

Jornada 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 fueron seleccionados para la jornada de seguimiento del año 2005. Dicha selección presenta una muestra significativa de la investigación nacional en Informática, reuniendo proyectos de distinta índole y naturaleza. La jornada se celebró en Granada el día 13 de Septiembre, como parte del I Congreso Español de Informática (CEDI).

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.

Septiembre 2005

José María Troya y Sascha Ossowski
Programa Nacional de Tecnologías Informáticas

Organización

La Jornada de Seguimiento de Proyectos en Tecnologías Informáticas ha sido organizada 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, y ha contado con la colaboración del Comité Organizador del I Congreso Español de Informática (CEDI), presidido por Alberto Prieto.

Programa Nacional de Tecnologías Informáticas

José María Troya Linero (Gestor del Programa TIN)

Ramón López de Arenosa (Jefe del Dep. de Tecnologías de la Producción y las Comunicaciones)

Sascha Ossowski (Colaborador del Programa TIN)

Índice

GLIRS-II: An information retrieval system based on fuzzy linguistic information and genetic algorithms TIC2003-007977	1
<i>Enrique Herrera Viedma</i>	
Agile Development and Maintenance of Applications based on Web Services (AgilWeb) TIC2003-02737-C02	9
<i>Miguel Toro Bonilla, Macario Polo Usaola</i>	
MODELADO DE PRESTACIONES DE SISTEMAS DE TIEMPO REAL TIC2003-06293	21
<i>Ramon Puigjaner, Carlos Juiz</i>	
Método para el desarrollo de materiales didácticos extendidos y su automatización TIC2003-03654	29
<i>Paloma Díaz Pérez</i>	
Web-i(2): Electronic Institutions: Extending the Core Notion TIC2003-08763-C02	37
<i>Pablo Noriega B.V., Sascha Ossowski</i>	
Web services, parallelism, evolution and complex systems: NadeWeb (nuevos algoritmos distribuidos y evolutivos en la web, new distributed evolutive algorithm on the web) TIC2003-09481-C04	52
<i>Juan Julián Merelo Guervós, Ken Sharman, Isabel Román</i>	
IDE-E: Technological advances for a Web-based National Spatial Data Infrastructure, convergence with the European initiative INSPIRE TIC2003-09365-C02	61
<i>Pedro R. Muro-Medrano, Miguel Angel Bernabé Poveda, Michael Gould Carlson</i>	
Connectionist and Rapid Prototyping for Analyzing and Forecasting Data Series TIC2003-08807-C02	80
<i>Miguel Delgado Calvo-Flores, Luis Jiménez Linares</i>	

Rules as a tool for the design and management of relational databases: algebraic foundations, deduction methods and efficient tools. TIC2003-08687-C02	92
<i>Inma P. de Guzmán, Olga Pons</i>	
Resolución de Problemas Combinatorios con Restricciones Suaves y Fuzzy Utilizando Algoritmos de Satisfactibilidad TIC2003-00950	104
<i>Teresa Alsinet</i>	
Study of the Systems with Uncertainty in Presence of a Fuzzy Similarity TIC2003-04564	112
<i>J. Recasens</i>	
Development and Evaluation of Metaheuristics for Constrain Satisfaction Problems TIC2003-04153	116
<i>Camino R. Vela, Jorge Puente, Cesar Alonso, Inés González, María Sierra and Ramiro Varela</i>	
TEFATE: Finite-state technologies applied to specialized translation TIC2003-08681-C02	123
<i>Mikel L. Forcada, Francisco Casacuberta Nolla</i>	
DELIA: Deduction and programming in non-classical logics for knowledge technology TIC2003-09001-C02	135
<i>Manuel Ojeda Aciego, David Pearce</i>	
The incorporation of the Mathematical Morphology, the Formal Concept Analysis and the Fuzzy Logic techniques and tools to the data cleaning, aggregating, reducing and mining stages of the Knowledge Discovery process TIC2003-08693	147
<i>Ramón Fuentes-González</i>	
Concept and Behavior Learning by Soft-Computing Techniques applied to the Mobile Robot Navigation problem TIC2003-04900	155
<i>Raúl Pérez</i>	

Architectures, Compilers and Applications in Multiprocessors TIC2003-06623	163
<i>Emilio López Zapata</i>	
Integrating qualitative and quantitative reasoning spatio-temporal models and their application to robotics TIC2003-07182	171
<i>M. Teresa Escrig</i>	
Improving the performance and services offered by PC clusters. Development of distributed multimedia applications TIC2003-08154-C06	182
<i>José Duato, Francisco Quiles, José Manuel García, Juan Manuel Orduña, Francisco Mora, José Manuel Claver</i>	
Computación natural con redes de procesadores evolutivos, y sus relaciones con las redes de neuronas y con los autómatas biomoleculares de Watson-Crick TIC2003-09319-C03	206
<i>Carlos Martín Vide, José María Sempere Luna, Luis Fernando Mingo López</i>	
GRID-IT: Investigación y Desarrollo de Servicios Grid - Aplicación a Modelos Cliente Servidor, Colaborativos y de Alta Productividad. TIC2003-01318	227
<i>Vicente Hernández García</i>	
Hybridizing Evolutionary Computation and Ant Colony Optimization. Applications to Fuzzy Rule Learning and Bioinformatics Problems. TIC2003-00877	235
<i>Oscar Cerdón</i>	
Development of High Level Software using Loosely Coupled Parallel Computers for the Optimal Design of Thermal Systems and Equipment. TIC2003-07970	243
<i>Asensio Oliva Llena</i>	
Sistema de Apoyo a la Visión mediante una Retina Artificial Neuromórfica a través de Realidad Aumentada TIC2003-09557-C02	251
<i>José Manuel Ferrández, Eduardo Fernández</i>	

Virtual trainers using low-cost platforms TIC2003-08933-C02	263
<i>Luis Pastor Pérez, Antonio Pérez Ambite</i>	
IVEM: Interfaces de Visión para Entornos Médicos TIC2003-00654	275
<i>Jordi Vitrià Marca</i>	
PASSIVIM: Panoramic based Summarization and Semantic Interaction with Digital Video for Multimedia Messages TIC2003-06075	283
<i>Xavier Binefa Valls</i>	
NIDO: Navegación e interacción con el usuario en el desarrollo de sistemas de información Web: Métodos, Técnicas y Herramientas TIC2003-00369	296
<i>Jesús Torre, María José Escalona, Antonia María Reina</i>	
Content-Based Music Processing (Procesamiento de Música Basado en su Contenido) TIC2003-07776-C02	304
<i>Xavier Serra Casals, Rafael Ramirez Melendez</i>	
MEDICI: Inpatient/Outpatient Monitoring for Diagnosis and Medical Research in Is- chaemic Cardiopathy TIC2003-09400-C04	312
<i>Roque Marín Morales, Ramón Ruiz Merino, Jesus María Rodríguez Presedo, Alfonso Palma Álvarez</i>	
The Application of Artificial Intelligence Techniques to Automatic Angiography Analysis using PACS to establish a Score for the effect of Coronary Stenosis TIC2003-07593	324
<i>Alejandro Pazos, Javier Pereira, Julián Dorado, Francisco Nóvoa, Alfonso Castro, Sabela Fernández, Bernardino Arcay</i>	

GLIRS-II: An information retrieval system based on fuzzy linguistic information and genetic algorithms TIC2003-07977

Enrique Herrera Viedma*
Dept. Of Computer Science and A.I
University of Granada

Abstract

In most existing Information Retrieval Systems (IRSs) users do not participate in the retrieval activity and systems have not appropriate tools to express correctly user queries. The main consequence is that the system performance and users' satisfaction degree are diminished. The Fuzzy Linguistic Modeling and Genetic Algorithms (GA) are Soft Computing tools useful for the design of IRS. The first one to manage the qualitative information in the formulation of queries by facilitating users the expression of their information needs, and GA for the optimization of queries as a tool that improves the relevance feedback (tool to define queries that allows the users participation in the retrieval processes). Our idea is to study the application of both tools in the design of IRSs to combine their advantages and to improve the effectivity of IRSs and the users' satisfaction degree.

Keywords: information retrieval, relevance feedback, query language, soft computing, fuzzy linguistic modelling, genetic algorithms.

1 Goals of the Project

This research project is a continuation of the project TIC-2002-03276 (GLIRS: Information retrieval systems based on fuzzy linguistic information and genetic algorithms). In that project a basic study on the development of IRS based on fuzzy linguistic information and GA is being carried out. In this occasion, we propose the development of a IRS based on fuzzy linguistic information and GA, called GLIRS-II (Genetic Linguistic Information Retrieval System), that incorporates more tools to help users to express better their information needs and better tools of relevance feedback based on GA. To do so, we focus on the developing the following specific goals:

- The design of a more expressive fuzzy linguistic multi-level weighted query subsystem that represents better the information needs by means of three elements: i) a scheme of multi-level weighting to associate weights with any element of a query (terms, sub-expressions, Boolean connectives and whole query), ii) the possibility of managing multi-granular and unbalanced fuzzy linguistic information to assess weights, and iii) an appropriate interface.

* Email: viedma@decsai.ugr.es

- The design of an evaluation subsystem of multi-level linguistic weighted queries that allows to manage multi-granular and unbalanced linguistic information.
- The design of relevance feedback tools based on GA for the optimization of multi-level linguistic weighted queries. We study two research areas: i) the use of new GA based strategies to model the relevance feedback tools, and ii) the design of new fitness functions, selection mechanisms and crossover operators more appropriate for relevance feedback.
- Implementation and evaluation of GLIRS-II. GLIRS-II will be a prototype of an IRS that integrates the above proposals. In such a way, GLIRS-II will be the issue of the project that stores all our proposals.

On the other hand, to develop our project we have the tools necessary to achieve the above goals. In particular, we have a good library to study the different advances about the development of information retrieval based on fuzzy linguistic modelling and genetic algorithms, we also have some computers to implement the different models of information retrieval systems, and finally, we also have some collections of documents to validate such models (CRANFIELD, TREC y CACM).

Finally, the approximate schedule of the project is the following:

Activities	First year (*)	Second year (*)	Third year (*)
	M1 M2		
Study of the art state	x x		
	M3 M7		
Design of the basic prototype GLIRS-II	x x x x		
	M8	M16	
Implementation of basic prototype GLIRS-II	x x x x x	x x x x	
		M13 M20	
Development of multi-level linguistic weighted query subsystem		x x x x x x x x	
		M16 M22	
Development of multi-level linguistic weighted query evaluation subsystem		x x x x x x	
		M23	M29
Development of relevance feedback tools		x x	x x x x x
Based on GA			
			M30 M36
Experimentation, evaluation of GLIRS-II			x x x x x x

2 Level of Success

The level of success achieved through the carrying out project has been important and very high. We have solved the nearly all the goals by implementing several prototypes of the different information retrieval

system models and an basic prototype of GLIRS-II. In the last phase of the project, we are integrating all the advances of those models in GLIRS-II.

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.

2.1 The Design of a More Expressive Fuzzy Linguistic Multi-Level Weighted Query Subsystem

We have developed a fuzzy linguistic multi-level weighted query subsystem by means of design of different models of information retrieval systems (IRSs).

Firstly, we have developed an IRS model that allows the use of linguistic multi-level weighted queries, in such a way, that a user could use any element of a query (atoms, operators, sub-expressions, whole query) to express his information needs. In this IRS model we use always the same linguistic term set. Then, we have designed some IRS models with multi-level weighted queries but with multi-granular linguistic information, that is assuming different term sets to assess the different concepts of the activity of a IRS, and with unbalanced linguistic information, that is assuming linguistic term sets whose distribution is not uniform and symmetrical with respect to the mid linguistic term. In both cases, the main problem is how to represent in a uniform way the different linguistic information. We have designed two methodologies to manage the multi-granular and unbalanced linguistic information using the 2-tuple linguistic representation model [8].

We have observed that these fuzzy linguistic query languages can be easily applied in the search engines on the Web, and then, we have proposed some models of information gathering distributed systems on the Web that use such query languages as a way to increase the expression possibilities of information needs.

Studying the information retrieval problem on the Web we find out that there exist two different approaches to focus the information access processes on the Web [1, 5, 9]: i) approaches based on the Classical Information Retrieval which retrieve information from user queries and provide relevance scores, and ii) approaches based on the Information Filtering tools which retrieve information from user profiles and provide recommendations. On the Web community it is known that information access models combining both approaches are a promising tool to improve the users' satisfaction degrees and the performance of information access systems. Then, in order to improve the performance of the Web search systems we have studied how to combine both, that is, IRS models together with filtering tools using linguistic modelling to represent the users' information needs and profiles. We also have studied how to generate recommendations in the filtering tools facilitating the user participation in the quality evaluation procedures of the objects/documents to recommend by means of the fuzzy linguistic modelling.

In all above models, the big problem to solve is to design adequate interfaces that allow users to provide their information needs by means of the linguistic information. We are designing a graphic interface that allows users to express their preferences by means of simple clicks on the mouse.

2.2 The Design of an Evaluation Subsystem of Multi-Level Linguistic Weighted Queries

For each one of the above fuzzy linguistic query languages we have defined its respective evaluation subsystem. To do that, we have defined in each IRS model three elements:

1. A strategy of evaluation of multi-level weighted queries to apply the different semantics associated with the query weights in a consistent way.
2. Some linguistic matching functions to interpret each one of the semantics of the query weights, and
3. appropriate soft aggregation operators of multi-granular and unbalanced linguistic information. We have defined such aggregation operators using 2-tuple linguistic hierarchies [3] and the family of the OWA operators [6, 7, 10, 11].

2.3 The Design of Relevance Feedback Tools Based on GA

We have studied how to improve the learning of the Boolean user queries and multi-weighted user queries by means of the multi-objective genetic algorithms [2, 4]. In such a way, we have provided new tools to improve the relevance feedback in the formulation of the queries. In particular we have used the multi-objective genetic algorithm called SPEA [12]. With these tools we improve the GA based procedures to learn queries existing in the literature.

We have observed that in the case of learning of multi-weighted user queries, sometimes we obtain a poor performance. Now, we are studying how to tune such algorithms by improving some of the usual operations of the GAs, as for example, the mutation, the crossover, the generation of the initial population, etc.

2.4 Implementation and Evaluation of GLIRS-II

Initially, we designed the basic architecture of system GLIRS-II, by incorporating the different parameters and variables (for example, different linguistic term sets with different cardinality, different fuzzy linguistic quantifier to compute the weighting vector of the OWA operators, different databases to validate our developments, etc) necessary to implement the IRS models that we have developed in the project. When we define a new IRS model we implement it and then, it is incorporated in the system GLIRS-II. As aforementioned, now we are designing an appropriate interface to facilitate users the use of the different query languages proposed and also to facilitate the use of the system GLIRS-II.

2.5 Relevant Results

Here, we are going to show the more relevant results that we have provided until this moment:

- We have introduced the concept of multi-weighted query language. We have defined different fuzzy linguistic query languages that allow to weigh the different levels of a query, to use linguistic weights, and to use multi-granular and unbalanced linguistic information. They can be consistently used in the search engines on the Web to improve the specification of the user information needs.
- We have defined for each multi-weighted fuzzy linguistic query language its respective evaluation method of queries that allows to evaluate coherently the different types of multi-weighted user queries.
- We have defined new semantic linguistic interpretation functions for the query weights to soften their evaluation.
- We have defined a method to help users in the formulation of their queries based on multi-objective GA, which improves the learning of the concept of user relevance with respect to other genetic proposals existing in the literature.

- We have defined a method based on multi-objective genetic algorithms to learn simple Boolean user queries and user queries with multiple weighting levels.

3 Indicators of Results

The main indicators that we can show about the results of the basic Research Project GLIRS-II are classified in the following categories: formation of students, publications, spanish projects, international and spanish collaborations.

3.1 Formation of Students

The results generated in this category have been important. A doctoral thesis has been defended and three PhD students have presented their research projects as a previous step to present the next year three doctoral thesis. In particular:

1. D^a. María Luque Rodríguez, member of Project and associate professor in Cordoba University, has finished her doctoral thesis titled "*Information Retrieval Models Based on Fuzzy Linguistic Information and Evolutionary Algorithms. Improving the Representation of Information Needs*" with Drs. Oscar Cordon and Enrique Herrera-Viedma as advisors. She presented it by March 2005.
2. D. Antonio Gabriel Lopez Herrera, PhD student in the Department of Computer Science and Artificial Intelligence (DECSAI), has presented his research Project titled "*Information Retrieval Models Based on Fuzzy Linguistic Information*" with the Dr. Enrique Herrera Viedma as advisor. He presented it by July 2005. He will finish his doctoral thesis the next year 2006.
3. D. Carlos Porcel Gallego, programmer in the UGR, has presented his research Project titled "*Information Access Models Based on Fuzzy Linguistic Information and Filtering Tools*" with the Dr. Enrique Herrera Viedma as advisor. He presented it by July 2004. He will finish his doctoral thesis at the end of this year 2005.
4. D. Jose Manuel Morales del Castillo, PhD student in the Department of Library Science, will present by September 2005 his research project titled "*Improvement of the Information Access on the Web Using Semantic Web and Fuzzy Linguistic Tools*" with the Drs. Eduardo Peis (member of the project) and Enrique Herrera Viedma as advisors. He will present his doctoral thesis the next year 2006.

3.2 Publications

The main publications correspond to the following categories:

- **Articles in international journals**

1. E. Herrera-Viedma, O. Cordon, M. Luque, A.G. Lopez, A.M. Muñoz. **An Information Retrieval System Based on Multi-Granular Linguistic Information.** International Journal of Approximate Reasoning 34 (3) (2003) 221-239.
2. E. Herrera-Viedma, F. Herrera, L. Martínez, J.C. Herrera, A.G. Lopez. **Incorporating Filtering Techniques in a Fuzzy Linguistic Multi-Agent Model for Gathering of Information on the Web.** Fuzzy Sets and Systems 148 (1) (2004) 61-83.
3. E. Herrera-Viedma, A.G. Lopez-Herrera C. Porcel. **Tuning the Matching Function for a Threshold Weighting Semantics in a Linguistic Information Retrieval System.** International Journal of Intelligent Systems 20 (2005) 921-937.

