Representation.**

We have suggested a cooperative system specification through an evolutionary hypermedia pattern based on a SEM-HP extension. We have developed a systematic process allowing to obtain a multi-level hypermedia structure, starting from the early outlines of an organisation, in terms of structure, tasks, activities, resources and objectives. We are working on the use of different navigation proposals delineated in SEM-HP pattern [22].

In the frame of the developed architecture, a Web service of tasks management has been considered. It allows the coordination of the workflow to be accomplished by the different system actors, via the placing into the system of information about the tasks to be carried out and on their restrictions.

We are approaching the creation of a set of evolutionary mechanisms allowing to execute reliable and flexible changes to a certain level of the hypermedia structure [21]. Several modifications to the JSEM-HP tool [23] have been applied in order to introduce some of these features.

- **Practical Application of the Project.**

As a practical application of the techniques and tools developed during the project, the results obtained in different earlier tasks have been applied. A system of augmentative communication for autistic children [35], based on the use of hypermedia techniques for the representation of the communication structures and adaptation techniques has been created, with the purpose of facilitating and/or restricting information navigation [13].

Subproject 2 (UdL). Formalization of the user centred process model applied to interactive cooperative systems based on web and its extension to new paradigms (IUANUPA)

- **Usability and Accessibility Engineering process Model that integrates with Software Engineering, and the adaptation of this with collaborative systems.**

A model of our own, for usability and accessibility engineering to develop interactive systems, has been proposed. The model, named MPIu+a, has been extensively tested in real cases. The detailed description of the process model has been the main task carried out during this activity. It serves as a base for the implementation of interactive systems under the principles of usability, accessibility and the user centred design, and it also integrates with the current software engineering standards [6]. In this section emphasis on seeing which activities are specific for the collaborative models of computation has been done. At this point, the main difficulties were about knowing the details that characterize collaborative contexts and, mainly, in the definition of points of contact between these collaborative contexts and usability attributes defined for traditional single-user interactive systems. In this aspect, usability guidelines and design strategies for collaborative systems have been proposed, mainly bound to Web context using the MPIu+a model [8, 9].

Moreover, a deep study of current evaluation techniques and requirements engineering and validation techniques has been made. At this point, special mention deserve the works related to optimizations with walkthroughs techniques or the experimentation carried out with the use of techniques coming from the scope of datamining to improve usability qualitative evaluation [9]. The experience allowed us to propose an adaptation for some evaluation techniques for collaborative environments (the characteristics of those systems makes difficult to apply the "traditional" usability evaluation techniques) [10].

- **New interaction paradigms of collaborative environments based on ubiquitous computing and augmented reality.**

This task has been carried out under two different perspectives, aimed at obtaining a wider view of the problem.

In the first case, a deep study of the methodologies used to develop virtual reality applications has been done, including methodologies that allow us to create inhabited environments in which several users can collaborate, and finally proposing a new methodology aimed at the design of 3D user interfaces, mainly virtual environments (TRES-D) [25]. Besides, different interaction techniques, which facilitate the manipulation and selection of objects in 3D environments, have been analyzed. In addition to that, several applications have been developed allowing us to evaluate an interaction model [20] for the design and implementation of the interaction in virtual worlds, especially also those that are multi-user. And finally, the use of a virtual environment, as a way to simulate the migration of user's interfaces among platforms, has been analyzed [26], offering also the possibility for this environment to be shared by several users.

The second approach, which is close to ubiquitous systems, a new perspective related to mobility and new devices has been analysed, looking for some techniques that make the adaptation to the context easier. The context is conceived as a combination of features from the user, the environment, the device, and finally from the working group restrictions and particularities. In the end, this adaptation also takes into account the quality characteristics defined by a set of usability properties. Therefore, we can refer to it as plasticity [39]. Regarding this point, a framework that incorporates plasticity criteria has been defined [41]. This framework includes an engine that solves the anticipation to foreseeable contextual changes in a proactive manner, which is called Implicit Plasticity Engine [42]. Among these changes the detection of the group dynamics in an automatic way is worth mentioning, trying to foster collaboration [40]. At the moment, and starting from these works, an extension of this framework to ubiquitous systems is being worked. Other further

works are: the formalization of the group shared-knowledge modelling; the development of an integrating conceptual framework to be used as a reference for the generation of tools aimed at supporting the collaborative software development.

Subproject 3 (UCLM). Collaborative Environments: Adaptive User Interfaces Development Methodologies (COMEDIA)

- **New models to make easier including adaptation features in collaborative environments.**

One of the main problems usually found when several people come together to work in collaboration is using a common language. Therefore, in this task different models have been devised to gather context and group characteristics, and those characteristics derived from collaboration. The first one, the context model [19], encompasses a user model, a platform model, an environment model and the current task the user is carrying out. These models collect those characteristics considered to be relevant for the adaptation processes proposed.

In the same way, some new approaches to model the organizational structure of the elements of a collaborative system have been proposed. These models provide an explicit representation, taking into account concepts such as: group, actor, role, user, agent, etc [38]. Furthermore, a user interaction relationship model has been proposed for collaborative systems. It does not reflect the interaction of user with the system, but the interaction between the users by means of machines and by using the network infrastructure [36]. These contributions are aimed at the development of a methodology [37] for the analysis and design of collaborative environments centered in the roles played by the users of the system, and driven by the tasks they need to perform in order to achieve the goals pursued, which lead to the consecution of the purpose the groupware application was implemented for. These models are the foundations for the creation of a development environment allowing the creation of adaptive and collaborative systems.

- **Adaptation mechanisms based on multi-agent systems.**

When a designer wants to include adaptation facilities in an application, one of the main problems encountered is how to do that in a way so the resources spent in the design of the adaptation capabilities, and the knowledge captured during the creation process can be reused. To do so, applying an engineered approach is required, which supports the extraction of the adaptation mechanisms, usually hardcoded directly into the applications, and moves it into a generic working framework. To tackle this problem in this project a multi-agent system has been developed allowing the management of the adaptation capabilities in a proper and reusable manner [18]. To provide the services required by an adaptative application the multi-agent system uses a graph grammar transformation approach to manipulate an XML-based user interface representation. This XML-based representation is described by using a user interface description language called usiXML. The team in this subproject is actively contributing to the definition of usiXML in collaboration with professor Jean Vanderdonckt and his team [16, 24, 27].

One of the greatest concerns in any adaptative system is the potential degradation of the system usability [17], and therefore of its quality, because of the adaptations applied. In the multi-agent system designed two usability control mechanisms have been created to control the usability of the user interfaces produced by means of adaptation. On the one hand, a set of metrics based on cognitive theory have been created that enable the evaluation of the potentially suitable rules for each situation to produce a ranking of the rules representing the quality of the rules. On the other hand, a quality model has been created[19, 29], based on goal-driven requirements specification

techniques, allowing taking a decision about to what extend the user interface generated in the adaptation process preserves a set of usability criteria previously specified by the designers.
One of the key issues we will be working on during this last year of the project is increasing and improving the number of usability criteria supported by the system, providing new metrics to assess each one of the criterion properly.

- **An interaction pattern language for collaborative environments.**

User interface development has been traditionally tackled in a non-systematic manner. Because of that, one of the objectives proposed in this project was the need to identify a way to document the experience related to user interface design by means of patterns, which have been successfully used in the field Software Engineering and Human-Computer Interaction.
The ways patterns are currently documented do not provide directly many advantages compared with guidelines, with respect of the manner they are used in practice. Because of this identified limitation, we have related pattern and model concepts, so now each pattern, along with the usual fields used to describe it, includes a set of models (mostly a domain and a task model) that describes the solution of each pattern properly documented [28].
On the other hand, by using the domain and task models specification, it is possible to derive, by means of a transformation and model-mapping, other models such as dialog or abstract user interface models. Later on, from this specification in the abstract level, by applying successive refinements, it is possible to obtain, in a systematic way, a concrete user interface. The specification of the user interface at the various levels of abstraction is achieved by using some of the models proposed for usiXML. A tool has been developed (IDEALXML) [30, 31] that allows managing the documented experience and using it in the development of interactive systems.
Extending this working philosophy to the specification and development of user interfaces for collaborative environments is almost straightforward, because the development of collaborative user interfaces is not different to the development of other user interfaces, except for the need to consider additional mechanisms related to coordination, collaboration and information sharing. At this moment, we are working on including these characteristics in the initial environment proposed.

- **Validating the developed techniques by using different case studies.**

The objective of this task is the definition and development of different prototypes that allow the validation of some of the ideas proposed, analyzing the pros and cons found. This task, although scheduled for this last year, has been made in some cases at the same time that the ideas where pushed [2]. On the other hand, it has fostered the analysis of different environments where collaborative and adaptative systems are especially relevant, such as e-learning environments [7], information system design environments or virtual collaborative environments. This task is raising new challenges, which have brought about a project proposal specifically centered on providing solutions for these kinds of environments.

3 Outcome Indicators

Next, we detail some indicators that show the relevance of the results achieved up to now. These indicators reflect training activities of the researchers, publications, relationships with the socio-economical context and other diffusion activities.

3.1 Students training

Firstly, it is important to highlight two PhD. Theses with European mention defended in the framework of this Project [16] [24]. Both of them got the maximum qualification.
Besides, other 8 PhD. Theses are predicted to be defended during 2007: María Paula González, *Integration of Knowledge Discovery Techniques in Databases within the Process Models of Usability Engineering.* Directed by: Dr. Toni Granollers & Dr. Jesús Lorés, UDL, March 2007; José Pascual Molina, *Three-dimensional User Interface Development. A Methodological Approach: TRES-D.* Directed by: Dr. Pascual González, UCLM, May 2007; José Luis Isla, *Conceptual Modelling of Cooperative Systems based on Patterns*, Directed by: Dr.Francisco L. Gutierrez & Dr. José Luis Garrido, UGR, June 2007; Manuel Noguera, *Modelling of CSCW Systems based on ontologies and its implementation by means of architectures.* Directed by: Dr. José Luis Garrido, y Dra. María V. Hurtado, UGR, Sept. 2007; Victor M. R. Penichet, *A User-Centred and Task-Driven Methodology for the Development of Collaborative Systems.* Directed by: Dr. José Antonio Gallud & Dra María Dolores Lozano, UCLM, Sept. 2007; Montserrat Sendín, *Software infrastructure to support plasticity of the user interfaces in mobile platforms under a dychotomic view.* Directed by Dr.Cesar Collazos & Dr. Jesús Lorés, UDL, Sept. 2007; Marcelino José Cabrera, *User Interfaces Adaptation*, Directed by Dr. Francisco L. Gutiérrez, UGR, Sept. 2007; Mario Anzures, *Adaptable and adaptive architectural model to develop groupware applications based on SOA.* Directed by: Dr. Miguel J. Hornos & Dra. Patricia Paderewski, UGR, Sept. 2007.
Finally, seven DEA degrees have been defended during the last two years and the defence of other seven is predicted by the middle of 2007.

3.2 Publications

The level of publications derived from the project has been very high. Perhaps, the main point to improve is the publication in specialized journals that is the goal of this last year as the accepted but pending of publication articles show. Some additional papers have been submitted but they are still in the reviewing process so they have not been included in this document. A resume of all accepted publications is shown in the next table:

	Journal JCR	Journals	LNCS/LNAI	Book chapters	Conferences	PhD
National				2	23	2+8
International	3	3	16	14	37	

This research activity has been reflected in different national and internationals reference congresses. We remark next some of them in which there is an active participation of some researchers who underwrite this project. In the international scope, we can highlight some ACM conferences such as. Web3D, IUI, CADUI, DSV-IS,ICMI, VRST, etc; IEEE conferences such as: CSCWD, MELECON, WBC, ICALT, etc; IFIP conferences such as Interact, and other international conferences such as CEEMAS, HCI-International, WEBIST, ICEIS, Meto-VR, CRIWG, VODCA, WBC, PSMD, SERP, EUROCAST, ICUC, etc. In the national scope: JISDB, INTERACCION, DSOA, SIHICA, JCSD, etc.
The impossibility to gather all the publications in this document obliges us to include only the ones we consider more relevant. To get a complete vision of all the publications, you can consult the project web page: http://adaco.dynalias.org.

3.3 Collaboration with other research groups

Currently, the research groups collaborate in the following Knowledge Networks and other European Actions related to the topics tackled in this project:

- European Network of Excellence "SIMILAR". http://www.similar.cc/
- European Action COST 294 "MAUSE".
- ERCIM Working Group on Software Evolution.
- National Research Network "Desarrollo Software Dirigido por Modelos (DSDM)".
- National Research Network "Creación de un entorno innovador para la comunicación de agentes inteligentes (AgentCities)".

On the other hand, four research stays, two doctoral and two postdoctoral, have been performed in prestigious research centres such as the Institute of Information Science and Technologies of the Italian National Research Council or the Department of Computer Science, University College London, UK.

Besides, Dr. Cesar Collazos from University of Cauca (Colombia) has performed a research stay in the University of Lerida as invited professor during the academic course 2005/2006.

In addition to this, some prestigious researchers have done some visits or short stays in our laboratories. For instance, Dr. Jean Vanderdonckt (Université Catholique de Louvain, Belgium), Dr. Fabio Paternò (National Research Council, Italy), Dr. Gustavo Rossi (Univ. Nacional de la Plata, Argentina), Dr Anthony Steed (University College London), Dr. Clive Fencott (University of Teesside, UK), etc.

3.4 Relationship with the socioeconomic context

In this context, we want to highlight the execution of different projects with enterprise participation. For instance, the application of the project to the augmentative communication system for children with autism has derived in the concession of an award to the innovation from the Spanish Association of Autism Professionals (AETAPI).

Other projects or relevant applications are: MyMobileWeb identified as "CELTIC flagship Project" CP04-020, regarding advanced technologies enabling multi-device mobile access to current and future web applications, services and information portals, coordinated by Telefónica I+D; ADONIS submitted to the CENIT call together with Atos Origin, Telefónica I+D, etc..

3.5 Other diffusion activities

Finally, several activities regarding the organization of conferences and workshops have been carried out in the scope of this project, apart form imparting some lecturers and postgraduate courses. Next, we detail some of these activities:

- Organization of conferences and workshops: *I Congreso Español de Informática CEDI'2005*; *X Jornadas sobre Ingeniería del Software y Bases de Datos, JISBD'2005*; *VI Congreso Internacional de Interacción Persona-Ordenador, 2005*; *1er Taller Tendiendo Puentes entre la Interacción Persona-Ordenador (IPO) y la Ingeniería del Software (IS)*; *XIII Jornadas de Concurrencia y Sistemas Distribuidos*; *III Taller de Sistemas Hipermedia Colaborativos y Adaptativos. CEDI 2005*; *IV Taller de Sistemas Hipermedia Colaborativos y Adaptativos. JISBD 2006*; Special sessions in *INTERACT'07* and *CASYS'07*; *CADUI 2008*.
- Invited lecture and organization of a Usability workshop in the *3er Congreso Nacional de Investigación y Tecnología, en las áreas de Usabilidad, Realidad Virtual y Robótica,* 2006, Colombia.

- Organization of several postgraduate courses in different foreign universities, such as: Universidad Nacional de La Plata (Argentina). June 2006; University of Petrosani (Romania) in October 2006; University of Magallanes, Punta Arenas, Chile in November, 2005 and a doctoral course organized by Universidad de Granada and Universidad de Cuba.
- Participation in the *Semana de la Ciencia de Andalucía* and *Semana de la Ciencia de Castilla-La Mancha* with the exhibition of several works.

4 Acknowledgements

We do not want to finish this document without a special mention to the memory of Dr.Jesús Lóres (main researcher of the subproject of the University of Lerida) who passed away due to illness at the end of 2006.

5 References

[1] Anzures, M., et al, Development of Extensible and Flexible Collaborative Applications Using a Web Service-Based Architecture. *International Workshop on Rapid Integration of Software Engineering techniques (RISE'06)*. LNCS, Springer, pp. 67-82, 2006. (to appear)

[2] García,F.,Lozano,M.,et.al. A controlled experiment for measuring Usability of webapps using Patterns. *Enterprise Information Systems VII*. Springer-Verlag. pp. 257-265, 2006.

[3] Garrido, J.L., Gea, M., et al. Requirements Engineering in Cooperative Systems. *Requirements Engineering for Sociotechnical Systems*. IDEA GROUP, USA. pp.226-244, 2005.

[4] Garrido, J.L., et. al. Leveraging the Linda Coordination Model for a Groupware Architecture Implementation.*Groupware:Design,Implementation and Use*. LNCS,Vol.4154,Springer,pp.286–301, 2006

[5] Garrido, J. L.,., et al. A Software Architecture Intended To Design High Quality Groupware Applications. *International Conference On Software Engineering Research And Practice*, pp. 59-65, 2005.

[6] Granollers, T., Lorés, J. Incorporation of users in the Evaluation of Usability by Cognitive Walkthrough. *HCI related papers of Interacción*. Springer.Verlag, pp. 243-255, 2006.

[7] Gascueña, J.M., Fernández-Caballero, A., González, P. Ontologies for student and domain models in an adaptive and collaborative learning system. *Research on Computing Science*, 17. pp. 33-42, 2006.

[8] González, M.P., et al. A Hybrid Approach for Modelling Early Prototype Evaluation under User-Centred Design through Association Rules. Proc. DSV-IS 2006. LNCS, Springer. (to appear)

[9] González, M.P., Granollers, T., Lorés, J. Detecting Usability Problem Patterns for a Context of Use under Qualitative Usability Testing with Datamining Techniques. *Information and Software Technology*. Elsevier. (to appear)

[10] González, M.P., et al. Guidelines and usability principles to design and test Shared-Knowledge Awareness for CSCL. *Groupware:Design,Implementation and Use*. LNCS,Vol.4154,pp.102-117,2006.

[11] Gutiérrez, F.L., Isla, J.L., et. al. An Architecture for Access Control Management in Collaborative Enterprise Systems Based on Organization Models, *Special Issue of the International Journal of Science of Computer Programming*, Elsevier, 2007. (to appear)

[12] Gutiérrez, F.L., et al. Organization Modelling to Support Access Control for Collaborative Systems, *Proc. International Conference on Software Engineering Research and Practice*. pp. 757-763, 2006

[13] Hurtado, M.,et al,. An Augmentative Communication System Based on Adaptive Evolutionary Hypermedia Systems. EUROCAST. Lecture Notes in Computer Science, Vol. 3643, pp. 223-228,2005.

[14] Isla, J.L., et al. Integration of Organisational Patterns into a Group-Centred Methodology. In *HCI related papers of Interacción*. Springer, pp. 137-146, 2006.

[15] Isla, J.L., et al. Social Organization Modelling in Cooperative Work Using Patterns. Computer Supported Cooperative Work in Design, LNCS Vol. 3865. pp 112-121, 2006.

[16] Limbourg, Q., et al. UsiXML: a Language Supporting Multi-Path Development of User Interfaces. *Eng. Human Computer Interaction and Interactive Systems*. LNCS, Vol.3425, Springer, pp.200-220, 2005
[17] López Jaquero, V., et al. A Seamless Development Process of Adaptive User Interfaces Explicitly Based on Usability Properties. *Eng. Human Computer Interaction and Interactive Systems*. LNCS, Vol.3425, Springer, pp.200-220, 2005.
[18] López Jaquero, V., et al. A Multi-Agent System Architecture for the Adaptation of User Interfaces. *Multi-Agents Systems and Applications IV*. LNAI 3690, Springer, pp.583-586, 2005.
[19] Lopez Jaquero, V. *Interfaces de Usuario Adaptativas Basadas en Modelos y Agentes Software*. PhD thesis. University of Castilla-La Mancha, 2005.
[20] Martínez, D.,García, A. S. et al. Towards an interaction model for CVEs. 12th International Conference on Human-Computer Interaction HCII'07. LNCS, Springer. 2007. (to appear)
[21] Medina, N., Molina, F., et al. Diversity of Structures and Adaptive Methods on an Evolutionary Hypermedia System, *IEE Proceedings–Software*,Vol.152,Issue3,pp.119-126, 2005.
[22] Medina, N., et al. An Adaptation Method by Feedback in an Evolutionary Hypermedia System. The Sixth International Conference On Web Engineering, ACM Press, pp. 161 - 168, 2006.
[23] Molina, F., et al. JSEM-HP: A Tool for the Development of Adaptive Evolutionary Hypermedia Systems, Proc. IADIS International Conference on Applied Computing'06, IADIS Press, 2006.
[24] Molina, J.P., Vanderdonckt, J., et al. Rapid Prototyping of Distributed User Interfaces. Chapter 12. *Computer-Aided Design of User Interfaces V*, Information Systems Series, Springer, pp. 151-166, 2006.
[25] Molina, J.P., et al. The development of glove-based interfaces with the TRES-D methodology. *13th ACM Symposium on Virtual Reality Software and Technology*. ACM Press, pp. 216-219, 2006.
[26] Molina,J.P., Vanderdonckt, J., Gonzalez, P., Direct manipulation of User Interfaces for Migration. *10th ACM Int. Conf. on Intelligent User Interfaces*, ACM Press, New York,. Pp. 140-147, 2006.
[27] Molina, J.P., Vanderdonckt, J., Montero, F., González, P. Towards Virtualization of User Interfaces based on UsiXML. Proc. of Web3D 2005 Symposium, ACM Press. pp. 169-179, 2005.
[28] Montero, F. *Integración de calidad y experiencia en el desarrollo de interfaces de usuario dirigido por Modelos*. PhD thesis. University of Castilla-La Mancha, 2005.
[29] Montero, F., Lozano, M., González, P. Usability-Oriented Quality Model based on Ergonomic Criteria. *Handbook of Research on Web Information Systems Quality*. IDEA GROUP Inc. (to appear)
[30] Montero, F., et al. IdealXML: An Interaction Design Tool and a Task-based Approach to User Interface Design. Chapter 20, *Computer-Aided Design of User Interfaces V*, Springer, pp. 245-252, 2006.
[31] Montero, F., López Jaquero, V., Vanderdonckt, J., et al. Solving the Mapping Problem in User Interface Design by Seamless Integration in IdealXML. Interactive Systems Design, Specification and Verification (DSV-IS'2005), LNCS 3941, Springer-Verlag, pp. 161-172, 2005.
[32] Noguera, M., et al.Use of Computation Independent Models in an MDA-Based Groupware Development Process. *Journal of Science of Computer Programming*, Elsevier, 2006. (to appear)
[33] Noguera, M., et al. An Ontology-Based Scheme Enabling the Modeling of Cooperation in Business Processes. On the Move to Meaningful Internet Systems, LNCS Vol. 4277, pp. 863–872, 2006.
[34] Noguera, M., et al. System Modeling for Systematic Development of Groupware Applications. *International Conference On Software Engineering Research And Practice*. pp. 750 - 756, 2006.
[35] Paredes, M.D., et al. A Platform For Creating Adaptive Communicators.Computers Helping People with Special Needs. LNCS, Vol. 4061. Springer-Verlag. pp. 847-854, 2006.
[36] Penichet, V.M.R., Paternò, F., et al. Collaborative Social Structures and Task Modelling Integration. Proceedings DSV-IS'06. LNCS Springer Verlag. (to appear)
[37] Penichet, V.M.R., et al. A Classification Method for CSCW Systems. the Second International Workshop on Views On Designing Complex Architectures. *Electronic Notes in Theoretical Computer Science*, Elsevier, pp.211-220, 2007.
[38] Penichet, V.M.R., Lozano, M., et al. Describing Group Tasks in Multi-User Systems Proceedings of the 4th Latin American Web Congress-La-Web 2006,IEEE Computer Society Press,pp.101-104, 2006
[39] Sendín, M., Lorés, J. Local Support to Plastic User Interfaces: an Orthogonal Approach. In *HCI related papers of Interacción*. Springer.Verlag, pp. 163-178, Feb. 2006.
[40] Sendín,M., et al.Implicit Plasticity Framework:a Client-Side Generic Framework for Collaborative