4. O. Cordon, E. Herrera-Viedma, M. Luque. **Improving the Learning of Boolean Queries by Means of a Multiobjective IQBE Evolutionary Algorithm.** Information Processing & Management. To appear, 2005.
5. E. Herrera-Viedma, G. Pasi, A.G. Lopez-Herrera, C. Porcel. **Evaluating the Information Quality of Web Sites: A Methodology Based on Fuzzy Computing with Words.** Journal of American Society for Information Science and Technology (JASIST). To appear, 2005.
6. E. Herrera-Viedma, A.G. Lopez-Herrera, M. Luque, C. Porcel. **A Fuzzy Linguistic IRS Based on 2-Tuple Modelling.** International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. Submitted, 2005.
7. E. Herrera-Viedma, E. Peis, J.M. Morales-del-Castillo, D. Gómez. **On the quality of Web sites based on XML documents: An evaluation model using fuzzy computing with words.** Journal of Information Science. Submitted, 2005.
8. E. Herrera-Viedma, A.G. Lopez-Herrera. **A Model of Information Retrieval System with Unbalanced Fuzzy Linguistic Information.** IEEE Trans. On Systems Man and Cybernetics, Part B. Submitted, 2005.
9. E. Herrera-Viedma. **Controlling the Retrieval of an Ordinal Fuzzy Linguistic Information Retrieval System Using a Multi-Level Weighting Scheme in the Formulation of Queries.** Journal of American Society for Information Science and Technology (JASIST). Submitted, 2005.

- **Book chapters**

1. O. Cordon, E. Herrera-Viedma, M. Luque. **A Realistic Information Retrieval Environment to Validate a Multiobjective GA-P Algorithm for Learning Fuzzy Queries.** In Soft Computing: Methodologies and Applications, Springer Verlag, 2005, pp. 299-310.
2. E. Herrera-Viedma, E. Peis, J.M. Morales-del-Castillo. **A Fuzzy linguistic multi-agent model based on Semantic Web technologies and user profiles.** In Soft Computing for Information Retrieval on the Web, Springer. Accepted, 2005.
3. E. Herrera-Viedma, C. Porcel, F. Herrera, L. Martínez, A.G. Lopez-Herrera. **Techniques to Improve Multi-Agent Systems for Searching and Mining the Web.** In Intelligent Data Mining: Techniques and Applications, Springer Verlag, 2005, pp. 463-486.
4. F. Herrera, E. Herrera-Viedma, L. Martínez, C. Porcel. **Information Gathering on the Internet Using a Distributed Intelligent Agent Model with Multi-Granular Linguistic Information.** In Fuzzy Logic and the Internet, Springer Verlag, 2005, pp. 95-116.

- **Articles in international and national conferences**

1. E. Herrera-Viedma, C. Porcel, A.G. Lopez, M.D. Olvera, K. Anaya. **A Fuzzy Linguistic Multi-Agent Model for Information Gathering on the Web Based on Collaborative Filtering Techniques.** Atlantic Web Intelligence Conference, AWIC'04. Cancún(México), 2004. Lect. Notes in Artificial Intelligence 3034, pp.3-12, 2004.
2. E. Herrera-Viedma, O. Cordon, M. Luque, A.G. Lopez. **A Model of Fuzzy Multi-Granular Linguistic IRS.** IPMU 2004, Vol II pp. 1365-1362, Perugia (Italia).
3. E. Herrera-Viedma, A.G. Lopez-Herrera, L. Hidalgo. **Tuning a Linguistic Information Retrieval System.** RASC 2004, Nottingham (UK). Proc. of the 5th Int. Conf. on Recent Advances in Soft Computing, pp. 506-511, 2004.
4. E. Herrera-Viedma, O. Cordon, M. Luque. **Improving the performance of ordinal fuzzy linguistic IRSs.** Proceedings of the XII Congreso Español sobre Tecnología y Lógica Fuzzy (ESTYLF 2004), Jaén (Spain), 509-514, 2004.

5. E. Herrera-Viedma, A.G. Lopez-Herrera, L. Hidalgo. **A New Linguistic Modelling of the Symmetrical Threshold Semantics**. Proceedings of the XII Congreso Español sobre Tecnología y Lógica Fuzzy (ESTYLF 2004), Jaén (Spain), 309-314, 2004.
6. E. Herrera-Viedma. **ON AGGREGATION OPERATORS FOR INFORMATION ACCESS ON THE WEB**. Proceeding of the 3rd International Summer School on Aggregation Operators (AGOP 2005), Lugano (Suiza), 141-146, 2005.
7. E. Herrera-Viedma, A. G. Lopez-Herrera, L. Hidalgo, C. Porcel. **Improving the Computation of Relevance Degrees in a Linguistic Information Retrieval System**. Proceeding of CEDI, Granada, 2005.
8. E. Herrera-Viedma, A. G. Lopez-Herrera, C. Porcel. **A New Model of Linguistic Weighted Information Retrieval System**, Proceeding of EUSFLAT, Barcelona, 2005.

3.3 Spanish Projects

We have participated with other spanish research groups and enterprises in two national projects:

1. **Network about Decisión Making, Modelling and Aggregation of Preferences (REDEMAP)**. Our research group is the coordinator of this spanish network in which we are collaborating with 20 or 22 research groups from different spanish universities and two enterprises (Telefonica I+D, Puleva) to join our proposals in the development of systems to help users in decision making, e.g. decision support systems, information retrieval systems, web multi-agent systems, recommendation systems, etc. We are studying with Telefonica I+D the possibility to ask an european project together with others international parnets about information filtering systems to help in decision making.
2. **Network about Soft Computing for information retrieval on the Web**. Our research group is an important research group of this spanish network. In this network, similarly, we collaborate with 15 or 20 research groups and an important enterprise, SOLUCIONA, in the development of information access systems based on Soft Computing tools. We are studying with Solucionona the possibility to ask an european project together with others international partners about information access systems on the Web based on Soft Computing tools.

On the other hand, at this moment we are organizing a regional regional project together with Dr. Jose Angel Olivas's research group, Solucionona enterprise about the use of Soft Computing tools in the design of information access systems. We shall ask it in September 2005 in the research program of the Castilla-La Mancha Community.

3.4 International and Spanish Collaborations

With respect to the international collaborations we have to point out two collaborations with two important european research groups, Dr. Gabriella Pasi's group from University of Bicocca (Milan) and Dr. Fabio Crestani's group from Univ. of the Strathclyde, Glasgow (UK). Both research groups are participating in several european projects about development of technologies to improve the information access tools on the Information and Knowledge Society. We are collaborating with them to organize in the future some european project. Until this moment our collaboration has produced the following results:

1. **Edition of a special issue** in the *Journal of American Society for Information Science and Technology (JASIST)*: E. Herrera-Viedma and G. Pasi (Eds.). Special issue on Soft approaches in information retrieval and information access on the Web. JASIST, to appear in 2006.
2. **Organization of two invited sessions together** with Dr. Gabriella Pasi about Soft Computing approaches in two International Conferences EUSFLAT 2003 (Zitau, Germany) and IPMU 2004 (Perugia, Italy).

3. **Edition of a book in Springer-Verlag:** E. Herrera-Viedma, G. Pasi and F. Crestani (Eds.). Soft approaches to information access on the Web. Springer Verlag. To appear in 2006.
4. **Edition of a special issue** in the *International Journal of Intelligent Systems (IJIS)*: G. Pasi and E. Herrera-Viedma (Eds). Special issue on Aggregation operators in the information management and access processes. IJIS, in preparation.

With respect to the national collaborations we have collaborated mainly with the Dr. Jose Angel Olivas's research group, in particular organizing two invited sessions about the use of Soft Computing Tools in Information Retrieval on the Web in the conferences:

1. National Conference on Fuzzy Logic and Technology (Estylft 2004) celebrated in Jaen in September 2004, and
2. European Conference on Fuzzy Logic and Technology (EUSFLAT 2005) to be celebrated in Barcelona in September 2005.

4 References

- [1] N.J. Belkin and W.B Croft. Information filtering and information retrieval: Two sides of the same coin?. *Communications of the ACM*, 35(12) (1992) 29-38.
- [2] C.A. Coello, D.A. Van Veldhuizen, G. B. Lamant. *Evolutionary Algorithms for Solving Multi-Objective Problems*. Kluwer Academic Publishers, (2002).
- [3] O.Cordón, F.Herrera, I.Zwir. Linguistic modelling by hierarchical systems of linguistic rules. *IEEE Transactions on Fuzzy Systems*, 10(1) (2001) 2-20.
- [4] K. Deb. *Multi-objective Optimization using Evolutionary Algorithms*. Wiley, (2001).
- [5] U. Hanani, B. Shapira, P. Shoval, Information filtering: Overview of issues, research and systems. *User Modeling and User-Adapted Interaction* 11 (2001) 203-259.
- [6] F. Herrera, E. Herrera-Viedma, Aggregation operators for linguistic weighted information. *IEEE Trans. on Systems, Man and Cybernetics, Part A: Systems and Humans*, 27 (5) (1997) 646-656.
- [7] F. Herrera, E. Herrera-Viedma, J.L. Verdegay, Direct approach processes in group decision making using linguistic OWA operators. *Fuzzy Sets and Systems*, 79 (1996) 175-190.
- [8] F. Herrera and L. Martínez, A 2-tuple fuzzy linguistic representation model for computing with words, *IEEE Transactions on Fuzzy Systems* 8:6 (2000) 746-752.
- [9] P. Resnick, H.R. Varian, Guest Editors. Recommender systems. *Communications of the ACM*, 40 (3) (1997), 56-89.
- [10] R.R. Yager. On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Transactions on Systems, Man, and Cybernetic*, 18 (1988) 183-190.
- [11] R.R. Yager. Quantifier Guided Aggregation Using OWA Operators. *Int. J. of Intelligent Systems*, 11 (1996) 49-73.
- [12] E. Zitzler, K. Deb, L. Thiele. Comparison of Multiobjective Evolutionary Algorithms: Empirical Results. *Evolutionary Computation*, 8(2) (2000)173-195.

Agile Development and Maintenance of Applications based on Web Services (AgilWeb) TIC2003-02737-C02

Miguel Toro Bonilla *
University of Sevilla

Macario Polo Usaola †
Univ. of Castilla-La Mancha

Abstract

The incredible successfulness of the Internet has paved the way for an industry devoted to developing and running web services. Our main goal in this research project is to provide a framework for the agile development and maintenance of applications that build on them. In this report, we provide an insight into its current status.

Keywords: methodologies, web services, automatic provisioning.

1 Introduction

The term web service is fairly self-explanatory since it refers to having access to services on the Web. However, there is more than this behind the scenes since the common use of the term also refers to the SOA reference architecture, standard protocols such as HTTP or SOAP, repositories such as UDDI, and data languages based on XML. The promise is that they will help simplify enterprise application integration and will bring new opportunities to outsource services on the Web. According to the Aberdeen Group, the percentage of companies opting for outsourced web services grew to more than 60% in 2004, and it is likely to increase as e-commerce becomes more mainstream and a must for almost every business.

As of the time of submitting AgilWeb, however, the process of building and maintaining a web application based on outsourced web services was not a small undertaking since it was a relatively new technology. This motivated us to work on an appropriate framework both from a conceptual and a practical point of view, and to provide software engineers with appropriate tools and infrastructure.

In this report, we present an insight into the current development of the 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; in Section 4, we provide some hints on the relevance of our contributions; finally, our main conclusions and the work that remains to be done are summarised in Section 5.

*Email: mtoro@lsi.us.es

†Email: macario.polo@uclm.es

2 Goals of the project

The main goal of AgilWeb was to devise a framework for the development and maintenance of applications based on web services. Their main feature is that time to market is usually very short, changes to the initial requirements are very frequent, and the risk of failure is high. The framework should then build on agile practices so as to keep changes under control, manage knowledge and expertise gained from previous projects to reduce development and maintenance costs, assess risks to help managers decide whether to embark or not on a new project, and provide an appropriate infrastructure for provisioning web services.

In summary, the framework relies on the following pillars and the schedule of tasks that is shown in Figure 2:

Methodology: Since we thought that the research on processes was rather mature and had reached the industry, we did not want to devise another process, but new guidelines and techniques in the context of web services. The idea was then to devise kind of methodological fragments ready to be integrated in well-established software processes. In summary, the main working areas and goals were the following:

1. Guidelines and notation: devise guidelines for requirements elicitation, analysis, design (USE-3.1), re-engineering (UCLM-2), and refactoring (UCLM-4); design a quality-driven reference architecture (USE-3.2); design an extension to UML to annotate software models with quality constraints (USE-3.3).
2. Knowledge management: turning intuitive knowledge about the applications to be built into formalised software product lines (USE-3.4); developing an explicit system to manage knowledge and expertise (UCLM-6).
3. Risk management: adapt a number of existing techniques to identify risks in our research context and react to them (UCLM-7).

Infrastructure: A methodological fragment is quite useless unless one can count on a reference infrastructure to support it. Summing up, the main goals were the following:

1. Provisioning system: design a trading system to outsource web services using quality criteria (USE-2.1, UCLM-5); make it interoperable by means of well-established industrial standards (USE-2.2); design automated testing techniques so that a trader can check automatically whether a web service can be trusted from a functional point of view (UCLM-3).
2. Advanced interaction models: refine previous results on multiparty interaction models to adapt them to a setting based on web services (USE-3.5).
3. Repositories: study current repositories of web services (USE-4.1); devise techniques to integrate human-friendly repositories (USE-4.2).

Tools: Last, but not least, engineers must have a set of tools to apply the guidelines, to reuse knowledge or to design complex multiparty interactions on the Web. The goal was to implement them during the third year (USE-5, UCLM-5, UCLM-8, UCLM-9).

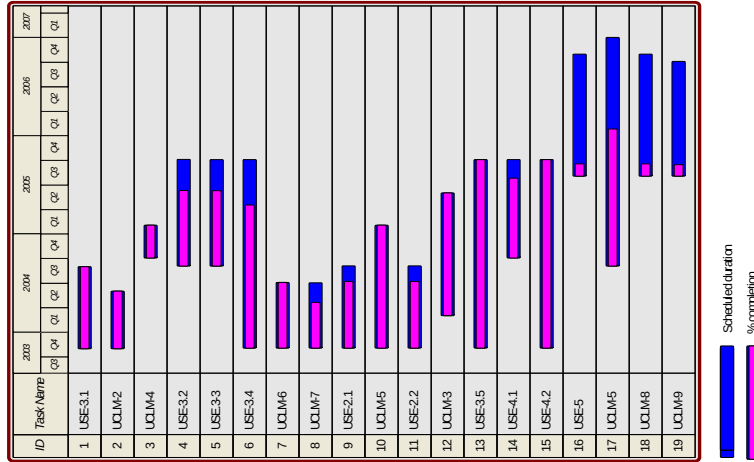


Figure 1: Chronogram of activities.

3 Current development

3.1 Difficulties and deviations

AgilWeb was awarded with budget to hire four technicians, but we soon realised that it was difficult to find qualified people. In addition, there were unexpected delays in the release of the final draft of WS-Agreement, WS-Resources, and WS-Negotiation. As a result, we began working on tasks USE-2.1 and USE-2.2 late. There was also a significant budget cut regarding the group of task UCLM, and this has had an important impact on the development of task UCLM-7. Currently, we are in arrears with them.

The project was assigned an FPI student. Unfortunately, he decided to move to the industry. This was not problematic since one of the collaborators from the industry (David Benavides) decided to work on the project full time, and a new junior researcher was invited to join it (Pablo Trinidad). The tasks in which the FPI student was involved have progressed at an adequate pace since then.

We have added a new task on studying ontologies since we found that current repositories of web services have sprouted chaotically. It is not surprising then that the terminology they use is inconsistent and makes it difficult to integrate them. This is why we decided to work on the design of a consensus ontology and divided the team in charge of task USE-4.1 into two groups: José M. Cañete and Francisco Galán to work on ontologies, and José L. Arjona to study current repositories and techniques to extract knowledge from them (USE-4.2).

There has been a little deviation regarding task USE-3.1. As of the time of submitting AgilWeb, it was clear that the role of use cases would gain importance, but we thought this was a well-established field. Soon after, we realised that there was a significant lack of techniques to verify use cases, and we then decided to invest a little effort on researching this topic. This deviation led to the development of a new PhD thesis by Beatriz Bernárdez.

Notice that not all of the deviations are negative. Tasks USE-3.5, USE-4.2 finished before it was expected, and the work on tasks UCLM-5, UCLM-8 and UCLM-9 is currently in progress.

This is the reason why we think we will be able to compensate for the negative deviations in tasks USE-2.1 and USE-2.2.

3.2 Guidelines and notations

USE-3.1: Here, our work focuses on multi-agent systems, since most web services back ends have an implicit degree of autonomy, proactivity and reactivity that allows to view them as agents. The goal was to define a methodological fragment to model, analyse, and design such systems, and the idea was to build on the results on requirements elicitation we developed in previous projects, namely GEOZOCO (TIC2000-1106-C02), and MENHIR (TIC97-0593-C05). We have also developed a number of metric-based heuristics to verify use cases.

UCLM-2: We have devised a technique that allows to re-engineer a legacy system so that it can be transformed semi-automatically into a set of web services ready to outsource. Most of the results build on the work that was done in projects MANTIS (TIC1FD97-1608) and MEDEO (TIC2000-1673-C06).

UCLM-4: We have developed a number of guidelines to refactor the micro-architecture of an object-oriented system and adapted them web services.

USE-3.2: The core service of our architecture is the quality trader, which offers the basic services for the procurement of web services, i.e., consistence, conformance and optimality checks. Key features of the proposal are that the language it uses is symmetric, it allows for two-way matching, temporal constraints, uncontrollable and controllable quality parameters. We have studied the limits for it to be efficient, and we are currently working on using OWL to define an ontology of quality concepts.

USE-3.3: We have studied how to use QRL to annotate the UML models with which we describe our reference architecture.

3.3 Knowledge management

USE-3.4: After a thoughtful study of the state of the art on feature models, we concluded that their most important drawback was the lack of automated support tools. Therefore, we started working on this, and the result is a prototype that is available at <http://www.tdg-seville.info/topics/spl>.

UCLM-6: We have developed a number of mechanisms to describe and organise both the sources and kinds of knowledge involved in the development of a typical software project. The validation was performed in the context of a software maintenance process using SPEM. As a result, we have defined an ontology and an agent-based support system. Furthermore, we have designed a system that uses CBR techniques to reuse maintenance knowledge.