Activities. *Groupware:Design,Implementation and Use*. LNCS, Vol.4154, Springer, pp 219-227, 2006.
[41] Sendín, M., Lorés, J. Towards the Design of a Client-Side Framework for Plastic UIs using Aspects. *International Workshop on Plastic Services for Mobile Devices* (PSMD05). Sept. 2005.
[42] Sendín, M. *Implicit Plasticity Framework: a Client-Side Generic Framework for Context-Awareness*. Proc. First International Conference on Ubiquitous Computing.CEUR-Workshop Proc.,Vol.208, 2006

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

IARV: Advanced Interfaces for Transparent Interaction with Highly Complex Models in Virtual Environments TIN2004-08065-C02

Isabel Navazo *
Departamento de Lenguajes y Sistemas Informáticos
Universitat Politècnica de Catalunya

Antoni Sellarès †
Instituto de Informática y Aplicaciones
Universitat de Girona

Abstract

The development of effective user interfaces for virtual reality environments must face different challenges. Not only must one evolve new paradigms that are adequate in those environments, but they must also be minimally invasive (not to spoil the immersion illusion), and they must rely on an adequate support from the underlying models and model-handling functions, so that the maximum complexity of usable scenes is not seriously limited. This project aims at pushing further the state of the art, developing new techniques at the modelling level (to allow handling models like the most complex currently available, which are too large to fit in core memory for a reasonably sized workstation). It also aims at perfecting the state of the art concerning the interfaces themselves, developing specific techniques adapted to the interaction in those media. Moreover, the results will be tested developing specific prototypes to verify the usability of the proposed techniques from the point of view of experts, chosen amongst those that may benefit the most from the expected results.
Keywords: Virtual Reality, Interaction, Geometric Processing, Deformable models, Image Registration, Computational Geometry, Interactive Navigation, GU Interfaces.

1 Project Goals

The current state of the art of the technology allows for the development of programs to interactively render scenes with a relatively large number of primitives —usually triangles— defined in a three-dimensional space. These successes hinge on advances in CPU's, memory

*Email: isabel@lsi.upc.edu
†Email: sellares@ima.udg.es

bandwidth, and very strongly on the advances of new GPUs of rapidly increasing performance. They are also based on the advances in fundamental research in Computer Graphics.

Existing applications span areas as diverse as designing the interiors of industrial buildings, the preservation and recuperation of cultural heritage artifacts, the aid to medical diagnosis of specific ailments, the visualization of the results of complex simulations, or the computer games.

An important aspect of all applications from the point of view of their usability, and ultimately their productivity, is the user interface. Virtual Reality systems push the interaction requirements further up because here the prime interest is in creating a feeling of immersion of the user in the synthetic scene. The user interface must then turn even thinner, even less noticeable, yet powerful enough to control a complex application. In fact the nature of the user interface may even enhance the impression of immersion, when it allows for objects to behave according to our real-world experience.

The illusion of immersion and the perception of the interface are further affected by the response time or speed of the application. If the latency or the rendering speed are too slow, the user looses the feeling of immersion, and can even have other adverse consequences (seasickness, for example). We therefore have a subtle compromise between the quality of the application and the user experience on one hand, and the complexity of the scene and of the user interface we may deploy on the other. However, we are becoming accustomed to ever more degrees of detail in the models used in synthetic scenarios, especially in applications where detail is paramount, like the preservation of cultural heritage, or the training of surgeons.

The main goal of this project is to study and develop solutions to these problems, specially adapted to virtual reality systems. We endeavor to develop new scene representations and object models designed to be scalable and suitable for interactive inspection in spite of their complexity, new paradigms of interaction that allow for the construction of non-invasive user interfaces (both for application control and for the direct manipulation of the scene and its components), and efficient algorithms to deal with deformations and topology changes because of the user interaction on the fly. We will seek to exploit the new resources available through the GPU's in these applications.

The following specific objectives are envisioned:

- *Advanced interfaces.*
  Traditional two-dimensional interfaces are not adequate for controlling virtual reality applications, which by definition are immersive, and hence the user perceives the synthetic world in three dimensions and sometimes all around him. We will pursue the development of minimally invasive interfaces for application control, and of new techniques of implicit interaction that can be also used in collaborative contexts. We will pay attention also to haptic feedback techniques in this context.

- *Techniques for modelling and manipulating objects.*
  Specific techniques for manipulating the objects in real time in a virtual reality environment and preserving acceptable frame rates are needed. In order to maintain an adequate user experience, deformations, cuts, splits and joins must be computed without falling below 20 frames per second per eye. Furthermore, we intend to continue our basic research in modelling techniques that insure that the resulting models are valid, in the context of the main applications of the project (see the last objective in this list).

- *Techniques for navigating very complex scenes.*
  For these applications to be truly scalable, the usual requirement that the model resides in main memory must be relaxed. We will study efficient algorithms for the automatic simplification of groups of objects preserving their appearance, algorithms that allow for less conservative computation of visibility for data culling, automatic computation of optimal trajectories, and rendering of out-of-core models. We envision the fine-tuning of these techniques together with the performance of modern GPU's to achieve interactive navigation of very complex scenes and scalability.

- *Validation of the results in test bed applications.*
  To test the performance and adequacy of the proposed algorithms, test bed applications will be developed in the main areas of applicability chosen: medicine, modelling of molecules and nautical engineering. Experts from those areas (be it members of the research team, like specialists in the Vall d'Hebron and Dr. Trueta Hospitals, or EPOs of the project) will test those applications and provide feedback for their evaluation and to orient further development beyond the end of this project

To achieve these goals, we enlist the expertise of research personnel of verified competence, the interest of companies to evaluate our results(Sener Ingeniería y Sistemas, Hosp. de Vall d'Hebron, Hosp. Dr. Josep Trueta), and the infrastructure of the participating groups, including the *Centro de Realidad Virtual UPC-gedas*.

The scheduling of the project contemplates tasks that extend for the full duration of the project, as they encompass doctoral dissertations in progress and the research of the teams involved. We will strive to have adequate results for the construction of the test bed software in the last year of the project.

2 Degree of success

Goals that have been attained so far are presented by briefly explaining some of the contributions produced grouped by project activities.

2.1 Advanced Interfaces

This activity aims at developing new tools and interaction techniques for application control on immersive Virtual Reality systems. The main goal is to provide VR applications with usable interfaces while minimizing the development efforts and maximizing performance, security and user comfort. We have also studied new techniques that allow having haptic perception of the objects and new low-cost virtual reality systems. The following results have been achieved:

- We have addressed the development of tools for providing Virtual Reality applications with complex GUIs (Grafical User Interfaces). We propose to extend current 2D toolkits such as Qt so that its full range of widgets can be displayed and manipulated either as 2D shapes on the desktop or as textured 3D objects within the virtual world. This approach allows 3D UI developers to take advantage of the increasing number of components, layout managers and graphical design tools provided by 2D UI toolkits [11].

- We have designed and evaluated a new interaction technique of GUIs being accessed through conventional VR devices. Two approaches have been developed, one using a raycasting technique ([12] and [1]) and new metaphor based on external GUIs displayed as texture-mapped rectangles ([13]).
- We have designed and constructed a new portable VR system for Cultural Heritage applications. The system has a 120 x 90 cm vertical screen with back-projection and passive stereo. The system can be easely disassembled (in about 10 minutes) and fits in the rear seats of a standard car for transportation. We plan to improve the system during 2007 by including a position tracking system and an optical finger tracking system [52].
- We have derived a complete component framework for transforming standalone virtual reality (VR) applications into full-fledged multithreaded collaborative virtual environments. Our main contribution is placing over the existing VR tool a very concise and extensible class framework as an add-on component that provides emerging collaboration features. [3].
- We have developed a method for rendering local haptic textures in triangle meshes, which allows a user to perceive correct surface details at several resolutions. Our method combines local heightfield maps with normal maps to model rugosity into much simpler triangle meshes [23].

2.2 Techniques for modelling objects

We have been working on the following topics:

- *Advanced modelling techniques* Within this activity we have developed techniques to extract isosurfaces and process them, which are based on the classical *Marching Cubes* algorithm, but can incorporate global criteria and different optimizations of the output surface. We highlight:
  - In [4] we present a novel approach to achieve optimal topologycal characteristics of the resulting isosurface, according to the user's desire. We can minimize or maximize, for example, the number of sheets of the surface, or its genus.
  - In a parallel effort, [44] presents a method to minimize the total surface area of the resulting isosurface.
  - When the initial data are points in three dimensional space (for example those acquired through laser scanners), [5] gives a procedure to construct a surface at a desired precision, and guaranteed to be water-tight. The algorithm does not require stringent constraints like a ε-sampling, but offers, naturally, less guarantees.
  - In [50], which is currently undergoing reviews for publication in a journal, an algorithm capable of extracting surfaces from binary grids with sub-voxel accomodation is presented.
- *Image Registration* Current medical imaging systems are able to provide accurate information of any organ of a patient. The images obtained from the devices represent different and, in some cases, complementary information of a patient. In this context,

we are investigating the application of information-theoretic tools to image registration. Some of the main achieved results are:

- We have proposed a new approach to compute the joint intensity histogram using uniformly distributed random lines to sample stochastically the overlapping volume between two 3D-images. This method provides us with an accurate, robust and fast mutual information-based registration. The interpolation effects are drastically reduced, due to the stochastic nature of the line generation, and the alignment process is also accelerated [28], [31].
- We have presented a new framework for image registration based on compression. The basic idea underlying our approach is the conjecture that two images are correctly registered when we can maximally compress one image given the information in the other.We show that the image registration process can be dealt with from the perspective of a compression problem and we demonstrate that the similarity metric performs well in image registration [30].
- We have realised a study of image simplification techniques as a first stage to define a multiresolution registration framework. We have proposed a new approach for image registration based on the partitioning of the source images in binary-space (BSP) [32].
- We have proposed a registration method for the aligment of SPECT images of the left ventricle. The algorithm is also based on the maximization of the mutual information, but we adapt and extend the standard techniques to operate on tree-dimensional images of low resolution [24].

- *Molecular Modelling*

  We have studied the approximated matching of sets of discs in the plane, we have designed algorithms for approximate computing generalized Voronoi diagrams and to determine a Center-Transversal line in the space. We think that these foundamental algorithms can be of utility to determine geometric properties of Proteins. Some of the results are:

  - Given two disks sets $\mathcal{A}$ and $\mathcal{B}$ be with $|\mathcal{A}| \leq |\mathcal{B}|$, we propose a process for determining matches between $\mathcal{A}$ and subsets of $\mathcal{B}$ under rigid motion, assuming that the position of all disks in both sets contains a certain amount of "noise". The process consists on two main stages: a candidate zone determination algorithm and a matching algorithm [38], [48].
  - In [35] we introduce a new approach to approximate generalized 3D Voronoi diagrams for different site shapes and different distance functions. The approach is based on an octree data structure denoted Voronoi-Octree (VO) that encodes the information required to generate a polyhedral approximation of the Voronoi diagram and also to solve proximity problems such as nearest neighbor queries.
  - In [34] we propose an algorithm that finds a center-transversal line for two finite point sets in $\mathbb{R}^3$in $O(n^{6+\epsilon})$, for any $\epsilon > 0$. We extend the concept of center-transversal line to that of bichromatic depth of lines in space, and give an algorithm

that computes a deepest line exactly in time $O(n^{6+\epsilon})$, and a linear-time approximation algorithm that computes, for any specified $\delta > 0$, a line whose depth is at least $1 - \delta$ times the maximum depth.

2.3 Techniques for manipulating objects

Within this activity we have been working on topological modifications of rigid solids and on deformable objects.

- *Simulation of topological modifications*
  - We propose algorithms to incrementally modify a mesh of a planar domain by interactively inserting and removing elements into or from the planar domain, keeping the quality of the mesh during the process. The changes applied to the mesh are local and the number of Steiner points added during the process remains low. Moreover, our approach can also be applied to the directly generation of refined Delaunay quality meshes [36], [37]. We also study the simulation of cuts in surgery in [46].
  - Given a tetrahedral mesh immersed in a voxel model, we present a method to refine the mesh to reduce the discrepancy between interpolated values based on either scheme at arbitrary locations. An advantage of the method presented is that it requires few subdivisions and all decisions are made locally at each tetrahedron. The method could be extended for simulating interactive cuts [25].
- *Phisical based deformations.*
  - Facial Animation. We introduce a facial deformation system that helps the character setup process and gives artists the possibility to manipulate models as if they were using a puppet. We find the correspondence of the main attributes of a generic rig, transfer them to different 3D face models and automatically generate a sophisticated facial rig based on an anatomical structure.[16],[17],[18].
  - Cloth Simulation. We present a Finite Element Method (FEM) implementation for cloth simulation on the GPU. The advantages of FEM are twofold: the realism of cloth simulations using this method is improved compared with other methods, and it has a wider application rank because it can be used for general triangulated cloth meshes. We are able to detect collisions between cloth and other objects and also we deal with self cloth collisions [15], [19].
  - Deformable Muscle Simulation. We propose a new hybrid approach for deformable models combining FEM and a Mesh Free Method, which is carried out in a virtual reality environment achieving real time performance with haptic interactions. [33][2],.

2.4 Techniques for navigating and interacting with very complex scenes

We have been working on the following topics:

- We have studying new techniques to provide interactive navigation through tools for 3D models containing thousands of objects and millions of primitives. In our first proposal [26] the scene is represented through an octree model which is able to differentiate among objects and from which we can obtain on-the-fly triangle meshes corresponding to view-dependent LODs. Given the performances of the graphic cards we are analysing the viability of an hybrid scene representation by using impostors (textured polygons) and appearance simplified triangle meshes [40].

- Another problem that has also been addressed is to automatically obtain good views for complex objects and scenes [7]. The selection of interesting views is of molecules is generally quite complex and requires a high degree of expertise for the user. We have used Information Theory to characterize two kinds of views: views which show most of the structure of a molecule and views which show a low amount of information of an arrangement of molecules. [6]

- We have focused on real time rendering of complex models from medical data. Concretely, we have faced the problem of real time cutting medical models with the aid of a haptic device. Our method interactively generates the cut of a complex triangular mesh with the help of graphics hardware by the use of an algorithm based on the Marching Tetrahedra [51].

- Another approximation to the real-time rendering of complex scenes has been addressed in the field of natural scenes. Concretely, we have focused on rendering falling leaves at real time. Our approach makes an intensive use of GPUs. The algorithm given a falling path and an initial position of a leaf, is able to calculate the positions of the leaf at different moments in time [39].

- We have designed an algorithm for multi-visibility computations in complex terrains modelled as TIN's. The algorithm exploits hardware graphics capabilities. We have also designed an algorithm for computing the widest empty slab through a set of points in the space, that can help us in determining regions with good visibility [33],[47],[9].

2.5 Software Prototypes

This activity has barely started, as it is scheduled for the last year of the project. Nonetheless, there have been some preliminary results in connection with medical applications (see [27, 29]).

3 Performance indicators

3.1 PhD and MSc students

Currently, 20 PhD students and 7 MSc students are enrolled in our PhD/MSc programme and are developing their thesis, partly or completely, within the project.

The following students have completed the PhD courses F. Argelaguet, A. Bardera, Y. Díez, M. Fort, G. Fortuny, M. Guerrieri, C. Iordache, X. Lligades, N. Madern, F. Prados, J.A. Rodríguez, V. Theoktisto, R. Trueba and H. Yela.

The following students have almost completed their PhD thesis and are expected to obtain their degrees in 2007: T. Chica (under review), V. Costa-Teixeira, L. Rodríguez, J. Rodríguez, M.G. Mero and S. Murillo.

The following PhD thesis were completed during the project:

- Coupling Marker Cell and Smoothed Particle Hydrodynamics for Fluid Animation. Nuria Suàrez de la Torre, advised by Antoni Susín, Dec 2006.
- Reflector Shape Design from Radiance Distributions. Gustavo Ariel Patow, advised by Alvar Vinacua and Xavier Pueyo, Oct 2005.

3.2 Publications

The Bibliography section lists the publications produced so far within the project. They include 10 papers in indexed journals (references [1] through [10]), 29 in international conferences with peer reviews (references [11] to [39]), one book chapter ([40]), 8 papers in national conferences with peer reviews (references [41] through [48]), one paper in a un-indexed journal ([49]) and three technical reports [50, 51, 52].

3.3 Technology transfer

As detailed in the project's work plan, most of the technology transfer efforts are concentrated in the third year, so we can provide only some preliminary results.

We have continued a fluid collaboration with the companies and institutions who showed their interest for the project from its proposal phase (*Sener*, *Gedas* and *Hospital Vall d'Hebron*). The collaboration has been particularly fruitful with Sener, which commercializes the stereoscopic system developed by our group and covered by patent No. P200102140.

Besides the above institutions, we have established a fruitful collaboration with new companies in the areas of medical visualization and interactive exploration of digital cartography.

Since 2006 our group collaborates with the *Instituto de Diagnóstico por la Imagen* to develop and apply scientific visualization techniques in the medical scenario.

From March 2005 to July 2006 our group has been involved in a project funded by *ALMA I.T. Systems* to develop a software application for medical diagnostics.

We have also signed a technology transfer agreement with *GeoVirtual, S.L.* to develop new algorithms supporting the interactive visualization of 3D digital cartography and huge databases of 3D buildings.

3.4 European projects

Our group has been involved in the project *Virtual Heritage: High Quality 3D Acquisition and Presentation (VIHAP3D)*, IST-2001-32641, funded by the EC from April 2004 to March 2005.

The ViHAP3D Consortium consisted of three research partners (Univ. Politécnica de Catalunya, Max-Planck-Institut für Informatik and the Visual Computing Group from the Istituto Elaborazione dell'Informazione) two private companies (Gedas Iberia and Minolta Europe) and one cultural heritage institution (the Soprintendenza per i beni ambientali, architettonici, artistici e Storici per le province Pisa, Livorno, Lucca e Massa Carrara). Our basic

contribution to the joint results included a mesh repair tool, a virtual builder and a virtual browser.

3.5 Collaboration with other research groups

We have cooperated, in the framework of the project, with a number of foreign groups and researchers. Among them, we would like to cite the following ones:

- GVU Group, Georgia Tech, US: Professor Jarek Rossignac
- CNUCE, CNR, Pisa, Italy: Dr. Roberto Scopigno, Dr. Claudio Montani
- Trinity College, Dublin, Ireland: Prof. Carol O'Sullivan
- INRIA, Grenoble, France: Prof. Marie-Paule Cani (who visited our group in Dec 2006)
- Fraunhofer Institute, Darmstadt, Alemania: Prof. Dr. Dieter Fellner
- Max Plank Institute for Computer Graphics, Saarbruecken, Germany: Prof. Dr. Hans-Peter Seidel

Mel Slater holds an ICREA Research Professorship in our group since January 2006. Mel Slater founded the Virtual Environments and Computer Graphics group in the University College London.

The project team has been active in a number of scientific journals and conferences. The International Conferences VRIPHYS'2006 and SIACG'2006, together with GRAPP 2006, were Co-Chaired by members of the project team.

Pere Brunet was a member (between April and July 2006) of the international INRIA evaluation committee. This included the participation in the presentation of the research activitites of the INRIA main teams in France (May 2006) and the preparation of the final report.

References

Papers in Indexed Journals

[1] C. Andujar and F. Argelaguet. Anisomorphic ray-casting manipulation for interacting with 2D GUIs. *Computers & Graphics*, 31(1):(in press), 2007.

[2] Rodríguez-Baeza A., Armengol-Carrasco M., Susín A., López-Cano M., and Rodríguez-Navarro J. Areal-time dynamic 3d model of the human inguinal region for surgical education. *Computers in Biology and Medicine*, (in press), 2007.

[3] V. Theoktisto and M. Fairén González. Enhancing collaboration in virtual reality applications. *Computers & Graphics*, 29(5):704–718, 2005.

[4] C. Andujar, P. Brunet, A. Chica, I. Navazo, J. Rossignac, and A. Vinacua. Optimizing the topological and combinational complexity of isosurfaces. *Computer-Aided Design*, 37(8):847–857, 2005.

[5] J. Esteve, P. Brunet, and A. Vinacua. Approximation of a variable density cloud points by shrinking a discrete membrane. *Computer Graphics Forum*, 24(4):791–808, 2005.

[6] P.-P. Vázquez, M. Feixas, M. Sbert, and A. Llobet. Realtime automatic selection of good molecular views. *Computers & Graphics*, 30(1):98–110, February 2006.

[7] P.-P. Vázquez. Automatic light source placement for maximum illumination information recovery. *Computer Graphics Forum, accepted*, page (in press), 2006.

[8] J. M. Díaz-Báñez, M. A. López, and J. A. Sellarès. On finding a widest empty 1-corner corridor. *Inf. Process. Lett.*, 98(5):199–205, 2006.

[9] J. M. Díaz-Báñez, M. A. López, and J. A. Sellarès. Locating an obnoxious plane. *European Journal of Operational Research*, 173(2):556–564, 2006.

[10] G. Patow, X. Pueyo, and À. Vinacua. User-guided inverse reflector design. *Computers and Graphics, accepted*. (in press).

Papers in International Conferences

[11] C. Andujar, M. Fairen, and F. Argelaguet. A cost-effective approach for developing application-control guis for virtual environments. In *Proc. of 1st IEEE Symposium on 3D User Interfaces, 3DUI'06*, pages 45–52, 2006.

[12] C. Andujar and F. Argelaguet. Friction surfaces: scaled ray-casting manipulation for interacting with 2d guis. In *Proc. of 12th Eurographics Symposium on Virtual Environments, EGVE'06*, pages 101–108, 2006.

[13] C. Andujar and F. Argelaguet. Virtual pads: Decoupling motor space and visual space for flexible manipulation of 2d windows within ves. In *Proc. of 2nd IEEE Symposium on 3D User Interfaces, 3DUI'07*, 2007.

[14] P. Brunet, I. Navazo, and A. Vinacua. Modelling very complex geometric assemblies: The use of discrete and hierarchical models. *Computer-Aided Design and Applications*, 3(5):639–644, 2006.

[15] A. Susín, Rodríguez-Navarro J., and M. Sainz. Fast cloth simulation on moving humanoids. In J. Dingliana and F. Ganovelli, editors, *Eurographics Short Presentations (EG'05)*, pages 85–88, 2005.

[16] Susín A., Costa-Teixeira V., and Zacur E. A maya plug-in for transferring a labeled generic rig to animate face models. In Ramires Fernades A. Dionisio J., editor, *International Digital Games Conference (iDiG'06)*, pages 12–20, 2006.

[17] Susín A., Costa-Teixeira V., and Zacur E. Transferring facial expressions to different face models. In Baranoski G Brunet P., Correia N., editor, *3th Ibero-American Symposium in Computer Graphics (SIACG-2006)*, pages 21–28, 2006.

[18] Susín A. Costa-Teixeira V., Zacur E. Transferring a labeled rig to animate face models. In Springer Lecture Notes Perales F., Fisher B., editor, *IV Conference on Articulated Motion and Deformable Objects (AMDO-06)*, pages 21–31, 2006.

[19] Susín A. Rodríguez-Navarro J. Non structured meshes for cloth gpu simulation using fem. In Navazo I. Mendoza C., editor, *Workshop On Virtual Reality Interaction and Physical Simulation (VRIPHYS'06)*, pages 1–7, 2006.