3.4 Risk management

UCLM-7: We have worked on risk management in the context of agile methodologies, but we are in arrears with this task. So far, we have prepared a research report with our main conclusions.

3.5 Provisioning system

USE-2.1: We are currently working on the design of a trading system to outsource web services. So far, we have identified the roles involved, the protocols to connect them, and a general purpose negotiation framework. We still need to refine the protocols and to study their properties regarding correctness, completeness and complexity.

UCLM-5: We have implemented a tool that helps determine whether two web services are interchangeable, despite their interfaces being different. It requires metadata to describe an interface, the constraints on each method, and the protocol that governs the order in which the methods can be invoked; building on this information, it is able to produce a similarity metric to decide whether two services are interchangeable. Currently, we are working on adjusting the similarity criteria and thresholds in practice.

USE-2.2: The trading system is based on current state-of-the-art industrial standards such as WS-Agreement, WS-Resource, WS-Negotiation, and UDDI. The work we have developed is fully compatible with these standards, which ensures that it will be easily interoperable.

UCLM-3: We have proposed a new technique to automate the testing of black-box components, which is the key to testing web services. We have also implemented an accompanying tool that builds on a metamodel and uses introspection to generate and execute test cases.

3.6 Interaction models

USE-3.5: In project GEOZOCO (TIC2000-1106-C02-01), we developed and implemented a synchronous multiparty interaction model that was suitable for open systems. Our improvements are twofold: first, we have worked on a new fairness notion that solves two well-known problems of previous approaches since it deals with both fair finiteness and conspiracies; second, we have worked on an asynchronous extension and we have developed a number of techniques to refine and/or abstract interactions and to compose and/or decompose interaction models.

3.7 Repositories

USE-4.1: We have studied current web services repositories, and the conclusion is that most of them are user friendly and lack the ability to search for services using quality criteria. Our conclusions helped us define our trading system and design a knowledge extraction system able to deal with a user-friendly repository programmatically.

USE-4.2: We have designed and implemented a system to extract knowledge from the Web that has proved useful to have access to user-friendly repositories programmatically. It has been validated with three popular web sites.

4 Relevance of our results

4.1 Preliminaries

Before the project was funded, we started working on a number of activities whose goal was to have a good understanding of the technologies involved in the construction of a typical business solution. We got in contact with Dr. Wrembel¹ and embarked on two projects whose goal was to compile a series of articles on innovative solutions to common development problems, with an emphasis on web services. The result was an special issue of Inderscience's Journal of Web Engineering and Technology [30] and a book that was published by Nova Science [29]. A little later, we embarked on editing a special issue of Springer's Journal of Universal Computer Science on challenges and breakthroughs on software engineering [28].

As a result, we learnt what the main problems in this area are, and we established a network of international contacts that has proved fundamental to the development of AgilWeb.

4.2 Guidelines and notations

USE-3.1: Regarding modelling and analysis, our publications are [53, 54, 55, 56, 57, 58]. They were cited in [33] and have been classified as recommended bibliography by the Modelling and Methodology Technical Committees of FIPA. We have collaborated with the research groups by Dr. Odell², Dr. Cossentino³, Dr. Levy⁴, and Dr. Hinchey⁵. These collaborations led to the publication of [56], a collaboration agreement between NASA and the University of Seville, and a three-month research visit to the NASA Goddard Software Engineering Lab and to Intelligent Automation Inc. to keep working on our methodological fragment and to explore its applications to the industry. Most of these results are part of Joaquín Peña's PhD thesis [59], which was defended in July 2005.

Regarding the verification of use cases, the publications are [14, 15, 16, 17, 18, 19, 20, 21, 32], and they constitute the core of Beatriz Bernárdez's PhD thesis [22], which was defended in December 2004 and is the result of a close collaboration between both groups involved in AgilWeb since it was co-advised Amador Durán and Marcela Genero. Given the good results obtained, we are now active members of the National Research Network on Empirical Software Engineering (TIC2002-11494-E), which is led by Dr. Juristo⁶.

UCLM-2: The publications related to this task are [40, 41, 42, 43, 44, 45, 46, 64]. We have contacted Mr. García⁷. and work in closed co-operation with him regarding this task,

¹Dr. Robert Wrembel is a Professor of Database Systems with the University of Poznań.

²Dr. James Odell is the chair of the OMG Analysis and Design Task Force.

³Dr. Massimo Cossentino is the chair of the Methodology Technical Committee of FIPA.

⁴Dr. Renato Levy is the research head of Intelligent Automation Inc.

⁵Dr. Michael Hinchey is the head of the Software Engineering Laboratory of the NASA

⁶Dr. Natalia Juristo is a Professor of Software Engineering at the Universidad Politécnica de Madrid.

⁷M.Sc. Ignacio García is a lecturer with the University Rey Juan Carlos.

UCLM-3, and UCLM-5. We have also established a co-operation agreement with Soluziona Software Factory, S.A. to keep working in this task. Many of the results we have produced were developed in the context of Félix Ó. García and Javier Garzás's PhD's thesis.

UCLM-4: The main results have been produced in the context of Javier Garzás's PhD thesis.

USE-3.2: The publications related to this task are [49, 50, 51], and they have got citations from [23, 74]. Furthermore, [52] was presented in a conference with an acceptance ratio of 21% and was selected for publication in a major journal [73]. Recently, we have collaborated with Dr. Tosic⁸ whose research is very related to ours. We have developed a prototype that is available at <http://www.tdg-seville.info/topics/procurement.html>. Most of these publications constitute the core of future Octavio Martín-Díaz's PhD thesis, which is expected to be defended during the first semester of 2006.

USE-3.3: See USE-3.2. Both tasks are intimately related.

4.3 Knowledge management

USE-3.4: Our main publications are [7, 8, 12, 13, 75, 31]. We would like to mention that [12] was presented in a conference whose acceptance ratio was 14%, and that it was referenced by Dr. Batory⁹ in [5, 6]. We have also established a network of contacts with whom we are working on this task, namely: Dr. Batory, Dr. Smith¹⁰, Dr. Montes de Oca¹¹, and Dr. Díaz¹². Some of the papers mentioned above were written in co-operation with them, and David Benavides is currently in the course of a three-month research visit to Dr. Smith's Laboratory.

UCLM-6: The publications related to this task are [25, 26, 27, 36, 47, 65, 66, 67, 68, 69, 70, 71, 76, 77]. We have established a collaboration agreement with Dr. Canfora¹³, and we are working with him in the development of this task. Félix Ó. García has recently paid a six-month research visit to his department, and he is currently on a new four-month visit. Furthermore, we have established a network of contacts with whom we are working within the context of Experimental Software Engineering Network (IST-2001-37482) and a Spanish network called "Métodos de Investigación y Fundamentos Filosóficos en Ingeniería del Software y Sistemas de Información" (TIC2002-12378-E). Many of the results we have produced were developed in the context of Félix Ó. García PhD's thesis.

4.4 Risk management

UCLM-7: So far, we have just produced a research report [62].

⁸Dr. Vladimir Tosic is with the Department of Computer Science of the University of Western Ontario.

⁹Dr. Don Batory leads the group on Software Product Lines at the University of Texas at Austin.

¹⁰Dr. Barbara Smith is with the Cork Constraint Computation Centre, Ireland.

¹¹Dr. Carlos Montes de Oca is a senior researcher with the CIMAT, México.

¹²Dr. Óscar Díaz is a Professor of Software Engineering with the University of the Basque Country

¹³Dr. Gerardo Canfora is a Professor with the Università degli Studi del Sannio.

4.5 Provisioning system

USE-2.1 and USE-2.2: The publications related to this task are [34, 63, 35]. We have established relationships with Dr. Ludwig¹⁴, Dr. Zhang¹⁵, and Dr. Alonso¹⁶. These collaboration has led to a three-month research visit to the IBM T.J. Watson Research Center that will be held during the forth quarter of 2005. Furthermore, we got additional budget to pay a six-month research visit to Dr. Tavangarian's institute¹⁷. There, we collaborated on setting up a testing bed for our trader in the context of web services for mobile e-learning services.

UCLM-5: We are working in co-operation with Dr. Augusto¹⁸. Andrés Flores will soon pay a research visit to his department to keep working on this task.

UCLM-3: The publications related to this task are [37, 38, 39, 48, 61] and the results are part of Javier Garzás's PhD thesis. Currently, we are members of the Experimental Software Engineering Network (IST-2001-37482) and are working in co-operation with other members in the development of this task. Our co-operation with Soluziona Software Factory, S.A. is also closely related to this task.

4.6 Interaction models

USE-3.5: Our main publication is [60]. We have established relationships with Dr. Fiadeiro¹⁹, and the results were published in [28]. Most of the work on fairness was at the core of David Ruiz's PhD thesis [72]. The results on our new version of the multiparty interaction model and the associated techniques for refinement/abstraction or composition/decomposition were developed in the context of Joaquín Peña's PhD thesis [54, 55, 56, 59].

4.7 Repositories

USE-4.1: Recall that this task was divided into two new tasks regarding the study of ontologies and the repositories. As for the former, our only publication is [24]. Keep reading for additional details on the latter.

USE-4.2: The publications regarding this task are [1, 2, 3]. We have established relationships with Dr. Sykara²⁰, and we got additional budget for José L. Arjona to pay a six-month research visit to her laboratory. Unfortunately, the visit was not carried out due to personal problems. However, the collaboration keeps alive and has led to José L. Arjona's PhD thesis [4], which was defended in March 2004.

¹⁴Dr. Heiko Ludwig co-leads the work on WS-Agreement at the IBM T.J. Watson Research Center.

¹⁵Dr. Lian J. Zhang leads the group on outsourcing web services at the IBM T.J. Watson Research Center.

¹⁶Dr. Gustavo Alonso leads the research group on web services architectures at the ETH Zürich.

¹⁷Dr. Djamshid Tavangarian leads a research group on mobile devices of the University of Rostock.

¹⁸Dr. Juan C. Augusto is a Lecturer with the University of Ulster at Jordanstown.

¹⁹Dr. José L. Fiadeiro is a Professor of Software Science and Engineering with the University of Leicester.

²⁰Dr. Katia Sykara is a Professor of Intelligent Systems with the Carnegie Mellon University

5 Conclusions and further work

Although we had problems with our initial schedule, we have been able to overcome them and currently almost all of the tasks that were planned are progressing adequately. Only a couple of tasks are a little delayed due to the problems we had to hire qualified people and the delay to publish the final release of the standard on which we rely. We do not think this will be a serious problem since we plan on finishing them in a few months.

We have produced a number of publications in major journals and conferences, and 11 PhD thesis are currently under development and will conclude before or little after the project is finished, namely: Ignacio García (cf. UCLM-2, UCLM-3, UCLM-4, UCLM-5), Gabriela Aranda (cf. UCLM-6), Andrés Flores (cf. UCLM-2, UCLM-3, UCLM-4, UCLM-5), Óscar Rodríguez (cf. UCLM-6), Juan P. Soto (cf. UCLM-6), Manuel Martínez (cf. UCLM-6), Mateus Ferreira (cf. UCLM-6), Pablo Fernández (cf. USE-2.1, USE-2.2), Manuel Resinas (cf. USE-2.1, USE-2.2), Octavio Martín (cf. USE-2.1, USE-3.2, USE-3.3), and David Benavides (cf. USE-3.4).

As a conclusion, the results we planned on producing are almost ready and we can embark on the final stage, that consists mostly of implementing the tools we have designed.

References

- [1] J.L. Arjona, R. Corchuelo, J. Peña, D. Ruiz. First International Atlantic Web Intelligence Conference AWIC 2003: Coping with Web Knowledge. Springer-Verlag LNAI 2663. Madrid, Spain. pp. 165-178. 2003
- [2] J.L. Arjona, R. Corchuelo, D. Ruiz, J. Peña, M. Toro. 15th Conference on Advanced Information Systems Engineering CAISE'03: Knowledge Channels. Bringing the Knowledge in the Web to Software Agents. 2003
- [3] J.L. Arjona, R. Corchuelo, M. Toro. International Conference on Web Intelligence IEEE/WIC 2003: A Knowledge Extraction Process Specification for Today's non-semantic web. pp. 61-68. 2003
- [4] J.L. Arjona. Bringing the Knowledge on the Web to Software Agents. University of Sevilla. 2005
- [5] D. Batory. Feature models, grammars, and propositional formulas. To be published at Software Product Line Conference (SPLC 2005), 2005
- [6] D. Batory. A tutorial on feature oriented programming, the ahead tool suite. In Summer school on Generative, Transformation Techniques in Software Engineering, 2005
- [7] D. Benavides, A. Ruiz-Cortés, R. Corchuelo, A. Durán. Seeking for extra-functional variability. In Proceedings of the ECOOP Workshop on Modeling Variability for Object-Oriented Product Lines, 2003
- [8] D. Benavides, A. Ruiz-Cortés, R. Corchuelo, O. Martín-Díaz. SPL needs an automatic holistic model for software reasoning with feature models. In International Workshop on Requirements Reuse in System Family Engineering, pages 27-33, Madrid, Spain, 2004. Universidad Politécnica de Madrid
- [9] D. Benavides, A. Ruiz-Cortés, M.A. Serrano, C. Montes de Oca. A first approach to build product lines of multi-organizational web based systems (mows). LNCS, Innovative Internet Community Systems, IICS 2004 3473:91-98, 2005
- [10] D. Benavides, A. Ruiz-Cortés. A first approach to build product lines of mows. In Proceedings of the ZOCO02 workshop, El Escorial, Madrid, 2002
- [11] D. Benavides, A. Ruiz-Cortés, P. Trinidad. Coping with automatic reasoning on software product lines. In Proceedings of the 2nd Groningen Workshop on Software Variability Management, 2004
- [12] D. Benavides, A. Ruiz-Cortés, P. Trinidad. Automated reasoning on feature models. LNCS, Advanced Information Systems Engineering: 17th CAISE, LNCS(3520):491-503, 2005
- [13] D. Benavides, A. Ruiz-Cortés, P. Trinidad. Using constraint programming to reason on feature models. In The Seventeenth International Conference on Software Engineering, Knowledge Engineering (SEKE'05), July 2005

- [14] B. Bernárdez, A. Durán, M. Genero. Metrics for Software Conceptual Models, chapter of Metrics for Use Cases: A Survey of Current Proposals. Imperial College, pp 59–98, 2005
- [15] B. Bernárdez, A. Durán, M. Genero. Empirical Evaluation, Review of a Metrics-Based Approach for Use Case Verification. *Journal of Research, Practice in Information Technology, Special Collection on Requirements Engineering*, 36(4):247–258 2004
- [16] B. Bernárdez, A. Durán, M. Toro, M. Genero. Una Propuesta para la Verificación de Requisitos basada en Métricas. *Procesos y Métricas de las Tecnologías de la Información*, (1) 2:13–24, 2004
- [17] B. Bernárdez, A. Durán, M. Genero. An Empirical Review of Use Case Metrics for Requirements Verification. *Actas del 1st Software Measurement European Forum SMEF'2004*. pp. 216–225. 2004
- [18] B. Bernárdez, A. Durán, M. Genero, A. Ruiz. Empirical Assessment of a Defect Detection Technique for Use Cases. *Actas del 2nd Workshop on Software Quality del 26th International Conference on Software Engineering (SWSQ-ICSE'2004)*. pp 72–77, 2004
- [19] B. Bernárdez, M. Genero, A. Durán, M. Toro. A Controlled Experiment for Evaluating a Metric-Based Reading Technique for Requirements Inspection. *Actas del 10th IEEE International Software Metrics Symposium (METRICS'2004)*. pp 257–269. 2004
- [20] B. Bernárdez, A. Durán, M. Toro. Una Revisión Empírica de Heurísticas de Verificación de Casos de Uso basadas en Métricas. *Actas de las VIII Jornadas de Ingeniería del Software y Bases de Datos*. pp. 147–156, 2003
- [21] B. Bernárdez, A. Durán, M. Toro, M. Genero. MBR: Una Técnica de Detección de Defectos para Casos de Uso basada en Métricas. *Actas de las IX Jornadas sobre Innovación y Calidad del Software*. pp. 61–68. 2004
- [22] B. Bernárdez. Una Aproximación Empírica al Desarrollo de Heurísticas basadas en Métricas para Verificación de Requisitos. E.T.S. de Ingeniería Informática. Dpto. de Lenguajes y Sistemas Informáticos. Universidad de Sevilla. 2004
- [23] D. Bianchini, V. Antonellis, M. Melchiori. Qos in ontology-based service classification and discovery. In 4th. DEXA Workshop on Web Semantics DEXA-WS'04, pages 145–150, Zaragoza, Spain, 2004. IEEE Computer Society Press
- [24] J.M. Cañete, F.J. Galán. Towards a Theory on the Role of Ontologies in Software Engineering Problem Solving. *Fundamental Approaches to Software Engineering, FASE'05*, 2005
- [25] G. Canfora, F. García, M. Piattini, F. Ruiz, A. Visaggio. A Family of Experiments to Validate Metrics for Software Process Models. *The Journal of Systems and Software*. To be published
- [26] A. Cechich, M. Piattini. Quantifying COTS Component Functional Adaptation. *Software Reuse: Methods, Techniques and Tools: 8th International Conference (ICSR 2004), LNCS(3107)*: 195-204, 2004
- [27] A. Cechich, M. Piattini. Balancing stakeholder's preferences on measuring COTS component functional suitability. *Sixth International Conference on Enterprise Information Systems (ICEIS 2004)*, vol. 3, pp. 115–122. 2004
- [28] R. Corchuelo, A. Ruiz-Cortés. Breakthroughs, Challenges in Software Engineering. *Journal of Universal Computer Science*, 10(4). 2004
- [29] R. Corchuelo, A. Ruiz-Cortés, R. Wrembel. *Technology Supporting Business Solutions*. Nova Science. 2003
- [30] R. Corchuelo, A. Ruiz-Cortés, R. Wrembel. Web, Database Technologies in Business Solutions. *Journal of Web Engineering, Technology*, 1(3). 2003
- [31] A. Durán, D. Benavides, J. Bermejo. Applying system families concepts to requirements engineering process definition. In *Software Product-Family Engineering*, volume 3014 of LNCS. Springer-Verlag
- [32] A. Durán, B. Bernárdez, M. Genero, M. Piattini. Empirically Driven Use Case Metamodel Evolution. *Actas del 7th UML, Lisboa, 2004. Lecture Notes in Computer Science*, 3273:1–11. Springer Verlag, 2004
- [33] L. Ehrler, S. Cranefield. Executing Agent UML Diagrams. *AAMAS 2004*. pp. 906-913. 2004
- [34] P. Fernández, M. Resinas, R. Corchuelo. On the automatic agreement procurement of web services. *Actas del I Congreso Español de Informática (CEDI)*, 2005. (To be published)
- [35] P. Fernández, M. Resinas, R. Corchuelo. Automatic creation of agreements in a service-oriented scenario. In Frank Columbus, editor, *Progress in Computer Network Research*. Nova Publishers, 2005. (To be published)