[20] Susín A. Mero M.G. Deformable hybrid approach for haptic interaction. In Navazo I. Mendoza C., editor, *Workshop On Virtual Reality Interaction and Physical Simulation (VRIPHYS'06)*, pages 8–16, 2006.

[21] Susín A. Suárez N. A mesh-particle model for fluid animation. In Baranoski G Brunet P., Correia N., editor, *3th Ibero-American Symposium in Computer Graphics (SIACG-2006)*, pages 13–20, 2006.

[22] Susín A. Suárez N. A mesh-particle model for the animation of large-volume fluids. In *International Conference on Engineering and Mathematics 2006 (ENMA06)*, pages 169–176, 2006.

[23] V. Theoktisto, M. Fairén, , E. Monclús, and I. Navazo. Rendering detailed haptic textures. In *Proc. of VRIPHYS'05, Second Workshop on Virtual Reality Interaction and Physical Simulation*, Pisa, Italy, November 2005.

[24] S. Murillo, I. Navazo, and A. Vinacua. Volume cardiac spect image registration. In Ch. Moore G. Clapworthy, editor, *Proc. of International Conference on medical Information Visualization, MediVis'06)*, London, UK.

[25] L. Rodriguez, I. Navazo, and A. Vinacua. Data-driven tetrahedral mesh subdivision. In G. Baranoski P. Brunet, N. Correiara, editor, *Proc. of SIACG'06, Ibero-American Symposium in Computer Graphics)*, Santiago de Compostela, Spain.

[26] M. Fairen and R.Trueba. Octree-based view-dependent triangle meshes. In J. Rossignac and V. Skala, editors, *Proc. of WSCG'07, 15th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, page in press, 2007.

[27] J.A. Hueto, I. Navazo, A. Casals, J. Gonzalez-Lagunas, and G. Raspall. Prototype of robotic assistant for craniofacial osteotomies. In *Proc. of 17th International Conference on Oral and Maxilofacial Surgery*, Vienna, Austria.

[28] A. Bardera, M. Feixas, I. Boada, and M. Sbert. Medical image registration based on random line sampling. In *IEEE International Conference on Image Processing (ICIP'05), Proceedings*, Genova, Italy, September 2005.

[29] A. Bardera, I. Boada, M. Feixas, Salvador Pedraza, and Joaquim Rodriguez. A framework to assist acute stroke diagnosis. In *Vision, Modeling and Visualization, Proceedings*, pages 359–366, Erlangen, Germany, November 2005.

[30] A. Bardera, M. Feixas, I. Boada, and M. Sbert. Compression-based image registration. In *IEEE International Symposium on Information Theory*, Seattle, 2006.

[31] A. Bardera, M. Feixas, I. Boada, and Mateu Sbert. High dimensional normalized mutual information for image registration using random lines. In *International Workshop on Medical Image Registration*, Utrecht, Netherlands, July 2006.

[32] A. Bardera, M. Feixas, I. Boada, Jaume Rigau, and Mateu Sbert. Medical image registration based on BSP and quadtree partitioning. In *International Workshop on Medical Image Registration*, Utrecht, Netherlands, July 2006.

[33] N. Coll, M. Fort, N. Madern, and J. A. Sellarès. Computing terrain multi-visibility maps for a set of view segments using graphics hardware. In *ICCSA (1)*, pages 81–90, 2006.

[34] P. K. Agarwal, S. Cabello, J. A. Sellarès, and M. Sharir. Computing a center-transversal line. In *FSTTCS*, pages 93–104, 2006.

[35] N. Coll, I. Boada, N. Madern, and J. A. Sellarès. Approximations of 3d generalized voronoi diagrams. In *21th European Workshop on Computional Geometry*, pages 163–166, 2005.

[36] N. Coll, M. Guerrieri, and J. A. Sellarès. Modifying delaunay refined two-dimensional triangular meshes. In *22th European Workshop on Computional Geometry*, pages 129–132, 2006.

[37] N. Coll, M. Guerrieri, and J. A. Sellarès. Mesh modification under local domain changes. In *15th International Meshing Roundtable*, pages 39–56, 2006.

[38] Y. Diez and J. A. Sellarès. Noisy disk set matching under rigid motion. In *22nd European Workshop on Computational Geometry*, pages 115–118, 2006.

[39] P. Vázquez and M. Balsa. Real time falling leaves. In *Proc. of GRAPP 2007*, page (in press), 2007.

Book Chapters

[40] C. Andujar, M. Fairen, P. Brunet, and V. Cebollada. *Simplification of Topologically Complex Assemblies*, pages 339–352. Springer-Verlag, 2005.

Papers in National Conferences

[41] A. Susín, Rodríguez-Navarro J., M. Sainz. Gpu based cloth simulation with moving humanoids. In D. Martín J. Regincós, editor, *Proc. XV Congreso Español de Informática Gráfica- (CEIG2005)*, pages 147–155, 2005.

[42] A. Susín and Mero M.G. Deformable hibrid model for haptic interaction. In D. Martín J. Regincós, editor, *Proc. XV Congreso Español de Informática Gráfica- (CEIG2005)*, pages 237–240, 2005.

[43] A. Susín and Suarez N. Desarrollo de un método mixto malla- partícula para la animación de fluidos. In D. Martín J. Regincós, editor, *Proc. XV Congreso Español de Informática Gráfica- (CEIG2005)*, pages 233–236, 2005.

[44] C. Iordache and P. Brunet. Minimizing isosurface area. In D. Martín J. Regincós, editor, *Proc. XV Congreso Español de Informática Gráfica- (CEIG2005)*, Granada, Spain.

[45] J.A. Rodríguez, P. Brunet, N. Ezquerra and J.E. Palomar. A Virtual Reality Approach to Progressive Lenses Simulation. In D. Martín J. Regincós, editor, *Proc. XV Congreso Español de Informática Gráfica- (CEIG2005)*, pages 43–52, 2005.

[46] N. Coll, M. Guerrieri, and J. A. Sellarès. Cutting two-dimensional triangular meshes. In *XI Encuentros de Geometría Computacional*, pages 39–46, 2005.

[47] N. Coll, M. Fort, and J. A. Sellarès. Muli-visibility in terrains. In *XI Encuentros de Geometría Computacional*, pages 71–78, 2005.

[48] Y. Diez and J. A. Sellarès. Approximate disk matching under rigid motion. In *XI Encuentros de Geometría Computacional*, pages 279–286, 2005.

Papers in un-indexed journals

[49] P.-P. Vázquez and Dani Susín. Generación de penumbras con hardware gráfico. *Novática*, pages 49–53, Jul-Aug 2005.

Technical Reports

[50] C. Andujar, P. Brunet, A. Chica, I. Navazo, J. Rossignac, and A. Vinacua. Smooth isosurfaces with flats form binary grids. *LSI Research Report (LSI-06-9-R)*, 2006.

[51] E. Monclús, I. Navazo, and P. Vázquez. Mtcut: Gpu-based marching tetra cuts. *Technical Report, Departament de LSI, UPC, LSI-07-10-R, submitted*, 2006.

[52] J. Boo. Sistema portable de realidad virtual. *Technical Report, Departament de LSI, available under www.lsi.upc.edu/ pere/PapersWeb/Portable.pdf*, 2006.

Jornadas de Seguimiento de Proyectos, 2007
Programa Nacional de Tecnologías Informáticas

E-learning adaptive services: methodologies, architectures and languages (MetaLearn) TIN2004-08367-C02

Martín Llamas Nistal[*]
Universidade de Vigo

Baltasar Fernández Manjón[**]
Universidad Complutense de Madrid

Abstract

New information and communication technologies improve education providing several new features, like time independence or content adaptation. The success of this approach to learning makes web-based education or e-learning a generally accepted concept. E-learning proliferates, but course development costs are generally high. As a consequence, the trend is to develop interoperable e-learning solutions, that is, e-learning platforms able to exchange information and work cooperatively. Besides, a higher degree of content reuse is also desirable to decrease content development costs. Standards are being developed in practically all fields of e-learning to support these features.
However, in spite of technical advances, adaptability has not been conveniently addressed in available commercial systems. However, adaptability at all levels – pedagogical, knowledge delivery, technical – is becoming an important feature, due to the heterogeneity of users and learning sceneries.
This project is aimed to contribute to the adaptability of state-of-the-art e-learning solutions, providing the corresponding support for e-learning application development through methodologies, architectures and languages. Obviously, this support will be independent of the underlying infrastructure and network environment. Adapted services and contents will be provided transparently insofar the user's network environment is concerned. Our proposals will likely have an impact in present-day and future e-learning standards. As a consequence, standardization-related aspects of adaptability in e-learning will also be addressed.

Keywords: e-learning, adaptability, architectures, methodologies

1 Project objectives

Main objectives are the definition and development of methodologies, architectures and languages to support adaptive e-learning. Our efforts are focused on the current standardization framework, with proposals for updates or new proposals.

[*] Email: martin@uvigo.es
[**] Email : balta@fdi.ucm.es

This general objectives are organized around two sub-projects:

The goals of the UCM (Universidad Complutense de Madrid) subproject were:

- Definition and development of a methodology for adaptive e-learning services provision
- Definition and development of support model for adaptive e-learning services provision
- Definition of a learning model for adaptive e-learning
- Standards and Adaptive e-learning
- Domain Specific Languages (DSL) Definition used for modelling processes in e-learning

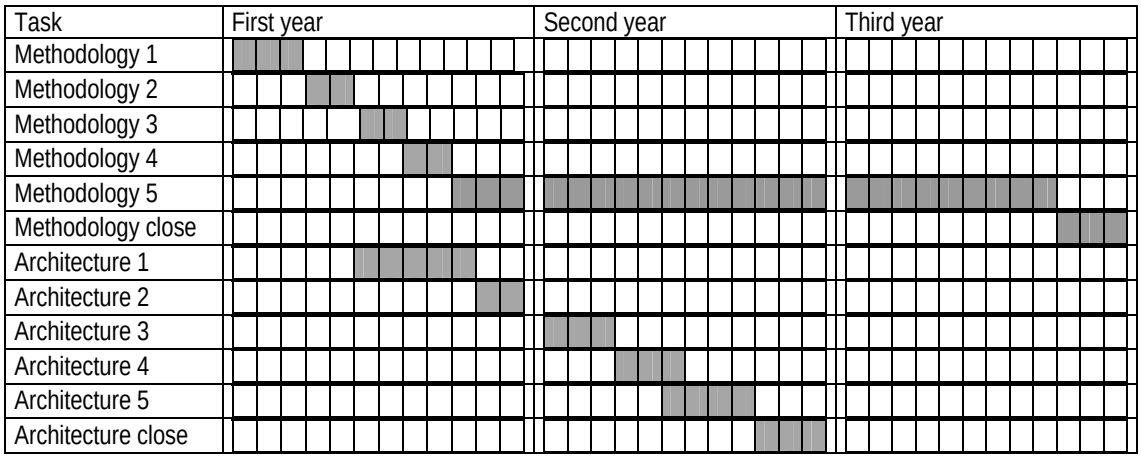
The goals of the UV (Universidad de Vigo) subproject:

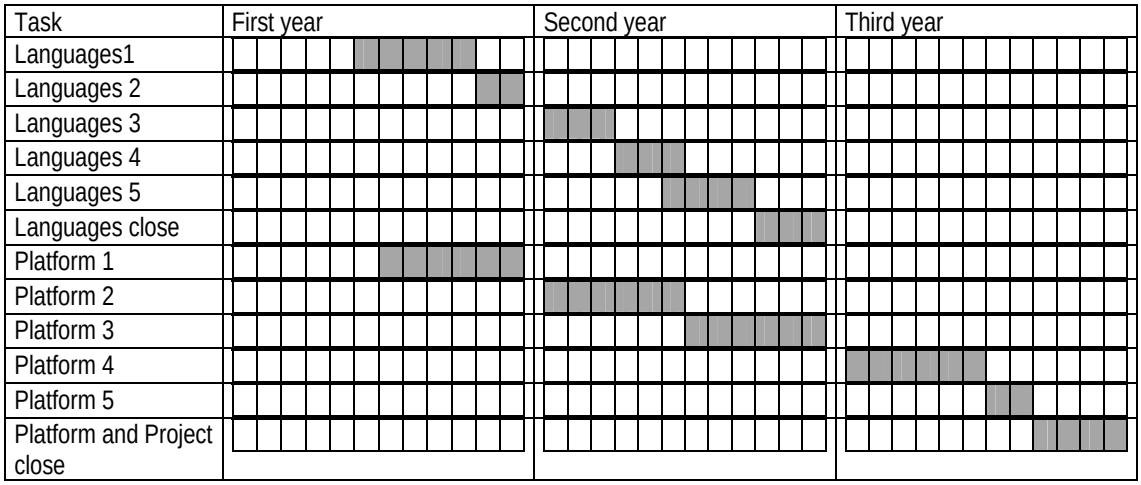
- Definition of a ontology for e-learning with support for adaptability.
- Definition of a adaptive multi-agent system for e-learning
- Definition of a framework to provide adaptive content and services in e-learning.
- Adaptation of current standards to support the previous objectives
- Development of the needed support tools to provide the previous objectives.
- Development, deployment and testing of a system to provide adaptive e-learning

To cope with the above defined objectives, the project was divided into four main modules. Each module was developed using an incremental methodology: (1) state-of-the-art, (2) analysis, (3) design, (4) development and (5) testing.