- [36] M. Ferreira, F. Ruiz, M. Piattini. Integración de WorkFlows en los Entornos de Ingeniería del Software. Informatics Technology Management/Revista Gerencia Tecnológica Informática (GTI), Instituto Tecnológico Iberoamericano de Informática, Colombia. To be published
- [37] A.P. Flores, M. Polo. Dynamic assembly and integration on component-based systems. 4as Jornadas Iberoamericanas de Ingeniería del Software e Ingeniería del Conocimiento, JIISIC. 2004
- [38] A.P. Flores, J.C. Augusto, M. Polo, M. Varea. Towards Context-aware Testing for Semantic Interoperability on PvC Environments. Conference on Software Engineering Research and Practice. 2004
- [39] A.P. Flores, M. Polo, A. Cechich. Towards Context-aware Testing for Semantic Interoperability on PvC Environments. Net.ObjectDays Workshops on Testing of Component-Based Systems (TECOS2004) and Software Quality (SOQUA), pp. 17-29. 2004
- [40] I. García Rodríguez, M. Polo, M. Piattini. Una herramienta basada en metamodelos para la generación automática de aplicaciones web. Informatics Technology Management/Revista Gerencia Tecnológica Informática (GTI), Instituto Tecnológico Iberoamericano de Informática, Colombia. pp. 49-57. 2004
- [41] I. García, M. Polo, M. Piattini. An integrated environment for reengineering. JOurnal on Software Maintenance and Evolution: Research and Practice. To be published. 2005
- [42] I. García, M. Polo, M. Piattini. Metamodels and architecture of an automatic code generator. 2nd Nordic Workshop on the Unified Modeling Language, NWUML 2004
- [43] I. García, M. Polo, M. Piattini. Automated generation of component-based web applications from databases. Conference on Software Engineering Research and Practice. 2004
- [44] I. García, M. Polo, M. Piattini. Generación Automática de Aplicaciones basadas en Componentes a partir de Bases de Datos. VII Workshop Iberoamericano de Ingeniería de Requisitos y Desarrollo de Ambientes de Software IDEAS. 2004
- [45] I. García, M. Polo, M. Piattini. Estado del arte de la reingeniería y la ingeniería inversa: 2001-2003. 4as Jornadas Iberoamericanas de Ingeniería del Software e Ingeniería del Conocimiento, JIISIC. 2004
- [46] I. García, M. Polo, M. Piattini. Un entorno integrado para reingeniería. JISBD'05. To be published. 2005
- [47] F. García, F. Ruiz, M. Piattini, F. Canfora, C.A. Visaggio. FMESP: A Framework for the Modelling and Evaluation of Software Processes. QUantitative TEchniques for SoftWare Agile Processes, QUTE-SWAP. 2004
- [48] J. Garzás, M. Piattini. Improving the object-oriented design process. 11th European Conference EuroSPI 2004. LNCS(3281). 2004
- [49] O. Martín-Díaz, D. Benavides, J. Peña, M. Toro. Un tratamiento sensible a la calidad para la adquisición de servicios web. In VIII Jornadas de Ingeniería del Software y Bases de Datos JISBD'03, pp. 209-220, 2003
- [50] O. Martín-Díaz, A. Ruiz-Cortés, D. Benavides, A. Durán, M. Toro. A quality-aware approach to web services procurement. In 4th. VLDB Workshop on Technologies for E-services TES'03, LNCS(2819):42-53, 2003
- [51] O. Martín-Díaz, A. Ruiz-Cortés, R. Corchuelo, M. Toro. A framework for classifying and comparing web services procurement platforms. In 1st. Web Services Quality Workshop WISE-WQW'03, pages 156-165, 2003
- [52] O. Martín-Díaz, A. Ruiz-Cortés, A. Durán, D. Benavides, M. Toro. 1st International Conference on Service Oriented Computing ICSOC'03: Automating the Procurement of Web Services. LNCS(2910): 91-103. 2003
- [53] J. Peña, R. Corchuelo. 2nd. International Workshop on Practical Applications of Agents, Multi-Agent Systems IWPAAMS'03: Towards a Coordination Specification of Complex Multi-Agent System, pp. 54-65. 2003
- [54] J. Peña, R. Corchuelo, J. L. Arjona. ACM Symposium on Applied Computing SAC'03: A Top Down Approach for MAS Protocol Descriptions. pp. 45-50. 2003
- [55] J. Peña, R. Corchuelo, O. Martín-Díaz. VIII Jornadas de Ingeniería del Software y Bases de Datos JISBD'03: Towards a Methodology for Multi-Agent System Coordination, pp. 315-324. 2003
- [56] J. Peña, R. Levy, R. Corchuelo. Revista Iberoamericana de Inteligencia Artificial. Towards clarifying the importance of interactions in Agent-Oriented Software Engineering. No.25 (2005), pp. 19-28. ISSN: 1137-3601
- [57] J. Peña, R. Corchuelo, M. Toro. IX Jornadas de Ingeniería del Software y Bases de Datos JISBD'04. Representing Complex Multi-Agent Organisations in UML, pp. 159-170. 2004

- [58] J. Peña, R. Corchuelo, M. Toro. 3rd. Intl. Workshop on Practical Applications of Agents, Multiagent systems IWPAAMS'04. Modelling Interaction Organisation in Large Multi-Agent Organisations. pp. 61-72. 2004
- [59] J. Peña. Mejorando las Capacidades para Modelar Sistemas Multiagentes Complejos. ETSI Informática. 2004
- [60] J.A. Pérez, R. Corchuelo, M. Toro. An Order-based Algorithm for Implementing Multiparty Synchronization. Concurrency, Computation: Practice and Experience. 16: 1173-1206. 2004
- [61] M. Polo, M. Piattini. Test case mutation. Verification, validation and testing in software engineering. Idea Group Publishing. To be published. 2005
- [62] M. Polo. Identificación de riesgos en proyectos ágiles. Informe interno, Grupo Alarcos. 2005
- [63] M. Resinas, P. Fernández, R. Corchuelo. Hacia un framework para la negociación automática orientada a servicios. Actas del I Congreso Español de Informática (CEDI), 2005. (To be published)
- [64] J.C. Riquelme, M. Polo, J.S. Aguilar-Ruiz, M. Piattini, F.J. Ferrer-Troyano, F. Ruiz. A comparison of effort estimation methods for 4GL programs: experiences with Statistics and Data Mining. International Journal of Software Engineering and Knowledge Engineering. World Scientific Publishing Company. (To be published)
- [65] O.M. Rodríguez, A. Vizcaíno, A.I. Martínez, M. Piattini, J. Favela. How to Manage Knowledge in the Software Maintenance Process. 6th International Workshop on Advances in Learning Software Organizations (LSO 2004), LNCS(3096):78-87, 2004
- [66] O.M. Rodríguez, A.I. Martínez, J. Favela, A. Vizcaíno, M. Piattini. Understanding and Supporting Knowledge Flows in a Community of Software Developers. X International Workshop on Groupware (CRIWG'04), LNCS(3198): 52-66, 2004
- [67] O.M. Rodríguez, A. Vizcaíno, A.I. Martínez, M. Piattini, J. Favela. Using a Multi-agent Architecture to Manage Knowledge in the Software Maintenance Process. 8th International Conference on Knowledge-Based Intelligent Information and Engineering Systems (KES'04), LNCS(3213), 2004
- [68] O.M. Rodríguez, A. Vizcaíno, A.I. Martínez, M. Piattini, J. Favela. Applying Agents to Knowledge Management in Software Management. International Workshop on Agent-Mediated Knowledge Management, AMKN'04. 2004
- [69] E. Rolón, F. Ruiz, M. Piattini. El Factor Humano en el proceso de Mantenimiento. Informatics Technology Management/Revista Gerencia Tecnológica Informática (GTI), Instituto Tecnológico Iberoamericano de Informática, Colombia. To be published
- [70] F. Ruiz, A. Vizcaíno, M. Piattini, F. García. An Ontology for the Management of Software Maintenance Projects. International Journal of Software Engineering and Knowledge Engineering. World Scientific Publishing Company, 14(3): 323-349, 2004
- [71] F. Ruiz, F. García, M. Piattini. Software Process Measurement. UPGRADE. The European Journal for the Informatics Professional. Vol. 7. 2004
- [72] D. Ruiz. On fairness, multiparty interaction. E.T.S. de Ingeniería Informática. Dpto. de Lenguajes y Sistemas Informáticos. Universidad de Sevilla. 2003
- [73] A. Ruiz-Cortés, O. Martín-Díaz, A. Durán, M. Toro. Improving the automatic procurement of web services using constraint programming. International Journal of Cooperative Information Systems. To be published, 2005
- [74] V. Tasic. WSOL versus related work. Research report SCE-04-07 by the Department of Systems and Computer Engineering, Carleton University, 2004
- [75] P. Trinidad, D. Benavides, A. Ruiz-Cortés. Putting the basis for automatic reasoning on software product lines. Desarrollo y Mantenimiento Ágil de aplicaciones basadas en Servicio Web (ZOCO'04), 2004
- [76] A. Vizcaíno, J.P. Soto, M. Piattini. Supporting knowledge reuse during the software maintenance process through agents. Sixth International Conference on Enterprise Information Systems (ICEIS 2004), vol. 1, pp. 397-402. 2004
- [77] A. Vizcaíno, F. Ruiz, M. Piattini, F. García. Using REFSENO to Represent Knowledge in the Software Maintenance Process. 5th International Workshop on Theory and Applications of Knowledge Management, TAKMA'04. 2004

MODELADO DE PRESTACIONES DE SISTEMAS DE TIEMPO REAL MOPRESIT TIC2003-06293

Ramon Puigjaner*
Main Researcher

Carlos Juiz**
Researcher

Abstract

Designing real time systems needs to consider the non-functional requirements of temporal type, if possible since the first phases of the design. If not, the fulfilment of the requirements is just verified after the implementation. If they are not met, it is necessary to go back and reprogram and even redesign the system.

This project intends to develop a tool allowing the prediction, since the first phases of the design, of what will be the temporal behaviour of the system and, in this way, reduce the risk of erroneous decisions taken during the design. Obviously, for doing this prediction, it is necessary to complement the design information with other concerning the temporal behaviour of the system, which will be included in the system as annotations.

In order to experiment the best way to annotate designs, two notably different design approaches will be used: MASCOT and UML. The first one, originated for the design of hard real-time systems, is a method oriented to the data flow and based on activities interchanging information through intercommunication data areas. The second one, result of merging a set of diagrams proposed by several authors, allows the description of the different dimensions needed for the design of real-time systems of great size.

From the annotated designs in these two methods, we intend to that the tool derives queuing network models able to be processed by simulation methods and, if possible, by analytical methods. Till here this is the original abstract of the project. However, as it will be explained in the body of this presentation, several goals have been modified.

Keywords: Soft real-time systems. Performance modelling. Software performance engineering.

1. Project goals

1.1 Original goals

1. To build a tool allowing the automatic derivation of performance models based on queuing networks in order to know the performance of soft real-time systems before their implementation and their test verification of their timing requirements. Building this tool will be done first with a high level proprietary software, and then, with free software in order to ease its diffusion.
2. Building this tool should be preceded by the study of how to incorporate to the design languages, like UML, the timing aspects related with the system performance. To do this, it

* Email: putxi@uib.es

** Email: cjuiz@uib.es

will be necessary to see how to annotate UML and how can be extended the different UML diagrams in order to correctly construct the performance models and then to execute them.

3. Another approach to build the tool is derived from MASCOT as design method for real-time systems. This requires to complete the taxonomy of data interchange protocols to collect all the aspects related to the performance modelling and then to set up the tool allowing deriving queuing network models.

1.2 Modified goals

Due to several reasons (several members of the research team left the university, discovering new aspects interesting to be researched, etc.), the current goals of the project are:

1. The goal 3 has been worked out but instead of creating a tool for deriving performance models, the annotations of a MASCOT design have been standardised and a simulation process has been set up directly using the design description together with the annotations to obtain the desired performance estimations.
2. The goal 1 has been modified in order to develop PMIF, Performance Model Interchange Format, a mechanism whereby system model information may be transferred among performance modelling tools. Also it has been demonstrated how Web services can be used to facilitate the use of modelling tools that can interface with the PMIF. Indirectly, it also includes part of the goal 1 because one type of the performance models considered is the one based on UML mainly as input to obtain some type of performance model able to be computed to obtain performance results.
3. A new topic that has been included in the project goals is the performance assessment of ambient intelligence systems through ontology engineering. Thus, firstly it has been necessary to find appropriate description methods for distributed intelligent applications. Derived from the system characterization, typical software performance engineering techniques are based on the augmented description of the model regarding performance annotations. However, these annotations are only related with the syntactical view of the software architecture. In the next generation of performance assessment tools for ambient intelligent systems, the description of the system has to be able of reasoning and acquiring knowledge about performance. Having an appropriate architectural description including performance aspects, any possible design options for intelligent distributed applications can be evaluated according to their performance impact.

2. Main results related to each goal

2.1 MASCOT performance models (MASCOTime)

Soft real-time systems differ from traditional software systems in that they have a dual notion of correctness. That is, in addition to producing the correct output, soft real-time systems must produce it in a timely manner.

One of the most usual approaches to software engineering is based on the concept of a system as a set of interacting components. In order to characterize the interaction between the components is also necessary to express a protocol representing the timing constraints on the interaction and to its support by an explicit interconnection in the future implemented system.

MASCOT (Modular Approach to Software Construction Operation and Test) is a design and implementation method for real-time software development and brings together a co-ordinated set of techniques for dealing with the design, construction (system building), operation (run-time execution) and testing software, as its acronym wants to depict [1]. At the heart of MASCOT there

is a particular form of software structure supported by complementary diagrammatic and textual notations. These notations give visibility to the design as it emerges during development, and implementation takes place via special construction tools that operate directly on the design data. MADGE (Mascot Design Generator) is one of these tools [2]. Design visibility greatly eases the task of development management, and the constructional approach ensures conformity of implementation to design, and allows traceability of performance back to design. The functional model allows a system to be described in terms of independently functional units interacting through defined transfers of information. Here is introduced the notion of protocol component which is used to annotate the interaction between processes with a MASCOT representation [3]. Unfortunately, MASCOT does not provide any kind of performance constraints annotation on its functional components that could estimate, running a performance evaluation tool, whether the system being designed will meet or not the performance requirements.

In this research, we prototyped MASCOTime as the main contribution to the integration of performance engineering practices into a traditional system engineering methodology through performance constraint annotations. MASCOTime interferes minimally with the functional description on original MDL (MASCOT Description Language). In this way, system/software engineers do not perceive any change in the functional description of the future system.

However, once the system design method has been augmented with performance annotations, the information provided for the explicit system design is not evaluated itself. It is necessary either to build a performance tool to directly compute the performance of the model that MASCOTime generates. Therefore, we also build a prototype of discrete-event simulator for MASCOTime designs. The implementation of the simulator has been developed using objects that represent directly the MASCOTime components. Thus, the simulated components that build a system and the objects that the simulator manages are isomorphic. Due to the restrictions that would be imposed to the functional (and performance) model to apply analytical or numerical algorithms, we decide to avoid transforming the MASCOTime model to a traditional formal model for performance evaluation. The discrete-event simulator uses a direct translation between the MASCOTime components and their isomorphic Java objects. With this newly started work our goal is to automatize the analysis, that is, to build some expert knowledge into the simulator and make it analyze the system, diagnose it and give concrete advice to the system designer.

2.2 Performance Model Interchange Format (PMIF)

The Software Performance Engineering (SPE) process uses multiple performance assessment tools depending on the state of the software and the amount of performance data available. Web Services are distributed software components that allow the communication among applications or application components in a standard way through common protocols that are independent of the programming language, platform, presentation format, and operating system.

If a software performance modelling tool, like *SPE-ED* [4], wants to access another performance model Web Service, the tool first exports the model to a predefined model interchange format and then makes an HTTP call (application-to-application) to this Web Service. A tool owner who wants to provide a performance model Web Service needs to implement the Web Service, the import mechanism from the interchange format to a model in the tool's format, solve the model and send back to the client the obtained results. A user of several tools that support a common interchange format can create a model in one tool, and later move the model to other tools for further work without the need to laboriously translate from one tool's model representation to the other and the need to validate the resulting specification. Web Services provide a mechanism to easily invoke other tools automatically rather than having to manually execute each one.

We could also use performance model Web Services for SPE. Standard model interchange formats are the foundation that enables a Performance model Web Service. A common set of XML based interchange formats lets one use a variety of different tools as long as they support the format. To use an interchange format, each tool must either provide an explicit import and export command, or provide an interface to/from a file. With a file interface, an Extensible Stylesheet Language translation (XSLT) [5] can convert between the interchange format and the file.

There are two performance model interchange formats: the Performance Model Interchange Format (PMIF) is for the exchange of Queueing Network Models (QNM); and the Software Performance Model Interchange Format (S-PMIF) is for the exchange of software performance models among (UML-based) software design tools and software performance engineering tools. The PMIF meta-model is a model of the information that goes into constructing a QNM model. The PMIF was subsequently enhanced, implemented in XML, and named PMIF 2.0.

Thus, the PMIF 2.0 is a common representation for system performance model data that can be used to move models among system performance modeling tools that use a queueing network model paradigm. Earlier work also defined an SPE meta-model that formally defines the information required to perform an SPE study [6]. This model is known as the SPE meta-model because it is a model of the information that goes into constructing an SPE model. The S-PMIF based on the SPE meta-model is a common representation that can be used to exchange information between UML-based (Unified Modeling Language, [7]) software design tools and software performance engineering tools.