- MODULE 1: Methodology to provide adaptive e-learning services.
- MODULE 2: Definition and development of a distributed architecture for adaptive agents for e-elearning.
- MODULE 3: Languages for the modelling domain and automatization of processes.
- MODULE 4: Development and implementation of a software platform according to tasks 1, 2 and 3.

Task temporal organization is shown below:





2 Success level achieved in the project

The UV was responsible for the definition of the ontology and architectural infrastructure as stated in the objectives. A high degree of integration was achieved with the works carried out at the UCM thanks to the identification of new working lines within the UV duties, namely the definition of adaptive Educational Modelling Languages.

Within these generic objective, we have Developer an architecture [24, 25, 26, 27, 28, 29] based in adaptive Multiagent Systems for using the Cased-Based Reasoning paradigm in order to simplify the development of new Intelligent Tutoring Systems (ITSs). This allows us to automate the CBR phases and to reduce the difficulty for evaluate the proposed solution and, in case it is needed, to repair it using domain specific knowledge. In this domain, we have proceeded to enhance the adaptability of Intelligent Tutoring Systems (ITSs) by developing an adequate student model. This allows to detect with enough precision the knowledge level of the student and determine the best pedagogical strategy to achieve adaptive ITSs. This work is supported by the publications: [21, 22, 23]. Within the last year of the project we will integrate the so-defined ITS with the ontology defined to support adaptive learning and adaptive brokerage. Ontologies and related semantic-based mechanism are essential to support the success of this project.

In fact, one of the main objectives of the METALEARN project was the definition of underlying mechanisms to solve the adaptive problems in the e-learning domain with the aim of getting more context-aware personalised services. These mechanisms must be aligned with the current learning Technologies standards and specifications in order to maintain the interoperability properties of the systems to develop. So, several of these standards, like the metadata models for describing learning resources, formats for defining competencies, schemas for represent learner information, data models for describing accessibility issues, are the basis for the definition of ELEARNING-ONT [10, 14, 19], a set of interconnected OWL ontologies that facilitate the automatic management of the implicit semantic present in the instances of the standardized data models.

ELEARNING-ONT includes the definition of the concepts, and their inter-relations, necessary to develop adaptive and personalised services in the e-learning domain. Due to the great quantity of identified terms, the ontology is organized in a range of namespaces (or sub-

ontologies). There exists a basic namespace, where fundamental concepts such as "Educational Resource", "Course" or "Educational Services Provider" are defined. A series of sub-ontologies include the properties, with their corresponding vocabularies, that can be used to describe in detail the instances of the most basic classes. ELEARNING-ONT is the basis of the PhD by Prof. Juan Santos, which will be delivered for presentation before the end of the project.

Besides the above presented results and in accordance with the global objectives of the project, additional work in the area of Educational Modelling Languages has been carried out The development of a new EML, named as PoEML (Perspective-oriented Educational Modelling Language), based on a separation of concerns principle intended to solve the identified problems. This task has captured more of the efforts carried out during 2006. The main issues involved in these activities have been: (1) The development of the conceptual meta-model of the language. We have been working in the development of a textual and a graphical notation. The textual notation has been developed in a XML schema. The graphical notation enables to visualize the models in several views accordingly with the separation of concerns; and (2) The development of an authoring tool to facilitate the creation of models of educational units in PoEML. This tool is being finished and it is possible to obtain information at http://www.poeml.com. These activities are directly related with the goal number 8 about the "definition of domain specific languages for process modelling". This goal had been considered in the module number (I) "Domain specific language for process modelling and automatization". POEML is the main support of the PhD by Prof. Manuel Caeiro, which will be delivered for presentation before the end of the project.

New working areas were identified along the development of this project. In particular, those underlying concepts being applied to obtain the original objectives can be applied to other domains. An initial research effort has been carried out in the area of e-government. In due course of the researching carried out in the frame of this project, this group has identified notable points in common in both domains. These similarities (problems of interoperability, citizen support, difficulties in semantic characterization, needs of adaptability …) had lead the team towards synergetic solutions in the development of solutions, at least, in a partial manner.

Outcomes of this parallel researching line are reflected in contributions to several international fora, see[2][30][31][32][33][34][35]. These publications reflect some advances in the design of extensible and feasible support in eGovernment transactions. As data level is clearly not anymore longer the battlefield for interoperability, the use of semantic plays a paramount role as a common ground to interchange services and information.

Interoperability, standardization and integration of heteregenous systems is the core of this project and the common ground for the work in both participating groups. Since 1999 the UCM group has worked on e-learning systems based on international standards, as well as in the definition of methodologies and domain-specific languages applied to the educational domain. Furthermore, from the previous <e-aula> project, this group has been working on different aspects of the conceptualization and prototyping of hypermedia applications (Ph.D. of Antonio Navarro, presented in 2002) and the <e-aula> e-learning platform [37] has been used as a base for further development [50, 52, 54]. Besides, in 2004 the Ph.D. dissertation of José Luis Sierra was presented. In this thesis, a document-oriented paradigm for the development of applications was proposed. This paradigm, which is based on Domain Specific Languages (DSL), is being used in the MetaLearn project for the formulation of methodologies and supporting tools for educational standards [56, 60, 61, 62].

In MetaLearn we are dealing with the adaptive e-learning infrastructure based on the application of e-learning standards but taking into account the whole process [53]. This includes the definition of methodologies to design educational content and systems, the application of

domain specific languages to automate the process, and the development of authoring tools for highly adaptable and interactive learning content [40, 64, 66].

As the main results of MetaLearn and using the document-oriented paradigm we have established an initial methodology for the language-driven development of software in the domain of adaptive e-learning. This methodology facilitates the adaptation, extension and combination of the descriptive markup languages included in the specifications proposed in the e-learning domain with the objective of adapting them to satisfy the special needs of adaptive learning scenarios [30]. It also facilitates the modular implementation of appropriate processors for the resulting languages. Also, we have developed a methodology for the production of standards-compliant educational contents with the aim of simplifying the workload of the instructors. This platform also supports some degree of adaptability. We have also developed a tool for the creation of assessments and exams with the main types of questions following the standard QTI v 1.2, which also provides with a correction engine [45] (currently being upgraded to QTI 2.0). We have also started to work in the application of the document-oriented paradigm to the systematic production of highly-interactive contents (e.g. conversational videogames) that can be also integrated in e-learning systems. These applications are described in greater detail later.

A relevant result of the second year is derived of the cross-fertilization with the UCM Project OdaVirtual where some of the members of the research team are also involved as researchers. In this project we have further applied and refined the methodology based on the document-oriented paradigm in the development of a repository of learning objects associated with two academic museums at UCM [51, 58, 59, 63]. The resulting system, which is called Chasqui, is currently being applied to knowledge domains of more abstract nature (language processors and computational linguistics) [55, 57]. In these experiences we are taking special care in enriching the system with several adaptation-based features.

The creation of high quality adaptive educational materials is a complex and expensive task. This effort could be useless if these materials can not be exchanged between different systems. In addition, there has been a lot of research about the concept of Learning Object, to decrease the effort of authoring high quality educational materials. The aim of the content creation methodology has been to ease the teacher task and to avoid vendor-locking problems with the developed materials, hence the teacher can migrate easily between platforms, but also he/she can improve the quality and reusability of the content applying the Learning Object paradigm during the creation process [44, 49].

Another initiative of the MetaLearn project has been the development of <e-QTI> [45], an assessment web-based tool compliant with the IMS Question and Test Interoperability specification (QTI version 1.2). Now the application supports a rich set of questions such as True/False, Multiple Choice with multiple elections, filling the blanks, essay, etc. The aim of this project has been the development of four tools: A player, an authoring tool, a question pool, and an Importation/Exportation tool. The application has been developed bearing in mind that the tool will not be used isolated, so the application now can be integrated with the <e-aula> LMS. In the future our idea is to integrate <e-QTI> with other LMS such as SAKAI, .LRN and Moodle. Finally, it is necessary to say that currently we are upgrading the application to support the new versions 2.0 and 2.1 of IMS QTI specification.

Another of the initiatives in the MetaLearn project has been the development of the <e-LD> subproject. <e-LD> tries to solve some of the issues identified when applying the Educational Modeling Languages paradigm [42], in particular, IMS Learning Design. This paradigm is based on the concept of Unit of Learning where the teacher describes all the activities during the learning process. First we try to ease the teacher task during the creation of Units of Learning

providing a graphical tool to author the instructional designs (this tool uses a UML metaphor to abstract the LD constructs). Second, we provide an engine based on workflows technologies to run those UoL [47, 48].

As previously mentioned, highly interactive and adaptive content was one of the goals of MetaLearn, and this was addressed in the <e-Game> subproject. The main objectives were to explore the use of computer games as a medium to deliver content with a highly adaptable format and to explore the potential synergies between Educational Modelling Languages and game programming languages. After the initial research, a third research question was raised: How to lower the excessively high development cost associated with educational games (when compared to alternative media). Thus, the result is the specification of a Domain-Specific Markup Language (the <e-Game> language) which can be used to define adaptive educational games (this is one of the application of DSL documental paradigm previously mentioned). These games can be played using the <e-Game> engine, which can be delivered to the learner through a Learning Management System.

The language is defined as an XML dialect to facilitate its comparison and interaction with other Educational Modelling Languages defined with the same technology. The language and its main features are described in [39, 41]. Additionally, the use of descriptive markup in the definition of the educational games is the starting point for a development process that can significantly reduce the development cost for these games by providing a methodical way of translating a game storyboard into a fully functional game. The details of this development process are described in [38, 46].

On the other hand, both the language and the engine were developed with adaptive education in mind. This means that the language includes a number of features related to adaptation of the games and assessment of the activity of the learner within the game. These features are described in [41, 43].

Regarding the synergies with Educational Modelling Languages, the <e-Game> language and engine have been integrated in an IMS Learning Design environment (the CopperCore engine) in close collaboration with with Dr. Daniel Burgos and Prof. Rob Koper from the Open University of the Netherlands. The results of that development have been described in two journal articles recently sent to high-impact journals for their peer revision.

Another important working line is the recent collaboration agreement with the National Centre of Educational Information and Communication (CNICE) as experts in standard technologies, with full access to their educational contents. We are in charge of developing an official report sponsored by the CNICE Ministry of Education "The use of standards in Information and Communication Technologies for ·Education" [36].

In the international side a relevant outcome of the MetaLearn Project is the collaboration with the Learning Technology Task Force (LTTF) of IEEE. Baltasar Fernández-Manjon has been positioned as the liason member among LTTF and WG 3.3 Research on Education Applications of Information Technologies of the IFIP (http://lttf.ieee.org/liaison.htm). This is an optimal situation to contact and collaborate with members of the two most relevant research organisms in the Educational Technologies arena.