Our ongoing project intends to automate SPE performance evaluation. We have developed a Web Service prototype that makes implementation of a Performance model Web Service viable. Results would then be exchanged in the reverse direction and ultimately the software specialist would be able to view suggestions for performance improvements and automatically update the UML to reflect selected changes.

2.3 Performance Modelling of Ambient Intelligent Systems (PA-Ai)

During the last years UML has been widely used to specify, construct and document the functionality of software systems. In order to reduce the gap between functional models and performance formalisms, a software and performance community has emerged to provide (automatically) accessible techniques and tools to include performance annotations for building performance models of software models, constituting a new research topic known as Software and Performance Engineering (SPE). However, the growing availability of wireless networks and the expansion of powerful mobile devices define new challenges for these software distributed systems. Moreover, ambient intelligent applications designed for mobile computing are expected to run in a highly heterogeneous, dynamic and self-controlled environment. In that sense, less attention has been paid to these last phenomena in the SPE community because the traditional software architectures for distributed applications are difficult to translate to ad-hoc communication environments [8].

The conceptual framework of ambient intelligent systems includes an ontological foundation that structures the adaptability of the system to the users. An ontology is a formal, explicit specification of a shared conceptualization [9]. Conceptualization is referring to an abstract model of some piece of the real world, where the concepts that build the ontology are clearly identified. Therefore, the conceptualization process means the utilization of a formal language for representing the ambient. Clearly, this is a new challenge for the performance engineering area.

The importance of using an ontology as a conceptualization tool resides on the potential power of reasoning over the represented data. The ability of reasoning will allow taking decisions about performance and other non-functional features of systems. Different languages have been defined to represent particular knowledge, but none was totally compatible with the semantic World Wide Web as the Web Ontology Language (OWL). OWL is built on RDF and RDF schema [10] with a set of vocabularies for describing properties and classes. Since OWL includes semantics it is possible to apply logical rules to performance variables and even to infer new performance knowledge.

This research is fully complementary to previous approaches. For example, UML annotated diagrams that provide key information required for performance analysis describes, in different ways, behaviour of the system and consumption of resources, together. A huge number of approaches have been proposed to derive performance models from software architecture specifications: from annotated UML diagrams performance models are generated in the corresponding formalism (QN, SPN, SPA, etc.) and then they are off-line evaluated through analytical, numerical or discrete-event simulation techniques (see section 2.2). Very interesting examples of this R+D process can be found in the literature even the UML augmentation approach is not unique in the SPE, other software design methodologies has been extended to be simulated (see section 2.1).

Even the off-line performance evaluation relies on the syntactical annotated constraint values; it seems to be interesting to get information on-line about the relations in the context and to reason about them (monitoring and reasoning). The expressiveness of the ontology language rarely substitutes the syntactic richness of other approaches for an off-line performance evaluation of the system. However, the semantics and the standardization of OWL persuades to further research on this direction.

3. Results' indicators

Results appearing in some of these sections are due to the task of Catalina Lladó, in addition to those of the two authors of this report. The rest of members of the research group are mainly involved in the project TIC2001-1374-C03-01, reported below.

3.1 Graduate and Undergraduate Students

3.1.1 Katja Gilly de la Sierra

After obtaining her Research sufficiency, she is continuing her task, started in TIC2001-1374-C03-01, towards her PhD thesis. She is *Profesor Titular de Escuela Universitaria (i)* at the Universidad Miguel Hernández of Elx.

3.1.2 Salvador Alcaraz

After obtaining his Research sufficiency, he is continuing his task, started in TIC2001-1374-C03-01, to find the niche of his PhD thesis. He is *Profesor Titular de Escuela Universitaria* at the Universidad Miguel Hernández of Elx.

3.1.3 Carlos Guerrero

After finishing his doctoral courses, he is starting his task to find the niche of his PhD thesis. He was the last two year contracted as *Ayudante*, and from next academic year as *Profesor Colaborador* at UIB.

3.1.4 Jerònia Rosselló

She has been working, first as undergraduate student and then as graduate student in the domain of PMIF, from the beginning of the project until February 2005. She has obtained her degree of Informatics Engineer during her stay in the research group and has participated in several publications.

3.1.5 Pere Pau Sancho de la Jordana

He has been working, first as undergraduate student and after as graduate student in the domain of MASCOTime, since the beginning of the project. Initially he was technician contracted by the project and then with a similar status but partially funded also by the UIB. He has obtained his degree of Informatics Engineer during his stay in the research group; currently he is working in his PhD Thesis and has participated in several publications.

3.1.6 Isaac Lera

He is working as undergraduate student in the domain of performance of Ambient Intelligent systems, since July 1st. He is contracted as technician by the project. He has participated in several publications.

3.2 Publications

- 1 Haring, G.; C. Juiz, C. Kurz, R. Puigjaner, J. Zottl, *Framework for the Performance Assessment of Architectural Options on Intelligent Distributed Applications*. PERMIS'04. Proceedings. Multimedia publication.
- 2 Molero, X., C. Juiz, M.J. Rodeño, *Evaluación y Modelado del Rendimiento de los Sistemas Informáticos*. Prentice-Hall. Book ISBN 84-205-4093-5.
- 3 Sancho, P.P., C. Juiz, C., R. Puigjaner, *Evaluating the Performance of Architectures in MASCOTime*. SERP'04. Proceedings CSREA Press Book Chapter ISBN 1-932415-28-9.
- 4 Sancho, P.P., C. Juiz, C., R. Puigjaner, *Integrating System Performance Engineering into MASCOTime Methodology through Discrete-event Simulation*. In *Applying Formal Methods: Testing, Performance, and M/E-Commerce* (N. Núñez, Z. Maamar, F. L. Pelayo, K. Pouttschi, F. Rubio, eds.). Proceedings of EPEW 2004. Book Chapter ISBN 3-540-23169-2. Journal article: LNCS 3236, ISSN 0302-9743. pp. 278-292.
- 5 Smith, C. U. and C. M. Lladó, *Performance Model Interchange Format (PMIF 2.0): XML Definition and Implementation*, Proc. 1st Int. Conf. on Quantitative Evaluation of Systems (QUEST), Enschede, NL, IEEE Computer Society, 2004. Book Chapter ISBN: 0-7695-2185-1
- 6 Juiz C., J. Zottl, G. Haring, R. Puigjaner, *Performance Assessment of Architectural Options on Intelligent Distributed Systems*. IWSSA'05. Proceedings. Book Chapter. CSREA Press, vol I, 2005. ISBN 1-932415-49-1. pp. 45-51.
- 7 Kurz, C., C. Guerrero, R. Puigjaner, G. Haring, *Extending TPC-W to Allow for Fine Grained Workload Specification*. Proceedings of the 5th International Workshop on Software and Performance, WOSP 2005. Book Chapter. ACM Press, 2005. ISBN 1-59593-087-6. pp. 167-174.
- 8 Lera, I., C. Juiz, R. Puigjaner, *Web Operational Analysis Through Performance-Related Ontologies in OWL for Intelligent Applications*. Web Engineering: 5th International Conference, ICWE 2005 (D. Lowe, M. Gaedke, editors). LNCS 3759. Springer S+B M 2005. Book Chapter. ISBN 3-540-27996-2. Journal article: ISSN 0302-9743

- 9 Lera, I., C. Juiz, C. Kurz, R. Puigjaner, G. Haring, J. Zottl, *Performance Assessment on Ambient Intelligent Applications through Ontologies*. Proceedings of the 5th International Workshop on software and Performance, WOSP 2005. Book Chapter. ACM Press, 2005. ISBN 1-59593-087-6. pp. 195-204.
- 10 Rosselló, J., C. M. Lladó, R. Puigjaner, C. Smith, *A Web Service for solving Queuing Network Models using PMIF*. Proceedings of the 5th International Workshop on software and Performance, WOSP 2005. Book Chapter. ACM Press, 2005. ISBN 1-59593-087-6. pp. 187-192.
- 11 Rosselló, J., C. M. Lladó, R. Puigjaner, C. Smith, *A Lexical and Syntactical Analyzer for the Exporting of QNAP to the Performance Model Interchange Format (PMIF)*. 31 Conferencia Latinoamericana de Informática, CLEI 2005. Accepted paper.
- 12 Sancho, P.P., C. Juiz, R. Puigjaner, *Evaluating the Performance of Architectures in MASCOT*. Science of Computer Programming, Volume: 57-1, Journal article ISSN 0167-6423. pp. 45-60.
- 13 Sancho, P.P., C. Juiz, R. Puigjaner, *Giving Feedback on MASCOTime Simulation Results*. SERP'05 Proceedings. Book Chapter. CSREA Press, vol. I, 2005. ISBN 1-932415-49-1. pp. 126-134.
- 14 Sancho, P.P., C. Juiz, R. Puigjaner, *Automatic Performance Evaluation and Feedback for MASCOT Designs*. Proceedings of the 5th International Workshop on software and Performance, WOSP 2005. Book Chapter. ACM Press, 2005. ISBN 1-59593-087-6. pp. 193-194.
- 15 Smith, C. U., C. M. Lladó, V. Cortellessa, A. Di Marco and L. G. Williams, *From UML Models to Software Performance Results: An SPE Process Based on XML Interchange Formats*, Workshop on Software and Performance, WOSP 2005. Book Chapter. ACM Press, 2005. ISBN: 1-59593-087-6

3.3 Joint Actions and Projects

3.3.1 Integrated Action with the University of Vienna (Austria)

This project aims to develop a methodology framework for performance engineering of applications with ambient intelligence, showing its applicability in a few test cases. As a result of the encouraging initial results obtained in this integrated action, the goals of the project have been redefined.

3.3.2 *Análisis de las prestaciones en los entornos web*

Coordinated Project (TIC2001-1374-C03-01) with the Universities *Politécnica de València* and *Oviedo*, initially planned to finished in December 2004 but prorogued until December 2005.

3.4 National and International Collaborations

Members of the group maintain regular collaboration actions with the following universities: Miguel Hernández, Politécnica de València, dell'Aquila (IT), Viena (AT), Paul Sabatier de Toulouse (FR), Imperial College (UK), Carleton (CA), etc.

3.5 Relevant Participation on Meetings

- 1 Galmés, S., C. Juiz, Lladó, C., Puigjaner, R., XXX Conferencia Latinoamericana de Informática. PC Members.
- 2 Galmés, S., C. Juiz, International Symposium on Performance Evaluation of Computer and Telecommunication Systems, SPECTS 2004. PC Members.
- 3 Juiz C., Workshop on Software and Performance, WOSP 2005. Program Chair, Steering Committee Member and Session Chair.
- 4 Juiz, C., Workshop on System/Software Architectures IWSSA'05. PC Member.
- 5 Juiz, C., Conference on Software Engineering Research and Practice, SERP'05. PC Member.

- 6 Juiz, C., R. Puigjaner, International Symposium on Computer and Information Sciences, ISCIS 2004. PC Members.
- 7 Juiz, C., 8th WSEAS International Conference on COMPUTERS. Guest Editor and Session Chair.
- 8 Juiz, C., 12th Conference of the IEEE / ACM MASCOTS 2004. Session Chair.
- 9 Juiz, C., R. Puigjaner, Symposium on Performance Evaluation of Computer and Telecommunication Systems, SPECTS 2005. PC Members.
- 10 Lladó, C. M., R. Puigjaner, 31 Conferencia Latinoamericana de Informática, CLEI 2005. PC Members.
- 11 Puigjaner, R., 12th IEEE / ACM International Symposium on Performance Evaluation of Computer and Telecommunication Systems, MASCOTS 2004. PC Member.
- 12 Puigjaner, R., Network Control and Engineering for QoS, Security and Mobility, Net-Con 2004. Conference Chair.
- 13 Puigjaner, R., XXX Conferencia Latinoamericana de Informática CLEI 2004. Session Chair.
- 14 Puigjaner, R., IFIP World Computer Congress, WCC 2004. PC Member.
- 15 Puigjaner, R., 1st International Conference on Quantitative Evaluation of Systems, QEST 2004. PC Member.
- 16 Puigjaner, R., 3rd International Conference on Networking, NETWORKING'2004. PC Member and Session Chair.
- 17 Puigjaner, R., XXX Conferencia Latinoamericana de Informática, CLEI 2004. Session Chair.
- 18 Puigjaner, R., EPEW'2004. PC Member.
- 19 Puigjaner, R., Workshop on Software and Performance, WOSP 2005. General Chair, Steering Committee Member and Session chair.
- 20 Puigjaner, R., 4th International Conference on Networking, NETWORKING 2005. PC Chair and Session Chair.
- 21 Puigjaner, R., EPEW'2005. PC Member.

4. References

- [1] Simpson, H.R.: The Mascot method, Software Engineering Journal, pp. 103-120, May 1986.
- [2] Harding, M.D.: MASCOT DESIGN GENERATOR (MADGE) v.6.0. Release Notification, n. 5096, issue 11, November 1998.
- [3] Simpson, H.R.: Protocols for Process Interaction. 1, 2, 3. Matra BAe Dynamics, 2000.
- [4] Smith, C.U., and Williams L.G., Performance Engineering Evaluation of Object Oriented Systems with SPE•ED, in Computer Performance Evaluation: Modelling Techniques and Tools, Springer-Verlag, 1997.
- [5] Object Management Group (OMG): XML Metadata Interchange (XMI) Specification, version 1.2, <http://www.omg.org/cgi-bin/>, 2001.
- [6] Smith, C.U., and Williams L.G., Performance Solutions: A Practical Guide to Creating Responsive, Scalable Software, Addison-Wesley, 2002.
- [7] Object Management Group (OMG): Unified Modeling Language Specification, version 1.3, <http://www.omg.org/uml>.
- [8] Beer, W., Christian, V., Ferscha, A. and Mehrmann L., Modeling Context-Aware Behavior by Interpreted ECA rules. In Proceedings of Euro-Par 2003, H. Kosch, L. Böszörményi and H. Hellwagner (Eds.), LNCS 2790, pp. 1064-1073, 2003.
- [9] Berners-Lee, T., Hendler, J. and Lassila, O. "The Semantic Web", Scientific American, May 2001.
- [10] OWL Web Ontology Language Overview <http://www.w3.org/TR/owl-features/>.

MD2: Método para el desarrollo de materiales didácticos extendidos y su automatización TIC2003-03654

Paloma Díaz Pérez *

Laboratorio DEI. Departamento de Informática. Universidad Carlos III de Madrid

Abstract

The MD2 project aims at providing tools to develop extended learning materials. This goal is being addressed by means of three activities: (1) to define a method for the collaborative generation of learning contents; (2) to extend current learning object standards to incorporate concepts such as ontologies and adaptation to the context, and (3) to elaborate an evaluation framework for testing the usability and utility of educational applications. A specific platform will be developed to test practically all these goals. This document describes the project status after eighteen months of work as well as the main results obtained till the moment.

1 Project goals and resources

The impressive rise in the use of CBTs is chiefly due to the need for supporting lifelong learning at all educational levels as well as to the use of Internet as a distribution platform that allows didactic resources to be accessed anywhere and anytime. Even though a great of effort is being devoted to proposing standard mechanisms for creating reusable learning objects and courses in an easy and useful way by different international groups and organizations (including LTSC, ADL, ...), there are still some issues that need to be properly addressed as the ones that shape the objectives of the MD2 project:

- O1 To propose **a method for the collaborative generation of learning contents**, offering a framework for cooperative knowledge production that will take into account different interaction styles with a view to improve efficiency and reduce conflicts and inconsistencies
- O2 To **extend current learning object standards** to incorporate advanced features such as the metadata cohesion by using accepted and shared concepts (ontologies) and supporting hypermedia concepts such as the adaptation to the context.

*Email: pdp@inf.uc3m.es

	UC3M	External Org.
Doctor	4	2
Phd. Students	2	2
Undergraduate Students	2	0

Table 1: MD2 team

O3 To **elaborate an evaluation framework for a priori testing** of usability and utility of educational applications that will include a method and a number of criteria, parameters and metrics concerning the educational usefulness and the quality of the interaction mechanisms.

These three theoretical endeavours will be practically tested into a platform that will be developed with this purpose.

1.1 Project Resources

To achieve the goals stated above the project was assigned 6 EDPs. Team composition is shown in table 1. Most members have a technical background though experts in pedagogical issues are also collaborating to provide the necessary multidisciplinary view. These pedagogical experts are Dr. Cristina Laorden Gutiérrez from *Universidad de Alcalá* and Dr. *Valentín Martínez-Otero* from Centro Universitario Villanueva, both of which are cooperating to define criteria and parameters making up the evaluation framework (O3).

1.2 Project Planning

At this moment there are no remarkable deviations from the initial planning shown in figure 1. Activities planned for the next eighteen months include: to improve the method for col-

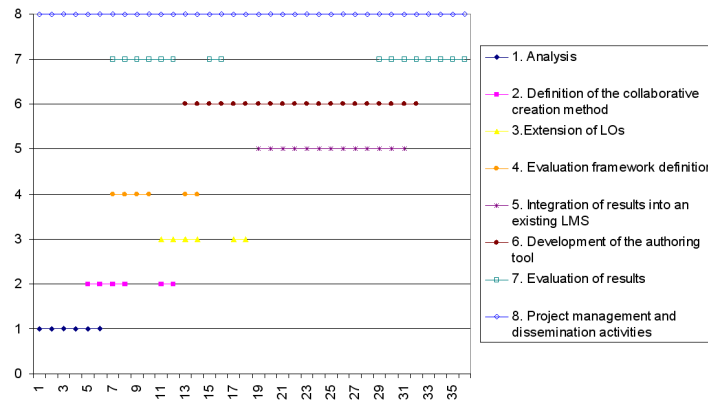


Figure 1: MD2 initial planning.

laborative authoring and the authoring tool based on evaluation results; to define a model of

dynamic adaptation for the units of learning (UoL); to define the criteria and metrics for aprioristic evaluation of learning material; to include meta-data related with aprioristic evaluation; and to integrate the authoring tool into an existing LMS.