Besides, currently at the UCM there are 5 Ph.D. theses in progress related to adaptive e-learning (3 of them will be presented before the end of MetaLearn project) and these results will be the main in the new project proposal.

All the PhD researchers at the UCM group are involved in the "E-learning" course of the official Master "Master de Investigación en Informática" included in the Plan Oficial de Posgrado "Ingeniería Informática" that has the "mencion de calidad" (referencia MCD2006-00500, Mención

de Calidad que se renueva MCD2004-00395 por resolución de 11 de agosto de 2006 de la Secretaría de Estado de Universidades e Investigación, B.O.E. 30/08/06).

3 Outcomes indicators

Coordination was carried out through two meetings, co-located with CTN71/SC36 de AENOR in Madrid, with SIIE2005 conference in Leiria (Portugal), with SIIE 2006 in León (Spain). One of the main activities was the collaboration of the group with the official standardization boby for e-learning in Spain: AENOR CTN71/SC36 (Subcomité CTN71/SC36 "Tecnologías de la información para el aprendizaje" whose main objectives are "Standardization of applications, products, services and specifications related with educational, training or learning technologies in the scope of individuals, organizations or groups, in order to enable interoperability and reuse of tools and resources".

Collaboration with other groups:

- REDAOPA – Thematic Network Objects and Designs for Learning (TSI2004-21263-E). Coordinator: Miguel Angel Sicilia, Salvador Sánchez , Univ. de Alcalá de Henares. Participants: 15 Spanish Universities, 3 foreign universities.

- Thematic Network of the Spanish Chapter of IEEE Education Society (TSI2005-24068-E). Coordinator: Martín Llamas Nistal (Universidad de Vigo). Participants: 74 researchers, 45 departments of 31 Spanish universities.

Manuel Caeiro has stayed in the University of Coimbra, Department de Informatics Engineering, during 3 months (01-03-06/31-05-06). During this stay, he has been working with Dr. Maria Jose Marcelino and Dr. Antonio Mendes in the development of the EMLs evaluation framework taking into account some of their lectures, courses and research experience.

Some of the features of <e-game> were developed in conjunction with the Laboratory of Computer Science at Harvard Medical School during a 3-month stay of Pablo Moreno Ger (May 1st 2006 through July 30th 2006) and further polished during Dr. Carl Blesius' stay at UCM in October 2006. Specific aspects of the integration with LD were developed in a 6-week stay of Pablo Moreno Ger at the Open University of the Netherlands working with Dr. Daniel Burgos and Prof. Rob Koper (November 1st thorugh December 15th).

Iván Martínez Ortiz has stayed 9 weeks at the Electrical and Computer Engineering department of the University of New Mexico with Prof. Gregory Heileman. His stay was related to adaptability (publishing and repurposing), security and rights management of Learning Objects.

Outcomes from this project lead to the following publications (10 JCR-journal papers, 5 other-journal papers, 10 LNCS/LNAI papers, 33 international congress papers and book-chapters, 7 national congress papers and 1 official report):

4 References

All the references included below have been the result of different outcomes from this project and may be considered as success indicators.

[1] Caeiro, M., Llamas, Anido, L., "Towards a Benchmark for the Evaluation of LD Expressiveness and Suitability, Journal of Interactive Media Education, (4), September 2005. ISSN: 1365-893X

[2] Luis Álvarez-Sabucedo and Luis Anido-Rifón. "A Proposal for a Semantic-Driven eGovernment Service Architecture". EGOV2005, Lecture Notes on Computer Science Vol. 3591, pp.: 237-248. ISSN: 0302-9743. Agosto 2005

[3] Manuel Caeiro-Rodríguez, Martín Llamas-Nistal and Luis Anido-Rifón. "Modelling Group-Based Education. A proposal for a Meta-model". Lecture Notes on Computer Science Vol. 3716, ISSN: 0302-9743. pp.: 96-111. Octubre 2005

[4] Carolina González , Juan C. Burguillo, Martin Llamas. "Case-based student modelling in multi-agent learning environments". CEEMAS 2005. Lecture Notes in Computer Science Vol. 3690. pp.: 72-81. ISSN: 0302-9743 September, 2005.

[5] Manuel Caeiro, Martín Llamas and Luis Anido. "From Contents to Activities: Modelling Units of Learning". Journal of Universal Computer Science, vol. 11, no. 9 (2005), pp.:1458-1469. ISSN 0948-69x.

[6] M. Caeiro-Rodríguez, M. Llamas-Nistal, L. Anido-Rifón. "An EML meta-model proposal for the modelling of Collaborative Educational Practices" VII Simposio Internacional de Informática Educativa (SIIE'05). Leiría, Portugal. ISBN 972-95205-4-7 Noviembre 2005.

[7] M. Caeiro-Rodríguez, L. Anido-Rifón, M. Llamas-Nistal. "Contributions to Educational Modelling Languages". II Simposio Pluridisciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos Reutilizables (SPDECE). ISBN 84-9788-350-0 Barcelona, España. Octubre 2005.

[8] Caeiro Rodríguez M.; Llamas-Nistal M.; Anido-Rifón L. "A Proposal for a Web Service-based Architecture to Support the Enactment Of Units of Learning" Enterprise Modelling and Information Systems Architectures. Lecture Notes in Informatics (LNI). Pp.180-193 ISBN 3-88579-404-7 ISSN 1617-54687 Austria, Octubre 2005.

[9] Manuel Caeiro, Martín Llamas and Luis Anido. "Standards Support for Different Learning Models in Web-Based Education" ICECE05, International Conference on Engineering and Computer Education. Madrid, Spain. 2005. ISBN 84-609-8149-5

[10] Juan M. Santos, Luis Anido, Martín Llamas. "Design of a Semantic Web-based Brokerage Architecture for the e-learning Domain. A proposal for a Suitable Ontology". 35th ASEE/IEEE Frontiers in Education Conference. October 19 – 22, 2005, Indianapolis, 2005. ISBN 0-7803-9078-4

[11] Manuel Caeiro-Rodriguez, Martin Llamas-Nistal, Luis Anido-Rifon, "Meta-modeling for Computer-supported Group-based Learning Design". 5th International Conference on Intelligent Systems Design and Applications (ISDA'05), pp. 84-89. ISBN 0-7695-2286-6. Polonia, Septiembre 2005.

[12] Manuel Caeiro-Rodriguez, Martin Llamas-Nistal, Luis Anido-Rifon, "Towards a Benchmark for the Evaluation of LD Expressiveness and Suitability". Current Research on IMS Learning Design, pp. 183-192, ISBN 9035823052, Holanda, Septiembre 2005.

[13] Manuel Caeiro-Rodriguez, Martin Llamas-Nistal, Luis Anido-Rifon. "Computer Support in Group-Based Learning. A Meta-Model Contribution to Educational Modelling Languages". V Jornadas de Ingeniería Telemática, pp.:57-64. Vigo, España. Septiembre 2005. ISBN 84-8408-346-2

[14] Juan Manuel Santos Gago, Luis Anido Rifón, Martín Llamas Nistal. "Elearning-Ont: Metodología y Subontologías". Simposio Nacional de Tecnologías de la Información y las Comunicaciones en la Educación, pp.: 87-93. Granada, 2005. ISBN 84-9732-437-4

[15] Manuel Caeiro Rodríguez, Martín Llamas Nistal, Luis Anido Rifón. "Instructional Design in Educational Modelling Languages". Simposio Nacional de Tecnologías de la Información y las Comunicaciones en la Educación, pp.: 111-118. Granada, 2005. ISBN 84-9732-437-4
[16] L. Anido, L. Sabucedo, M. Caeiro, J. Santos, M. Llamas, M. J. Fernández, F. Mikic. "An Observatory for E-learning Technology Standards". Proceedings of the Eighth IASTED International Conference on Computers and Advanced Technology in Education. ISBN: 0-88986-520-5, Aruba, Agosto 2005.
[17] M. Caeiro-Rodríguez, L. Anido-Rifón, M. Llamas-Nistal. "Capturing Requirements for the Modelling of Collaborative Learning Experiences". Proceedings of the Eighth IASTED International Conference on Computers and Advanced Technology in Education. ISBN: 0-88986-520-5, Aruba, Agosto 2005.
[18] M. Caeiro-Rodríguez, L. Anido-Rifón, M. Llamas-Nistal. "Improving the Modelling of Heterogeneous Learning Activities". Proceedings of the Eighth IFIP World Conference on Computers in Education. ISBN: 1-920-01711-9, Cape Town, SudÁfrica, Julio 2005.
[19] Santos, J.M., Anido, L., Llamas, M.. "Semantic Brokerage in Standards- driven e-Learning". Proceedings of the Eighth IFIP World Conference on Computers in Education. ISBN: 1-920-01711-9, Cape Town, SudÁfrica, Julio 2005.
[20] Manuel Caeiro-Rodriguez, Luis Anido-Rifon, Martin Llamas-Nistal, "A Perspective and Pattern-Based Evaluation Framework of EMLs' Expressiveness for Collaborative Learning: Application to IMS LD". Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05), pp. 51-53, ISBN 0-7695-2338-2. Taiwan, Julio 2005.
[21] Carolina González, J.C. Burguillo-Rial, Juan C. Vidal, Martín Llamas. SINCO: Intelligent System in Disease Prevention and Control. An Architectural Approach. Biological and Medical Data Analysis. Lecture Notes in Computer Science (LNCS) 3337. Springer-Verlag . Barcelona (España). 2004
[22] C. González, J.C. Burguillo, M. Llamas, J.C. Vidal. Sistemas tutores inteligentes: propuesta de una arquitectura para aprendizaje en salud pública. Congreso: RIBIE. Monterrey (México). Octubre 2004
[23] Carolina González, Juan C. Burguillo, Martín Llamas. Case-based Student Modelling in a Multi-agent Learning Environment. Multi-Agent Systems and Applications IV. Lecture Notes in Artificial Intelligence, 3690. Springer-Verlag. Budapest (Hungría). 2005.
[24] González, C., Burguillo, J.C., Vidal, J.C., Llamas, M., López D.M. Web Based Intelligent Medical Tutoring Systems. Technology Health Care. 2005;13 (5): 394-395. 2005.
[25] Carolina González, Juan C. Burguillo, Martín Llamas. A Case-Based Approach For Building Intelligent Tutoring Systems. 7th International Conference on Information Technology Based Higher Education and Training (ITHET) IEEE catalog Number: 06EX1387. Sydney (Australia). Julio 2006
[26] Carolina González, Juan C. Burguillo, Martín Llamas, Roberto Suenaga. Towards an Intelligent Tutoring System for training in HealthCare. 17th EAEEIE Annual Conference on Innovation in Education for Electrical and Information Engineering (EIE). Craiova (Rumanía). Junio 2006
[27] Carolina González, Martín Llamas, Juan C. Burguillo. Sistemas Inteligentes para Educación Médica. VI Congreso Iberoamericano de Telemática. (CITA 2006). Monterrey (Mexico). Mayo 2006
[28] Carolina González, Juan Carlos Burguillo-Rial and Martin Llamas-Nistal. A Qualitative Comparison of Techniques for Student Modeling in Intelligent Tutoring Systems. 36th Anual Frontiers in Education Conference (FIE). San Diego, California (EEUU). Octubre 2006.