2 Level of achievement

The study of the state of the art gave rise to the selection of the technological platform and standards with which we have worked out. Some results of this study have been:

- The selection of standard specifications like Learning Object Metadata (LOM), as well as structured descriptions like IMS Learning Design (IMS-LD)
- The incorporation of web services as a new technology that can drive a more efficient integration with learning systems and other related systems, in a more accessible fashion for end users, including companies who expressed their interest in the project results.
- The second goal of the project has been revised to a more complete but feasible one: the dynamic adaptation of units of learning to facilitate personalization of the environment to the students' profile, maintaining as well the semantic annotation as the basis for this goal.

The analysis activity is considered as complete, though new works may emerge in the future, whose relevance should be taken into account within the project goals.

2.1 A method for the collaborative generation of learning contents (O1)

A first version of the method has been devised, including the semantic annotation process as the basis to the generation and re-using of learning objects. This goal is not considered as finished, since the definition is incremental, to be continuously supported and evaluated by means of the corresponding software tool whose interface is shown in figure 2. In particular, to assist learning object authors in the collaborative creation process, an integrated development environment has been built, in which two major modules have already been developed and tested (see section 3.3). The results of using the method through that software tool will be used to improve and refine it iteratively.

2.2 Extension of the LO (O2)

Two different fronts have been tackled with respect to this goal:

- On the one hand, ontologies have been used to provide semantics to learning objects. Ontology annotations constitute the most basic mechanism of the authoring process. It is a process based upon the reuse of existing material and their semantic extension.
- On the other hand, there is a research effort on the adaptation of IMS-LD units of learning to different contexts, with the purpose of facilitating the reuse of didactic material in various situations, as well as to achieve the fitting of such material to the users' needs, so providing a dynamic and transparent adaptation to the user.

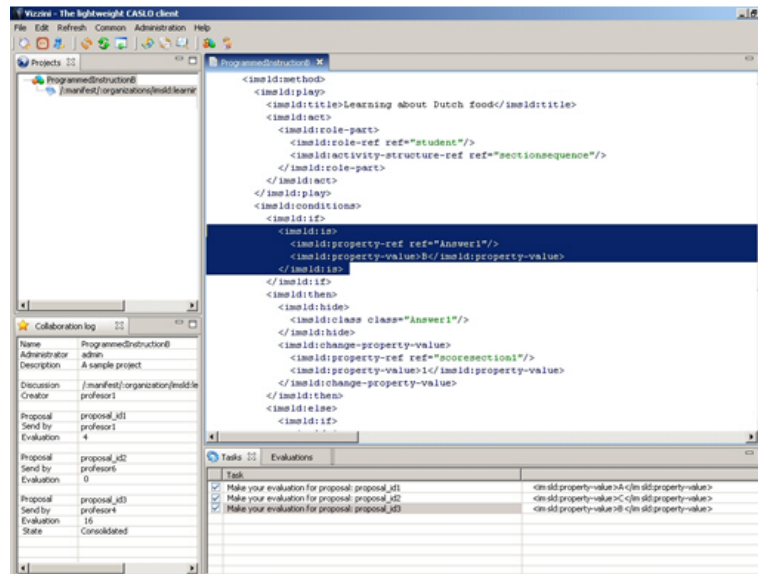


Figure 2: CASLO interface

In principle, the subject of extension concerned only hypermedia links. Nevertheless, since it has been extended to the adaptation of the educational process, this goal is not complete yet. The inclusion of ontologies has been carried out, but the adaptation of units of learning is still under ongoing. The main extension to learning technology specifications that has been provided is the way of integration of domain ontologies as learning object annotations [14, 22]. Another extension has been proposed, concerning the integration of decentralized learning service descriptions into a learning design specification [11]. Such extension will be subject of testing in the immediate stages of the project.

2.3 Apriori evaluation framework (O3)

A first analysis on the learning object evaluation model was elaborated, where the results were published in [1]. After some work with both software engineers and pedagogue experts (Dr. Cristina Laorden and Antonio Sarasa), such model and criteria were further elaborated, structuring the evaluation criteria into four dimensions: (1) *Technical criteria*, that assess the LO as a software object (interoperability, reusability, flexibility, etc.); (2) *Topic-specific criteria* that assess the LO as a piece of knowledge concerning a specific topic (eg. coherence, completeness, durability...); (3) *Pedagogical criteria* that assess the LO as a pedagogical resource (eg. learning goals, method, etc.); and (4) *Use criteria* that assess the LO as an interactive item (eg. presentation features, interaction, etc.).

Some preliminary works have been also carried concerning the formalization of elements from these four dimensions. In particular we are trying to translate criteria into elements of an ontology to describe explicitly the quality evaluation model of LO. An OWL-based ontology has been defined using tests based on the IMS QTI specification.

	Int. Journals	Conferences	Invited Talks	Int. Awards	Thesis
2004	0	5 [22, 4, 8, 2, 1]	3 [7, 12, 15]	1	
2005	4 [5, 14, 6, 16]	7 [17, 11, 20, 10, 19, 9, 21]	3 [18, 3, 13]	1	4
Total	4	12	6	2	4

Table 2: MD2 mid-term results

3 MD2 mid-term results

This section summarises the main results of the MD2 project including a description of personnel in training, publications and thesis and, finally, cooperation with other groups.

3.1 Personnel in training

As indicated in table 1, there are four PhD students preparing their thesis in issues related to the project goals. Such students are: Telmo Zarraonandía Ayo (UC3M), Carmen Luisa Padrón Nápoles (UC3M), Jorge Torres (*Instituto Tecnológico de Monterrey*, Mexico) and Antonio Sarasa (*Universidad Complutense de Madrid*). Some undergraduate students are also preparing their minor projects including:

- Jesús Lanchas Sampablo is finishing his project, which consists in the software design and implementation of a web services-based collaboration server and a prototype authoring tool.
- Carlos Salguero is extending the authoring tool to provide ontology navigation on the subjects of collaborative annotation.

3.2 Publications

Table 2 summarises the main publications derived from this project in this 18 months. The members of the project have received two international awards directly related with their work in MD2:

- The subproject "CASLO: Collaborative Annotation Service for Learning Objects" was awarded as the IEEE TCLT Small Scale Research Project 2004 in recognition of its contribution in the area of learning technologies.
- Dr. Dodero was awarded with the IEEE TCLT Young Researcher Award 2005 in recognition of his work in the area of Advanced Learning Technologies.

In addition to the publications, the members of the team are supervising four thesis under development:

- Telmo Zarraonandía Ayo. "Modelo de adaptación de diseños de aprendizaje en tiempo de ejecución". Supervisor: Camino Fernández.
- Jorge Torres. "Un modelo de soporte a la composición y ejecución distribuida de objetos y servicios de aprendizaje". Supervisor: Juan Manuel Dodero.

- Carmen L. Padrón Nápoles. "Procedimiento generativo para el diseño de objetos de aprendizaje". Supervisor: Paloma Díaz.
- Antonio Sarasa. "Un modelo de calidad para la composición de objetos de aprendizaje". Supervisors: Juan Manuel Dodero and Alfredo Fernández-Valmayor (Universidad Complutense de Madrid).

3.3 Technology transfer

The first version of the main tools supporting the collaborative authoring have been developed and tested [6]:

- CASLO is a web services-based collaboration server used to coordinate and negotiate the collaborative development of learning objects [9]. It is the collaboration provider of the integrated learning object development environment.
- *Vizzini* is a thin client graphical authoring tool, which is used to provide and discuss annotations on a learning object [13]. The provision of collaborative annotations is carried over CASLO server.

The authoring tool provides raw access to all functions of the collaboration server on any kind of XML manifest-described learning object, as all learning technology standards provide. Nevertheless, editing facilities are being extended with specialized views for some learning object specifications, particularly for question and test items (QTI), and learning design (LD) specifications.

3.4 Cooperation with other groups

The MD2 team has established and strengthen several liaisons with different research groups and organizations as a consequence of its work in this project.

The MD2 team has contacted the members of the UNFOLD european project consortium made up of Universidad Pompeu Fabra, The Open University of the Netherlands (OUNL), and European Continuing Education Network. Cooperation with this consortium includes the participation in several meetings and events as reported above. Moreover, the potential integration of MD2 technological results into the RELOAD tool is being analysed with the members of UNFOLD project. Another line of work is the extension of learning design specifications to be able to incorporate decentralized learning services. For this aim, discussions have been taken with two main players in the learning design arena:

- Prof. Rob Koper, director of learning technologies R&D at the OUNL, who is responsible for the development of Educational Modelling Language (EML), which currently takes the form of IMS Learning Design Specification.
- Dr. Ernie Ghihlione, LAMS project manager at McQuarie e-learning centre of excellence (MELCOE), McQuarie University (Sydney, Australia).

Researchers Nora Helena Villa and Dora Inés Chaverra from *Facultad de Comunicaciones* and *Facultad de Educación* respectively of Antioquia University (Colombia) were cooperating at

the laboratory from october 2004 to december 2004.

The project team also cooperates with the IEEE Technical Committee on Learning Technologies. In particular, Paloma Díaz is vice-chair of the committee and other team members take part in different events including editorial boards and program committees.

A closer cooperation has been established with the *Ingeniería de Sistemas Hipermedia* group at *Universidad Complutense* led by Prof. Dr. Alfredo Fernández-Valmayor to work in the area of usability criteria whose main result is the co-supervision of a thesis. The cooperation with this group and with Dr. Félix Buendía from Universidad Politécnica de Valencia has given place to a research project related with MD2 and prepared by both groups.

Other minor collaborations, mostly as expressions of interest about the results of MD2 project, were provided by Francesco Orciuoli, from the Centro di Ricerca in Matematica Pura ed Applicata, University of Salerno (Italy); and Cesar Moura, from the Université de Lille (France).

References

- [1] Sarasa A. and Dodero J.M. Towards a model of quality for learning objects. In *Proceedings of 4th IEEE Int. Conf. on Advanced Learning Technologies ICALT*, pages 822–824. IEEE Computer Science, 2004.
- [2] Sarasa A., Dodero J.M., Díaz P., and Padrón C.L. Integración de ontologías de dominio en la anotación de objetos de aprendizaje. In *Actas del 6º Simposio Internacional de Informática Educativa SIIE'04. Cáceres, Noviembre*, pages 34–39, 2004.
- [3] Padrón C.L. Learning web services composition and learner communities support for the deployment of complex learning processes., September 2005. Learning Designers CoP at UNFOLD Barcelona Communities of Practice (CoP) meetings.
- [4] Padrón C.L., Torres J., Dodero J.M., Aedo I., and Díaz P. Learning web services composition and learner communities support for the deployment of complex learning processes. In *Proceedings of 4th IEEE Int. Conf. on Advanced Learning Technologies ICALT*, pages 390–393. IEEE Computer Science, 2004.
- [5] Padrón C.L., Dodero J.M., Aedo I., and Díaz P. The development process of learning material. *Computer Science and Information Systems*, 2005. Invited paper.
- [6] Padrón C.L., Dodero J.M., and Lanchas J. CASLO: Collaborative annotation service for learning objects. *IEEE Learning Technology Technical Committee Newsletter*, 7(2):2–7, April 2005.
- [7] Aedo I. Tecnologías de la información para el desarrollo de materiales didácticos., Julio 2004. 7º Congreso Colombiano de Informática Educativa. Bogotá.
- [8] Aedo I. and Díaz P. Tecnologías de la información para el desarrollo de materiales didácticos. In *7º Congreso Colombiano de Informática Educativa. Bogotá*, Julio 2004.
- [9] Lanchas J., Dodero J.M., and Padrón C.L. Servicios web para la anotación colaborativa de objetos de aprendizaje. In *Actas de las Jornadas Técnico-Científicas en Servicios Web. Granada, Septiembre*, 2005.

- [10] Torres J., Dodero J.M., Aedo I., and Díaz P. A characterization of composition and execution language for complex learning processes. In *Proceedings of IASTED Int. Conference on Web-Based Education*, pages 255–260, 2005.
- [11] Torres J., Dodero J.M., Aedo I., and Zarraonandía T. An architectural framework for composition and execution of complex learning processes. In *Proceedings of Int. Conference on Advanced Learning Technologies ICALT*, Kaohsiung, Taiwan, July, 2005.
- [12] Dodero J.M. De la integración de las TIC en el desarrollo de material didáctico, Octubre 2004. I Simposio Pluridisciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos Reutilizables, Universidad de Alcalá (Guadalajara).
- [13] Dodero J.M. The CASLO collaborative authoring system, June 2005. "Implementing the solutions session" at UNFOLD "Communities of practice meeting, Braga, Portugal".
- [14] Dodero J.M., Aedo I., Díaz P., and A. Sarasa. Integration of ontologies into the collaborative authoring of learning objects. *Journal of Universal Computer Science*, 2005. accepted for publication.
- [15] Díaz P. Las TIC en el proceso de enseñanza-aprendizaje, Mayo 2004. I Jornada Campus Virtual UCM. Universidad Complutense de Madrid.
- [16] Díaz P., Aedo I., and Montero S. Applying flexible model processes in web development. *WSEAS Transactions on INFORMATION SCIENCE and APPLICATIONS*, 2005. Accepted for publication.
- [17] Díaz P., Aedo I., and Montero S. Web engineering in practice a development case. In *Proceedings of WSEAS International conference on Distance Learning and Web Engineering. Corfu*, 2005.
- [18] Zarraonandía T. Context sensitive adaptations on learning design executions, April 2005. "Learning Design Research Colloquium" at UNFOLD "Exploring authoring tools for IMS Learning Design".
- [19] Zarraonandía T., Fernández C., and Dodero J.M. Context sensitive adaptations on learning design executions. In *Proceedings of 5th Int. and Interdisciplinary Conference on Modeling and Using Context CONTEXT'05*, 2005.
- [20] Zarraonandía T., Fernández C., Díaz P., and J. Torres. On the way of an ideal learning system adaptive to the learner and her context. In *Proceedings of Int. Conference on Advanced Learning Technologies ICALT*, 2005.
- [21] Zarraonandía T., Dodero J.M., and Fernández C. Adaptaciones al contexto en tiempo de ejecución de unidades de aprendizaje. In *Actas del Simposio Nacional de Tecnologías de la Información y las Comunicaciones en la Educación. Granada, Septiembre 2005*, 2005.
- [22] Zarraonandía T., Dodero J.M., Aedo I., Díaz P., and A. Sarasa. Domain ontologies integration into the learning objects annotation process. In *Proceedings of Semantic Web for E-Learning Workshop at ITS'2004*, 2004.

***Web-i(2)*: Electronic Institutions: Extending the Core Notion TIC2003-08763-C02**

Pablo Noriega B.V. *
IIIA-CSIC
Barcelona

Sascha Ossowski †
Universidad Rey Juan Carlos
Madrid

Abstract

Electronic Institutions are means to implement interaction conventions for agents — human or software— who can establish commitments in a virtual environment.

The *Web-i(2)* project explores different ways of using and extending the primitive notion of electronic institution. It is organized around three themes: the study of the ideas underlying electronic institutions (EIs), the development of tools and methodologies for the construction of EIs, and the actual construction of examples of EIs. The project draws inspiration from two main sources: the work on coordination in agent societies developed by the URJC team and the work on electronic institutions of the IIIA team.

This document reports the results of the first twenty months of the project life.

Keywords: Multi-agent systems, electronic institutions.

1 Background

Agent-based technologies were originally focused on the features that enabled individual agents to interact, namely agent rationality and communication. The appearance of multi-agent systems (MAS) brought forward the social component of the interactions of groups or populations of agents. Although emphasis in MAS research has largely been put in the internal architecture of those agents and the communication capabilities, issues like task allocation, cooperative and competitive behavior, coordination mechanisms and their constitutive elements have gradually become part of the mainstream.

The notion of *electronic institution* (EI) focuses on the *environment* where agent interactions take place and structures them mirroring the way traditional institutions have articulated interactions among human agents. EIs have been conceived as coordination artifacts where the social elements of the coordination are made explicit and the subjective elements are left to the individual agents that join the electronic institution. The term *electronic institution* was coined at the IIIA to refer to the computational analogue of traditional institutions. Those concerns

*Email: pablo@iia.csic.es

†Email: sascha.ossowski@urjc.es

and ideas behind that proposal have motivated alternative and complementary developments. Nowadays, the notion of virtual organizations, and electronic institutions in particular, has become one of the standard topics of interest in MAS. In the meanwhile the original IIIA proposal has evolved over seven years and is stable enough to deserve revisiting its foundations and goals.

In this context, the *Web-i(2)* project aims at studying the intuitions behind EIs in order to have a sounder theory and better tools to construct open multiagent systems building on the previous work of the two participating groups of researchers. The URJC team has centered their previous research around the use of agent technologies for decision support systems and in particular in communication and coordination among agents for that purpose. In this project they contribute along two lines: (i) the coordination of agent societies through normative structures and (ii) objective and subjective elements in coordination mechanisms in agent societies. The IIIA team has been interested in organizing agent interactions through electronic institutions (EIs) and developing tools and examples around that concept.

The EI proposal is based on the following assumptions:

1. *Agent-architecture neutrality*: Nothing is assumed about the rationality, capabilities or intentions of participating agents.
2. *Dialogical stance*: All agent interactions are construable as speech acts.
3. *Repetitive interactions*: Agent interactions may be organized into some hierarchical system composed by sets of speech acts and relations among these sets.
4. *Institutional commitments*: The institution is the warrant of the intended conditions for illocution utterance and their effects.

These assumptions allow us to think of institutions as a set of conventions that constrain the class of possible dialogues to a set of admissible ones and their intended effects. Likewise, an EI can be seen as a set of norms that regulate agent interactions. We operationalize these assumptions through three simple concepts: the language for illocutions (*dialogical framework*); the structuring of interactions (*performative structure*); and *rules of behavior*, the conventions that regulate commitments (cf. <http://e-institutions.iiia.csic.es/>).

What we call the *core notion*, EI_0 , is one way of defining those three concepts. Roughly speaking, in EI_0 , the dialogical framework establishes a simple illocutory language that involves roles and a rather simple social structure among roles. The performative structure is composed of scenes and transitions between scenes. Scenes are finite-state graphs whose arcs are admissible illocutions and nodes reflect the state of social commitments as they change through the admissible utterances; transitions are a set of standard scenes that are used to control conversation flow. Finally, rules of behavior are bundled up as *obligations* that establish the conditions and social effects of illocutions uttered at some point in the enactment of an institution.

EI_0 produces institutions with explicit, fixed *a priori* conventions whose compliance is obligatory and strictly enforced for participating agents. Although that is a large and quite interesting set of institutions, in the present project we intend to go beyond this state of the art and investigate essentially along three major lines: (i) to develop a better understanding of the core notions of EIs, (ii) to develop adequate tools for building EI_0 -based institutions, and (iii) to be able to define and implement other types of institutions.

2 Project objectives

The *Web-i(2)* project aims at making explicit and precise the intuitions of the EI proposal by looking carefully into EI_0 and exploring some extensions that are formally or empirically interesting, in order to have better tools for modeling institutional environments that are more complex and require more flexibility than those EI_0 may give us. We translated these intentions into five tasks that organize the analysis of EI_0 and its potential extensions, the construction of tools that enable the design, construction and refinement of actual EIs and the exploration of application domains that may serve to elicit and test the extensions of EI_0 . The tasks, subtasks and deliverables that correspond to these objectives of project *Web-i(2)* are summarized in Table 1.

Coord	Task	Subtask	Resp.	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	q11	q12
PN	T1: Core Notion Extensions	a) Core Notion	PN	X	X	X	R	X	X	X	R	X	g	g	R
		b) Alternative Normative Notions	PN	X	X	X	R	X	X	X	R	X	X	X	R
		c) Dynamic. Institutions	CS	X	g	g	R	X	X	X	R	X	X	X	R
		d) Negotiation	FE	X	X	X	R	X	X	X	R	X	X	X	R
		e) Formal treatment	LV	X	X	X	R	X	X	X	R	X	X	X	R
		f) Institutions <i>ABSM</i>	SO	g	g	X	R	X	X	X	R	X	X	X	R
JR	T2: Tools for an Institutional Layer	a) <i>ISLANDER</i> extensions	CS	X	X	X	RP	X	X	X	RP	X	g	X	RP
		b) <i>Governors</i>	JR	g	g	g	RP	g	g	g	RP	g	g	g	RP
		c) Governor validation	LV	g	g	g	RP	g	g	g	RP	g	g	g	RP
		d) Activation	JR	g	g	g	RP	g	g	g	RP	g	g	g	RP
		e) Executability tests	JR	g	g	g	RP	g	g	g	RP	g	g	g	RP
SO	T3: Visualization and Modelling Tools	a) Visualization	JR	g	X	X	RP	X	X	X	RP	X	g	X	RP
		b) Simulation	AF	g	X	X	RP	X	X	X	RP	X	X	X	RP
		d) <i>ABSM</i>	JMS	g	X	X	RP	X	X	X	RP	X	X	X	RP
JR	T4: Application Domains	a) Mechanism Design	JR	g	g	X	RP	X	g	X	RP	X	X	X	RP
		b) <i>DSS-ABSM</i>	SO	g	g	X	RP	X	g	X	RP	X	X	X	RP
		c) Multifunctional EIs	PN	g	g	X	RP	X	g	X	RP	X	X	X	RP
PN	T5: Dissemination and Technology Transference	a) Coordination	PN	C	C	C	C	C	C	C	C	C	C	C	C
		b) Prototype evaluation	JR	g	X	X	R	X	X	X	R	X	X	X	R
		c) TT entities	PN	M	g	C	C	C	g	C	C	C	C	C	C
		d) European Space	SO	X	X	X	X	X	X	X	X	X	X	X	W
		e) Dissemination of results	PN	R	R	R	R	R	R	R	R	R	R	R	R

Keys:

Researchers:

IIIA: a) J. Agustí, Francesc Esteve (FE), Pablo Noriega (PN), Juan Antonio Rodríguez (JR), Carles Sierra (CS), Lluís Vilà (LV)

URJC: a) Sascha Ossowski (SO), A. Pradera, Juan Manuel Serrano (JMS)

q1, ..., q12: quarters

X = Coordinated activity for both groups in this task
P = Prototype

g = Independent activity of one group at least
C = Coordination meeting

R = Report
W = Workshop

Table 1: *Web-i(2)* Workplan

3 Accomplishments

The project has proceeded along the expected path. The two groups have interacted fruitfully, holding regular coordination meetings, plus exchanging visits, co-organizing events and producing complementary and joint developments as the following paragraphs show. Most of the original project objectives have either been accomplished already or are well on the way of being accomplished. The overall balance is very positive and, since the project is about halfway through, we anticipate a successful conclusion.

Current achievements suggest productive and challenging collaborative work beyond the *Web-i(2)* project time-frame.

Theoretical extensions of the core notion of electronic institution (Task T1). In the last twenty months the groups have been involved in a careful discussion of the intuitions behind each of the components of EI_0 following three complementary lines of work. First, there has been a characterization of different structural operations on *performative structures*, second, a first attempt on defining standard —reusable— scenes and, finally, on expressing conventions as rules (or norms).

- The idea behind *structural extensions* of EI_0 is to be able to splice new scenes into existing structures, or even whole performative substructures assembled from other substructures. The purpose is twofold, on one hand to facilitate a modular construction and testing of EIs and, on the other, enabling a straightforward mechanism for EI updating or modification. In particular, we have characterized hierarchic *performative structures* so that one institution may be embedded into another provided some consistency constraints hold. This feature is being incorporated into the forthcoming release of the EIDE platform mentioned below.
- We have attempted an alignment of the RICA conceptual framework developed by the URJC team with EI_0 . In that process we looked into generic extensions of the RICA communication frameworks and their meta-model. That effort has led us to explore *standard scenes* that may be instantiated or refined to fit particular requirements while being readily spliced into a *performative structure*. This line of research is based on the social communication contexts proposed by J. M. Serrano in his dissertation and the splicing techniques just mentioned. We expect to move in the direction of some sort of compositional language to specify and modify performative structures. We are now using these ideas plus the structural extensions mentioned above in the problem domains mentioned in Task 4.
- The idea of expressing *conventions as norms* is really a generalization of the behavior rules that underlie the notion of commitment in EI_0 . The purpose is to formalize EIs as theories in a formal language (with some type of deduction). We have, so far, proposed a very conservative extension to the current notion of obligation and with that we have defined and implemented an interpretable language that is expressive enough to capture many types of normative expressions. This topic is theoretically very fertile and has attracted the attention of many researchers in MAS. It has also been the motivation for us to organize three different meetings that we mention below.

A somewhat complementary line of research was spawned during the visit of Michael Luck and Mark D’Inverno to the IIIA this winter: a *Z-specification* of EI_0 for which a tentative first version has already been produced.

Tools for an institutional layer in MAS (Task T2). Our aim is to produce a toolbox for building EIs in a principled and practical way. We started from *ISLANDER* and produced an extension of it plus a complete suit of tools for specification and deployment of EIs. The tools have been organized into an electronic institutions development environment, EIDE, whose third release is now underway. The following are the tools in EIDE:

- *Tools for specification and validation of EIs.* *ISLANDER* has become a graphic specification language that includes some automated verification functionalities of syntactic

features (completion, consistency and coherence). Extensions to RICA-J for AUML-based organizational models, and further extensions to incorporate normative elements in the specification are underway.

- *Tools for deployment.* *AMELI* is a middleware layer that enables the enactment of *ISLANDER* specified EIs on top of a FIPA-compatible agent communication infrastructure like JADE. *AMELI* generates all staff agents for an EI, including the *governor* agent wrappers for external agents. Additionally, we are working on extensions to RICA-J to deal with scalability in EIs.
- *Tools for the generation of agents.* Although *AMELI*-generated *governors* handle all the communication constraints and flow control conventions that the EI imposes on external agents —without assuming any particular external agent architecture— users still need to have their agents decide what to communicate and when. *AGENTBUILDER* is a tool that may be used to construct agents that comply with *ISLANDER* spawned governors, hence liberating the external designer from any interfacing concerns and allowing her to focus only in the decisional aspects of her agents.

Visualization and modeling tools (Task T3). Following the original workplan we are working along two other lines of tool building activities that complement those in Task T2.

- *Visualization tools.* In order to observe the activities of agents in an institutional-based MAS, we are building a visualization tool —based on the RICA organization and communication concepts— to show in an intuitive way the organizational activity, the roles agents are playing and their interactions. For the same purposes, we have developed (In collaboration with the University of Technology of Sydney) a tool that produces a 3-D animation of any *ISLANDER*-specified EI.
- *Tools for simulation and monitoring.* Given an *ISLANDER* specification of an EI, the *SIMDEI* tool produces a lightweight (object-based) version of it —including simple agent skeletons for all agents involved— that the designer may use to realistically explore the functionalities of the EI and go back to the drawing board if needed. *SIMDEI* is built around a REPAST simulator. In addition to *SIMDEI* and the visualization tools mentioned above, we have also built, *EIMONITOR*, a tool that keeps track of all the interactions that take place in an EI during a run-time enactment of it. *EIMONITOR* works on top of the *AMELI* middleware.

Application domains (Task T4). We are working in the following directions:

- *Mechanism design.* Analysis and implementation of a new type of negotiation mechanism: generalized combinatorial auctions. Conceived as non-trivial EI_0 -definable EIs are now undergoing the typical analytical testing for mechanism design, plus the empirical testing through its EIDE implementation. We have also implemented several examples of other EIs.
- *Negotiation.* Work has proceeded in two lines. First the design and testing of decisional models for fish-buying agents that participate in the federated auctions of the MASFIT

market institution. Second, the development of a theory and experimental methodology of Information-based negotiation, in collaboration with J. Debenham (U. Technology Sydney).

- *Modeling agent-based social systems.* We were involved with the World Tourism Organization and a tourism consultant firm in the development of an agent-based DSS for the management of tourism destinations, TMS. A prototype, based on the SADDE methodology, was released in August 2004. For the time being, we do not foresee further work in this direction.
- *Complex-flexible institutional environments.* This line is where we are putting most of our effort. We are exploring three problem domains that have many significant features in common: Conference workflow management; Help desk services and local government actions; corporate hotel management. The three domains are amenable to EI treatment and require the type of structural and normative flexibility we have mentioned above.

Dissemination and technology transfer (Task T5).

- *Coordination tasks.* Both groups have been in touch regularly. There have been three scheduled group meetings (two in Barcelona, one in Madrid) and short reciprocal visits from the PhD students. In addition, the many international events organized by the teams have provided extra opportunity to collaborate. A project webpage is up (<http://webi2.urjc.es/englishIndex.html>).
- *Prototype evaluation.* The EIDE open-source platform has had two different releases. Furthermore, a *beta* version of a new EIDE upgrade and five implemented EI_0 examples are now publicly available on-line (<http://e-institutor.iiia.csic.es/islander/pub>). More than one hundred copies of EIDE have been distributed in courses, conferences, demos and upon request.
- *Contacts with potential user firms.* In January 2005 we organized a meeting in Madrid with firms that are working on the problem domains we are interested. The meeting served to exchange experiences and consider possible joint activities. We are also preparing a technology transference contract with a software house to develop an EI-platform.
- *Contacts in the European space.* The group members have been very active in this sub-task. Most actions are listed below – here we only outline the most significant ones. The *Web-i(2)* partners co-sponsored and headed the local organization of the European Conference in Multi-agent Systems (EUMAS04); co-organized and co-sponsored a round table meeting on virtual organizations and organized two workshop on electronic institutions that are co-located with large international conferences. Group members participated in the organization of several workshops and meetings whose content is associated with this project, and are part of the agent-related European networks of excellence. Both partners are involved in VI Framework Program projects that are closely related to the *Web-i(2)* subjects. Both groups have stable academic relationships with foreign universities and research groups, and have exchanged several visits over the last twenty months.

- *General dissemination activities.* Group members taught seminars and short courses on *Web-i(2)* subjects in the European Summer and Spanish schools on agents (EASS-04, 05; AgentCities.ES-05) and in universities in the UK, Holland, Spain and New Zealand. There was a significant number of papers presented in international conferences, students attended courses on topics related to the project's content and there were demonstrations of EI developments in four international conferences (ECAI-04, AAMAS-04, AAAI-04, MATES-04). The project sponsored two international events and received public recognition for it.

4 Result indicators

This section summarizes the results obtained until July 2005, i.e. the first 20 months of the *Web-i(2)* project lifetime.

4.1 Scientific and technological production

Publications. At the time of this writing, 43 publications have been realized within the *Web-i(2)* project:

- 9 books (edited): [18, 6, 17, 19, 9, 10, 11, 39, 16]
- 12 book chapters (Springer, Kluwer, etc): [30, 21, 23, 13, 8, 20, 26, 28, 4, 42, 43, 25]
- 7 articles in international journals: [24, 22, 33, 41, 1, 36, 38]
- 9 papers at international conferences (ACM, IEEE, etc): [32, 3, 34, 12, 37, 5, 7, 27, 40]
- 6 other publications: [2, 31, 14, 15, 35, 29]

Software developments. One software platform, *EIDE* (with several tools: *AMELI*, *aBuilder*, *ISLANDER*, *SIMDEI*, *EIMONITOR*); the *EI in 3-D* visualization software; the RICA-J specification and visualization tools; *IBundler* combinatorial auctions server and an *AMELI*-compatible norm language interpreter prototype.

4.2 Education of human resources

Theses. The following seven PhD theses are linked to the *Web-i(2)* project.

- Eva Bou. Dynamic and reconfigurable electronic institutions. Under development.
- Alberto Fernández. Semantic service coordination in open environments. Under development.
- Andres García-Camino. Norm-based electronic institutions.
- Andrea Giovannucci. Combinatorial auctions. Under Development.
- Ramón Hermoso. Agent-based social simulation. Under development.

- Armando Robles: EI-based corporate information systems.
- Sergio Saugar. Coercive multiagent infrastructures. Under development.
- Juan Manuel Serrano. Pragmática de los agentes software: análisis y diseño de lenguajes de comunicación artificiales. (URJC, Jul 2004).

In addition, several undergraduate thesis have been carried out within the project. Recently, the IIIA team received a grant from CSIC for funding PhD research assistantships on electronic institutions. Four students will be chosen from the more than one hundred candidates who submitted CVs for those assistantships.

Courses taken. PhD students of the *Web-i(2)* teams have enrolled in the following advanced courses:

- AAI-2004, (S.José, CA), [Andrea Giovannucci]
- AAGTSLN-2004 on complex systems, co-organized by the EXYSTENCE European network of excellence (Santorini) [Jorge San Martín].
- AAMAS2004 Americas' School on Agents and Multiagents Systems (New York) [Andrea Giovannucci]
- EASSS-2004 on agent technologies, AgentLink III (Liverpool) [Eva Bou, Andrea Giovannucci, Andrés García, Sergio Saugar].
- EASSS-2005 on agent technologies, AgentLink III (Amsterdam) [Ramón Hermoso].
- ECAI 2004, (Valencia) [Andrea Giovannucci]
- Escuela Española de Agentes, AgentCities.ES (Sevilla) [Andrés García, Ramón Hermoso].

Courses taught. The research performed within the project has found its way into several graduate and final-year undergraduate lectures. In addition, the following courses have been taught by senior members of the teams

- Agent Mediated Interactions, PRIMA (New Zealand) [Carles Sierra].
- Agent-mediated Electronic Commerce, EASS04 (Liverpool) [Juan Anonio Rodríguez].
- Agent technology applied. Escuela Superior de Administración y Dirección de Empresas (ESADE) [Juan Antonio Rodriguez-Aguilar]
- Desarrollo de Sistemas Multiagente, Ikerlan (Mondragón) [Marc Esteva and Juan Antonio Rodríguez].
- Engineering open environments with electronic institutions (University of Aberdeen) [Juan Antonio Rodriguez-Aguilar]
- Instituciones Electrónicas, Escuela Española de Agentes (Sevilla) [Josep Lluís Arcos].

4.3 Relations with the socio-economic context

Technology transference.

- Contract proposal with a software house, Grupo TCA.
- DMR grant for applied research (DMR Decision Engineering Lab)
- Contacts with Telefónica, Barcelona City Government, Blanes City Hall, AUTECH, World Trade Organization, Mazars Consulting, Grupo TCA.
- More than 150 downloads of the EIDE platform from the *Web-i(2)* site and more than two hundred cd-copies of it distributed.

Participation in related projects with industrial participation.

- Context-Aware Business Application Service Coordination in Mobile Computing Environments (CASCOM). Funded by the 6th FP of the EU as STREP. 2004-2007
- Ever-growing global scale-free networks, their provisioning, repair and unique functions (EVERGROW). Funded by the 6th FP of the EU as IP. 2004-2007
- OK (Open Knowledge). EI-based P2P web-services. Funded by the 6th FP of the EU as STREP. 2006-2009

Participation in networks

- European network on agent technology (AgentLink III)
- European network on complex systems (EXYSTENCE)
- Spanish network on agent technology (AgentCities.ES)

4.4 Relations with European and international teams

International collaborations The project has greatly helped establishing and consolidating links with other research groups. These links have materialized on joint publications, joint organization of events, and joint projects. The following is a list of the main researchers and groups that the *Web-i(2)* teams are currently collaborating with:

- Huib Aldewereld (U. Utrecht, NL)
- Ron Ashri (U. of Southampton, UK).
- Anton Bogdanovych (U. of Technology Sydney, AS).
- Paul Davidsson, (Blekinge Institute of Technology, SE)
- John Debenham (U. of Technology Sydney, AS).
- Frank Dignum (U. of Utrecht, NL)

- Les Gasser (U. of Illinois Urbana, US)
- Mark d’Inverno (U of Westminster, UK)
- Matthias Klusch (DFKI, DE)
- Michael Luck (U. of Southampton, UK)
- Ronaldo Menezes, (Florida Institute of Technology, US)
- Marcos Oliveira (U. of Otago, NZ)
- Andrea Omicini, (U. of Bologna, IT)
- Julian Padget, (U. of Bath, UK)
- Sarvapali Ramchurn (U. of Southampton, UK)
- Wamberto Vasconcelos (U. of Aberdeen, UK)

4.5 Other dissemination activities

Organization of international events.