[29] Carolina González, Juan C. Burguillo, Juan C. Vidal, Martín Llamas, Daniel A. Rodríguez. ITS-TB: An Intelligent Tutoring System to provide E-learning in Public Health. Comunicación seleccionada presentada por un coautor. 16th EAEEIE Conference. Lappeenranta (Finlandia). Junio 2005.
[30] Luis A. Sabucedo, Luis Anido. Hacia una plataforma semántica para servicios de eGovernment V Jornadas de Ingeniería Telemática Universidad de Vigo. ISBN: 84-8408-346-2. Vigo (España) 12-14 de septiembre de 2005
[31] Luis Álvarez Sabucedo, Luis Anido Rifón. Semantic Service Oriented Architectures for eGovernment Platforms. American Association for Artificial Intelligence (AAAI) Spring Symposium. AAAI Press. ISBN: 978-1-577335-267-9. Stanford University CA (USA) 27-29 de marzo de 2006
[32] Luis Álvarez Sabucedo, Luis Anido Rifón. Semantic support for eGovernment transactions. Developing services for a better public service.. Proceedings of the Interoperability of Government Services Conference (eGovInterop'06). Bordeaux (France) 20-24 de marzo de 2006
[33] Luis Álvarez Sabucedo, Luis Anido Rifón. Life Event Oriented Architecture for eGovernment Solutions. Proceedings of the IEEE EDA-PS 2006 workshop proceedings (EDA-PS). Chicago(EE.UU.) 18-19 de Septiembre de 2006
[34] Luis Alvarez Sabucedo, Francisco García-Sánchez, Luis Anido Rifón, Rodrigo Martínez-Béjar. Plataforma basada en anotación y procesamiento semántico para eGovernment
Proceedings of the JS-WEB 2006. ISBN: 84-690-2398-5. Santiago de Compostela (Spain) 16-17 de Noviembre de 2006
[35] Luis Álvarez Sabucedo, Luis Anido Rifón. Semantic Life Event Environment for eGovernment Services. : WebIST 2007. Pendiente de publicación. Barcelona (Spain) 3-6 de Marzo de 2007
[36] Fernández-Manjón, B., Sierra, J.L; Martínez Ortiz, I., Moreno Ger, P. Uso de estándares aplicados a TIC en Educación. Informe oficial n 16. Centro de Información y Comunicación Educativa (CNICE), Ministerio de Educación y Ciencia (2007, en prensa).
[37] Fernández-Manjón, B., López Moratalla, J., Martínez Ortiz, I., Moreno Ger, P. Objetos educativos y estandarización en e-learning: experiencias en el sistema <e-Aula>. RED. Revista de Educación a Distancia (ISSN1578-7680) Volumen: III (disponible on-line en http://www.um.es/ead/red/fernandez_manjon10.pdf) 2005
[38] Moreno Ger, Pablo; Martínez Ortiz, Iván; Sierra Rodríguez, José Luis; Fernández Manjón, Baltasar. Language-Driven Development of Videogames: The <e-Game> Experience.Proceedings of the 5th International Conference in Entertainment Computing (ICEC 2006).Lecture Notes in Computer Science, vol. 4.161, pp. 153-164. Springer-Verlag, 2006.
[39] Moreno-Ger, P; Sierra, J.L; Martínez-Ortiz, I; Fernández-Manjón, B. A Documental Approach to Adventure Game Development. Science of Computer Programming. En prensa
[40] Moreno Ger, Pablo; Martínez Ortiz, Iván; Sierra Rodríguez, José Luis; Fernández Manjón, Baltasar. A Descriptive Markup Approach to Facilitate the Production of e-Learning Contents. Proceedings of the 6th IEEE International Conference on Advanced Learning Technologies (ICALT 2006), pp. 19-21. IEEE Computer Society Press, 2006.
[41] Moreno-Ger, P., Martinez-Ortiz, I., Sierra-Rodríguez, JL:, Fernandez-Manjón, B. The (E-Game) Project: Facilitating The Development Of Educational Adventure Games Proceedings of Cognition and Exploratory Learning in Digital Age (CELDA 2005), pp 353-358, 2005.
[42] Martínez-Ortiz, I; Moreno-Ger, P; Sierra, J.L; Fernández-Manjón, B. Educational Modelling Languages: A Conceptual Introduction and a High-Level Classification. Computers and Education E-learning – from theory to practice. Springer. En prensa

[43] Martínez-Ortiz, I; Moreno-Ger, P; Sierra, J.L; Fernández-Manjón, B. Production and Deployment of Educational Videogames as Assessable Learning Objects. Proceedings of the First European Conference on Technology Enhanced Learning (EC-TEL 2006), Lecture Notes in Computer Science 4227, pp. 316 – 330. Springer. 2006
[44] Martínez Ortiz, Iván; Moreno Ger, Pablo; Sierra Rodríguez, José Luis; Fernández Manjón, Baltasar. Using DocBook and XML Technologies to Create Adaptive Learning Content in Technical Domains. International Journal of Computer Science and Applications, vol. III, nº II, pp. 91-108, 2006.
[45] Martínez Ortiz, Iván; Moreno Ger, Pablo; Sierra Rodríguez, José Luis; Fernández Manjón, Baltasar.<e-QTI>: A Reusable Assessment Engine. Postproceedings of the 5th International Conference on Web-based Learning (ICWL 2006). Lecture Notes in Computer Science, vol. 4.181, pp. 134-145. Springer Verlag, 2006.
[46] Martínez Ortiz, Iván; Moreno Ger, Pablo; Sierra Rodríguez, José Luis; Fernández Manjón, Baltasar. Production and Maintenance of Content Intensive Videogames: A Document-Oriented Approach. Proceedings of the International Conference on Information Technology: New Generations (ITNG 2006), Las Vegas, NV, USA, pp. 118-123. IEEE Computer Society Press, 2006.
[47] Martínez-Ortiz,I; Moreno-Ger,P; Sierra, J.L; Fernández-Manjón, B. A General Architecture for the Authoring and the Operationalization of e-Learning Applications with Educational Modelling Languages. Proceedings del 8º Simposio Internacional de Informática Educativa, SIIE 2006
[48] Martínez-Ortiz,I; Moreno-Ger,P; Sierra, J.L; Fernández-Manjón, B. Facilitando la Creación y Ejecución de Unidades de Aprendizaje a partir de Lenguajes de Modelado Educativo y de Objetos de Aprendizaje Reutilizables. Proceedings del III Simposio Pluridisciplinar Sobre Objetos y Diseños de Aprendizaje Apoyados en la Tecnología – OdA 2006
[49] Martínez-Ortiz, I., Moreno-Ger, P., Sancho-Thomas, P., Fernandez-Manjón, B. Using DocBook to Aid in the Creation of Learning Content. New Trends and Technologies in Computer-Aided Learning for Computer-Aided Design. pp 11-24, Springer (IFIP Series), 2005.
[50] Navarro Antonio, Fernández-Valmayor Alfredo. Conceptualization of Hybrid Websites, Internet Research (ISSN: 1066-2243), Emerald (en prensa, aceptado para su publicación en 2007).
[51] Navarro Martín, Antonio; Sierra Rodríguez, José Luis; Fernández-Valmayor Crespo, Alfredo; Hernanz, Héctor. >From Chasqui to Chasqui II: an Evolution in the Conceptualization of Virtual Objects. Journal of Universal Computer Science, vol. 11, n. 9, pp. 1.518-1.529. Springer Verlag, 2005.
[52] Navarro Martín, Antonio; Sierra Rodríguez, José Luis; Fernández-Valmayor Crespo, Alfredo; Fernández Manjón, Baltasar. A First Step Towards the Web Engineering Body of Knowledge. Proceedings of the 5th International Conference on Web Engineering (ICWE 2005). Lecture Notes in Computer Science, vol. 3.579, pp. 585-587. Springer-Verlag, 2005.
[53] Sancho Thomas, P., Martínez, I., Fernández-Manjón, B.. Semantic Web Technologies Applied to e-learning Personalization in <e-aula>. Journal of Universal Computer Science (ISSN: 0948-69x), 11 (9), pp 1470 -1481, 2005.
[54] Sierra, J.L; Moreno-Ger, P; Martínez-Ortiz, I; Fernández-Manjón, B. A Highly Modular and Extensible Architecture for an Integrated IMS based Authoring System: The <e-Aula> Experience. Software: Practice & Experience. En prensa. Publicado on-line en acceso avanzado el 2 de Noviembre de 2006 (DOI: 10.1002/spe.779)

[55] Sierra,J.L, Fernández-Valmayor, A. Universalizing Chasqui Repositories with a Flexible Importation / Exportation System. Computers and Education E-learning – from theory to practice. Springer. En prensa.
[56] Sierra Rodríguez, José Luis; Fernández-Valmayor Crespo, Alfredo; Fernández Manjón, Baltasar. A Document-oriented Paradigm for the Construction of Content-Intensive Applications. The Computer Journal, vol. 49, n. 5, pp. 562-584. British Computer Society, 2006.
[57] Sierra, J.L; Fernández-Vamayor, A. A Prolog Framework for the Rapid Prototyping of Language Processors With Attribute Grammars. Electronic Notes in Theoretical Computer Science 164(2), 19-36. 2006
[58] Sierra Rodríguez, José Luis; Fernández-Valmayor Crespo, Alfredo; Guinea, Mercedes; Hernaz, Héctor. >From Research Resources to Learning Objects: Process Model and Virtualization Experiences. Journal of Educational Technology & Society, vol. 9, n. 3, pp. 56-68. 2006.
[59] Sierra Rodríguez, José Luis; Fernández-Valmayor Crespo, Alfredo; Guinea, Mercedes. Exploiting Author-Designed Domain-Specific Descriptive Markup Languages in the Production of Learning Content. Proceedings of the 6th IEEE International Conference on Advanced Learning Technologies (ICALT 2006), pp. 515-519. IEEE Computer Society Press, 2006.
[60] Sierra Rodríguez, José Luis; Fernández Manjón, Baltasar; Fernández-Valmayor Crespo, Alfredo; Navarro Martín, Antonio. Document Oriented Development of Content-Intensive Applications.Int'l Journal of Software Engineering and Knowledge Engineering, vol. 15, n. 6, pp. 975-993. World Scientific, 2005.
[61] Sierra Rodríguez, José Luis; Fernández Manjón, Baltasar; Fernández-Valmayor Crespo, Alfredo; Navarro Martín, Antonio. Document-Oriented Software Construction Based on Domain-Specific Markup Languages. Proceedings International Conference on Information Technology: Coding and Computing (ITCC 2005), vol. II, pp. 392-397. IEEE Computer Society Press, 2005.
[62] Sierra Rodríguez, José Luis; Fernández-Valmayor Crespo, Alfredo; Fernández Manjón, Baltasar; Navarro Martín, Antonio. El Desarrollo Documental. Ingeniería de la Web y Patrones de Diseño, pp. 131-161. Pearson Prentice-Hall, 2005.
[63] Sierra Rodríguez, José Luis; Fernández-Valmayor Crespo, Alfredo; Guinea, Mercedes; Hernanz, Héctor; Navarro Martín, Antonio. Building Repositories of Learning Objects in Specialized Domains: The Chasqui Approach. Proceedings of the 5th IEEE International Conference on Advanced Learning Technologies (ICALT 2005), pp. 225-229. IEEE Computer Society Press, 2005.
[64] Sierra Rodríguez, José Luis; Fernández-Valmayor Crespo, Alfredo; Fernández Manjón, Baltasar; Navarro Martín, Antonio. Developing Content-Intensive Applications with XML Documents, Document Transformations and Software Components. Proceedings 31st Euromicro Conference on Software Engineering and Advanced Applications (EUROMICRO/SEAA 2005), pp. 4-11. IEEE Computer Society Press, 2005.
[65] Sierra Rodríguez, José Luis; Moreno Ger, Pablo; Martínez Ortiz, Iván; López Moratalla, Javier; Fernández Manjón, Baltasar. Building learning management systems using IMS standards: Architecture of a manifest driven approach. Proceedings of the International Conference on Web-based Learning (ICWL 2005). Lecture Notes in Computer Science, vol. 3583, pp. 144-156. Springer Verlag, 2005.
[66] Sierra Rodríguez, José Luis; Navarro Martín, Antonio; Fernández Manjón, Baltasar; Fernández-Valmayor Crespo, Alfredo. Incremental Definition and Operationalization of Domain-Specific Markup Languages in ADDS. ACM SIGPLAN Notices, vol. 40, n. 12, pp. 28-37. ACM, 2005.