- ACM SAC-2004 Special Track on Coordination Models, Languages and Applications (Mar 04, Nicosia) [Sascha Ossowski]
- Agent Link II. Technical Forum Group in Trust (Jun 04, Rome) [Juan Antonio Rodriguez-Aguilar]
- AAMAS04. (Jul 04, New York) [Carles Sierra]
- AMEC-VI 2004. (Jul 04, New York) [Juan Antonio Rodriguez-Aguilar]
- ATT 2004 (Jul 04, New York) [Sascha Ossowski]
- ECAI 2004 (Aug 04, Valencia) [Carles Sierra]
- ECAI Workshop on Coordination in Emergent Agent Societies (Aug 04, Valencia) [Sascha Ossowski]
- Eighth International Workshop CIA 2004 on Cooperative Information Agents (Sep 04, Erfurt, Germany) [Sascha Ossowski]
- EUMAS04 European Conference on Multiagent Systems (Dec 04, Barcelona) (with *Web-i(2)* sponsorship) [Pablo Noriega]
- ACM SAC-2005 Special Track on Coordination Models, Languages and Applications (Mar 05, Santa Fe) [Sascha Ossowski]
- Barcelona Round Table on Virtual Institutions and Legal Theory: The confluence of legal and multi-agent systems perspectives (May 05, Barcelona) (with *Web-i(2)* sponsorship) [Pablo Noriega]

- XXII World Congress of Philosophy of Law and Social Philosophy IVR 2005. Workshop on AI and LAW: The regulation of electronic social systems. (May 05, Granada) [Pablo Noriega]
- AAMAS-2005 Workshop on Agents, Norms and Institutions for Regulated Multiagent Systems (ANIREM) (Jul 05, Utrecht) [Sascha Ossowski]

Invited talks

- PRICAI 2004, (Auckland, NZ) Electronic Institutions: Methodology of Multi-agent Systems development [Carles Sierra];
- I-Sourcing Forum, (Barcelona) *Tendencias Futuras en I-Sourcing* [Juan Antonio Rodríguez];
- Foro Tecno-atlántico 2004 (Vigo) *MAFIT y subastas intermediadas por agentes* [Pere Garcia];
- ESAW 2004, (Toulouse) *Electronic Institutions an Explorer's Notebook* [Pablo Noriega];
- ANIREM 2005 (Utrecht) *Fencing the Open Fields: Empirical Considerations on Electronic Institutions* [Pablo Noriega].

References

- [1] Josep Lluís Arcos, Marc Esteva, Pablo Noriega, Juan A. Rodríguez-Aguilar, and Carles Sierra. Environment engineering for multiagent systems. *Engineering Applications of Artificial Intelligence*, (submitted), October 2004.
- [2] Holger Billhardt, Ramón Hermoso, Sascha Ossowski, Sergio Saugar, and Juan Manuel Serrano. Infrastructures for virtual institutions. In *First Round Table on Virtual Institutions and Legal Theory. The Confluence of Legal and Multi-Agent Systems Perspectives*, Barcelona, 2005.
- [3] Holger Billhardt and Sascha Ossowski. Towards agent societies for information retrieval. In *IEEE Int. Conference on System, Man and Cybernetics*, volume VI, pages 5584–5589. IEEE Press, 2004.
- [4] Anton Bogdanovych, Helmut Berger, Simeon Simoff, and Carles Sierra. 3D Electronic Institutions: Social Interfaces for E-Commerce. In Mark Billinghurst and Andy Cockburn, editors, *Sixth Australasian User Interface Conference (AUIC'2005)*, volume 40 of *Conferences in Research and Practice in Information Technology*, Newcastle, Australia, 2004.
- [5] Guifré Cuní, Marc Esteva, Pere Garcia, Eloi Puertas, Carles Sierra, and Teresa Solchaga. MAFIT: Multi-Agent System for Fish Trading. In *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI 2004)*, Valencia, Spain, 2004.

- [6] José Luís Pérez de la Cruz, Francisco Garijo, Juan Manuel Corchado, Ana García Serrano, Sascha Ossowski, Juan Pavón, and Vicente Botti. *Agentes Software y Sistemas Multiagente – Conceptos, Arquitecturas y Aplicaciones*. Pearson Prentice Hall, 2004. ISBN: 8420543675.
- [7] Marc Esteva, Juan A. Rodríguez-Aguilar, Bruno Rosell, and Josep Lluís Arcos. AMELI: An agent-based middleware for electronic institutions. In *Third International Joint Conference on Autonomous Agents and Multi-agent Systems (AAMAS'04)*, pages 236–243, New York, USA, July 19-23 2004.
- [8] Marc Esteva, Wamberto Vasconcelos, Carles Sierra, and Juan A. Rodríguez-Aguilar. Norm consistency in electronic institutions. In *Proceedings of the XVII Brazilian Symposium on Artificial Intelligence - SBIA'04*, volume LNAI 3171 of *Lecture Notes in Artificial Intelligence*, pages 494–505. Springer, 2004.
- [9] Peyman Faratin, David Parkes, Juan Antonio Rodríguez-Aguilar, and William Walsh, editors. *Agent Mediated Electronic Commerce V: Designing Mechanisms and Systems*, volume 3048 of *Lecture Notes in Artificial Intelligence*. Springer, 2004.
- [10] Peyman Faratin and Juan Antonio Rodríguez-Aguilar, editors. *Agent Mediated Electronic Commerce VI: Theories for and Engineering of Distributed Mechanisms and Systems*, volume (in press) of *Lecture Notes in Artificial Intelligence*. Springer.
- [11] Michael Fisher, Dov M. Gabbay, and Lluís Vila, editors. *Handbook of Temporal Reasoning in Artificial Intelligence*. Elsevier, 2005.
- [12] Andres Garcia-Camino, Pablo Noriega, and Juan Antonio Rodríguez-Aguilar. Implementing Norms in Electronic Institutions. In *Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, 2005. (forthcoming).
- [13] Andres Garcia-Camino, Juan Antonio Rodríguez-Aguilar, Carles Sierra, and Wamberto Vasconcelos. A Distributed Architecture for Norm-Aware Agent Societies. In *Fourth International Joint Conference on Autonomous Agents and Multiagent Systems. Declarative Agent Languages and Technologies workshop (DAL'T'05)*, 2005. (forthcoming).
- [14] Ramón Hermoso, Rubén Ortiz, Sergio Saugar, and Juan Manuel Serrano. Instrumentación de sistemas multiagente mediante un entorno organizativo/lingüístico: Un caso práctico. In *JISBD Workshop sobre el desarrollo de sistemas multiagente (DESMA 2004)*, Málaga, 2004.
- [15] Ramón Hermoso, Sergio Saugar, Juan Manuel Serrano, and Sascha Ossowski. Un modelo de integración de conceptos institucionales en una plataforma de programación de agentes. In *CEDI - II Workshop sobre Desarrollo de Sistemas Multiagente (DESMA-2005)*, Granada, 2005.
- [16] Nick Jennings, Carles Sierra, Liz Sonenberg, and Milind Tambe. *Proceedings of the Fourth International Conference on Autonomous Agents and Multiagent Systems*. IEEE Press, New York, 2004.

- [17] Franziska Klügl, Ana Bazzan, and Sascha Ossowski, editors. *Applications of Agent Technology in Traffic and Transportation*. Number VIII in Whitestein Series in Software Agent Technologies. Birkhauser, 2005. ISBN: 3-7643-7258-3.
- [18] Matthias Klusch, Sascha Ossowski, Vipul Kashyap, and Rainer Unland, editors. *Cooperative Information Agents VIII, 8th International Workshop, CIA 2004, Erfurt, Germany, September 27-29, 2004, Proceedings*, volume 3191 of *Lecture Notes in Computer Science*. Springer, 2004.
- [19] Gabriela Lindemann, Sascha Ossowski, Julian Padget, and Javier Vazquez, editors. *Agents, Norms and Interactions for Regulated Multiagent Systems, ANIREM 2005, Utrecht, The Netherlands, Proceedings*, 2005.
- [20] Pablo Noriega, Juan Antonio Rodriguez-Aguilar, and Carles Sierra. Aplicaciones de tecnología de agentes en comercio electrónico. In *Agentes software y sistemas multiagente: conceptos, arquitecturas y aplicaciones [6]*. Pearson, 2004. (forthcoming).
- [21] Andrea Omicini, Sascha Ossowski, and Alexandro Ricci. *Methodologies and Software Engineering for Agent Systems – The Agent-Oriented Software Engineering Handbook*, chapter Coordination Infrastructures in the Engineering of Multiagent Systems, pages 273–296. Kluwer, 2004.
- [22] Sascha Ossowski et al. Multiagent system for decision support: A case study in the transportation management domain. *Transportation Research C*, 2005.
- [23] Sascha Ossowski, Alberto Fernández, Sergio Saugar, and María Victoria Belmonte. *Agentes Software y Sistemas Multiagente – Conceptos, Arquitecturas y Aplicaciones*, chapter La Coordinación en los Sistemas Multiagente, pages 105–144. Pearson Prentice Hall, 2004.
- [24] Sascha Ossowski and Ronaldo Menezes. On coordination and its significance to distributed and multi-agent systems. *Journal of Concurrency and Computation: Practice and Experience*, 2005.
- [25] Han Reichgelt and Lluís Vila. Temporal qualification in artificial intelligence. In Dov M. Gabbay Michael Fisher and Lluís Vila, editors, *Handbook of Temporal Reasoning in Artificial Intelligence*, pages 537–556. Elsevier, 2005.
- [26] Antonio Reyes-Moro and Juan Antonio Rodriguez-Aguilar. iAuctionMaker: A decision support tool for mixed bundling. In *Agent Mediated Electronic Commerce VI: Theories for and Engineering of Distributed Mechanisms and Systems*, volume (forthcoming) of *Lecture Notes in Artificial Intelligence*. Springer, 2004.
- [27] Juan Antonio Rodriguez-Aguilar, Andrea Giovannucci, Antonio Reyes-Moro, Francesc X. Noria, and Jesús Cerquides. Towards automated procurement via agent-aware negotiation support. In *Third International Joint Conference on Autonomous Agents and Multi-agent Systems (AAMAS’04)*, pages 244–251, New York, USA, July 19-23 2004.

- [28] Juan Antonio Rodríguez-Aguilar, Antonio Reyes-Moro, Jesús Cerquides, Andrea Giovannucci, and Francesc X. Noria. Negotiation support in highly-constrained trading scenarios. In *Selected Papers from (CAEPIA'2003)*, volume 3040 (forthcoming) of *Lecture Notes in Artificial Intelligence*. Springer, 2004.
- [29] Sergio Saugar, Rubén Ortiz, Juan Manuel Serrano, and Sascha Ossowski. Una arquitectura multiagente para asesoramiento bursátil on-line. In V. Botti and E. Corchado, editors, *Third International Workshop on Practical Applications of Agents and Multi-Agent Systems (IWPAAMS'04)*, pages 41–50. Univ. de Burgos, Oct 2004.
- [30] Juan Manuel Serrano and Sascha Ossowski. On the impact of agent communication languages on the implementation of agent systems. In Klusch, Ossowski, and Kashyap, editors, *Cooperative Information Agents VIII*, LNCS, pages 92–100. Springer, 2004.
- [31] Juan Manuel Serrano and Sascha Ossowski. RICA-J – a dialogue-driven software framework for the implementation of multiagent systems. In González, Cuesta, and Pavón, editors, *JISBD - I Taller en Desarrollo de Sistemas Multiagente (DESMA-2004)*, pages 48–61, Málaga, Nov 2004.
- [32] Juan Manuel Serrano and Sascha Ossowski. A semantic framework for the recursive specification of interaction protocols. In *Coordination Models, Languages and Applications. Special Track of the 19th ACM Symposium on Applied Computing (SAC 2004)*, pages 451–456. ACM Press, 2005.
- [33] Juan Manuel Serrano and Sascha Ossowski. A computational framework for the specification and enactment of interaction protocols in multiagent organizations. *Journal of Web Intelligence and Agent Systems*, forthcoming. forthcoming.
- [34] Juan Manuel Serrano, Sascha Ossowski, and Sergio Saugar. Engineering agent interactions from acl-based reusable connectors. In *Fourth International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'05)*. ACM Press., 2005. Poster Paper.
- [35] Juan Manuel Serrano, Sascha Ossowski, and Sergio Saugar. Reusability issues in the instrumentation of agent interactions. In *Third International Workshop on Programming Multi-Agent Systems: Languages and Tools (ProMAS'05)*, Utrecht, Jul 2005.
- [36] Carles Sierra. Agent-mediated electronic commerce. *Autonomous Agents and Multi-Agent Systems*, 9(3):285–301, 2004.
- [37] Carles Sierra. Electronics institutions: Methodology of multi-agent systems development. In Chengqi Zhang, Hans W. Guesgen, and Wai-Kiang Yeap, editors, *PRICAI 2004: Trends in Artificial Intelligence, 8th Pacific Rim International Conference on Artificial Intelligence, Auckland, New Zealand, August 9-13, 2004, Proceedings*, volume 3157 of *Lecture Notes in Computer Science*. Springer, 2004.
- [38] Carles Sierra, Juan Antonio Rodríguez-Aguilar, Pablo Noriega, Marc Esteva, and Josep Lluís Arcos. Engineering multi-agent systems as electronic institutions. *European Journal for the Informatics Professional*, V(4), August 2004.

- [39] Carles Sierra, Jordi Sabater, Jaume Augustí, and Pere Garcia. SADDE: Social agents design driven by equations. In F. Bergenti, M.P. Gleizes, and F. Zambonelli, editors, *Methodologies and software engineering for agent systems*. Kluwer Academic Publishers, 2004.
- [40] Wamberto Vasconcelos, Marc Esteva, Carles Sierra, and Juan A. Rodriguez-Aguilar. Verifying norm consistency in electronic institutions. In *Proceedings of the AAAI-04 Workshop on Agent Organizations: Theory and Practice*, pages 8–14, San Jose, CA, July 25-29 2004.
- [41] Wamberto Vasconcelos, Dave Robertson, Carles Sierra, Marc Esteva, Jordi Sabater, and Michael Wooldridge. Rapid prototyping of large multi-agent systems through logic programming. *Annals of Mathematics and Artificial Intelligence*, 41(2–4):135–169, 2004.
- [42] Lluís Vila. Formal theories of time and temporal incidence. In Dov M. Gabbay Michael Fisher and Lluís Vila, editors, *Handbook of Temporal Reasoning in Artificial Intelligence [11]*, pages 1–24. Elsevier, 2005.
- [43] Lluís Vila and Hajime Yoshino. Time in automatic legal reasoning. In Dov M. Gabbay Michael Fisher and Lluís Vila, editors, *Handbook of Temporal Reasoning in Artificial Intelligence [11]*, pages 537–556. Elsevier, 2005.

Web services, parallelism, evolution and complex systems: NadeWeb (nuevos algoritmos distribuidos y evolutivos en la web, new distributed evolutive algorithm on the web)

TIC2003-09481-C04

Juan Julián Merelo Guervós *
Departamento ATC, Universidad de Granada

Ken Sharman †
ITI, Universidad Politécnica de Valencia

Isabel Román ‡
Departamento Economía Financiera y Contabilidad, Universidad de Granada

Abstract

In this paper we present a half-time evaluation of results obtained in the NadeWeb coordinated project. This projects intends to evaluate and enhance the performance of the DRM Distributed Resource Machine, a highly scalable distributed virtual machine, as well as use it for evolutionary computation tasks and apply it to real-world problem such as the evaluation of company risks or the analysis of word streams looking for some temporal structure.

Keywords: Distributed computation, web services, performance evaluation, evolutionary computation, human-computer interface, adaptive computation applications.

1 Project Objectives

The main objectives of the NadeWeb project has been deepening the understanding of the DRM, Distributed Resource Machine, a P2P virtual machine for general computation, and, specifically, to measure its performance in a distributed computation environment. DRM was released as a part of the DREAM project [Paechter. et al., 2000], but the IST-FET project that spawned it did not include extensive testing and evaluation. Further studies proved that its scaling behavior did not correspond to what theoretically should be expected [Arenas, 2005], and

*Email: jmerelo@geneura.ugr.es

†Email: ken@iti.upv.e

‡Email: iroman@ugr.es

specially Maribel Garcia Arenas' PhD thesis; thus, the objective of this project was to extensively evaluate its performance, debug and correct flaws, and finally, use it for evolutionary computation, as the original project intended. DREAM also included the release of the JEO library [Arenas et al., 2004, Arenas et al., 2002b, Arenas et al., 2002a, Arenas et al., 2001], which was adapted for distributed computation, but had several problems. First, it used DRM for statistics collection, but not for migrating objects under evolution; second, it used a classic evolutionary computation paradigm: a generational island model, which did not really take advantage of the ad-hoc and asynchronous nature of the DRM virtual machine. That is why a second objective of the project was to design a new evolutionary/adaptive algorithm that really leveraged DRM. Since this project is a coordinated one, it included several subgroups more interested in the application of these algorithms, as well as advancing the state of the art in those particular applications. These applications were human-computer interfaces, natural language processing and bankruptcy prediction. The objective was to improve results in those areas, as well as use them as testbeds for the algorithms developed.

2 Success level reached in the project

Since we are barely halfway through the project, there have not been many results; mainly, researchers that have been hired for the project are getting up-to-date on the project topics. One of them was hired by March 2004, and the other one by September 2004. There have been also some changes in the project staff; a new researcher has entered the project, another one has changed status and location, and finally, one has left the project. However, a few objectives have been reached.

2.1 DRM performance evaluation

Two problems have been solved using evolutionary computation on the DRM: the classical OneMax and knapsack problem. The main objective was to check what were its scaling properties, when new nodes were added. Results for up to 32 nodes are shown in figure 1.

However, at this point in time, it is not clear what are the causes of this lack of speedup. It could be due to an implementation defect, or it could be essential to the nature of the newscast algorithm: items on each node cache are sent time and again while they have not been substituted by another; this could be fixed by changing the cache size, but it could indicate a deeper problem.

At any rate, the newscast algorithm has been extensively studied, and as an alternative for further work, its reimplementation using web services or libraries such as JXTA [JXTA,] will be considered.

2.2 Genetic programming on the DREAM platform

With support from the Granada-ATC subproject, the ITI-UPV subproject has focused on genetic programming [Riolo and Worzel, 2003], updating the GP modules in JEO and applying them to their main research axis: brain-computer interfaces (BCI).

In fact, many of the structures obtained would have been competitive in the BCI competition that is held every year.

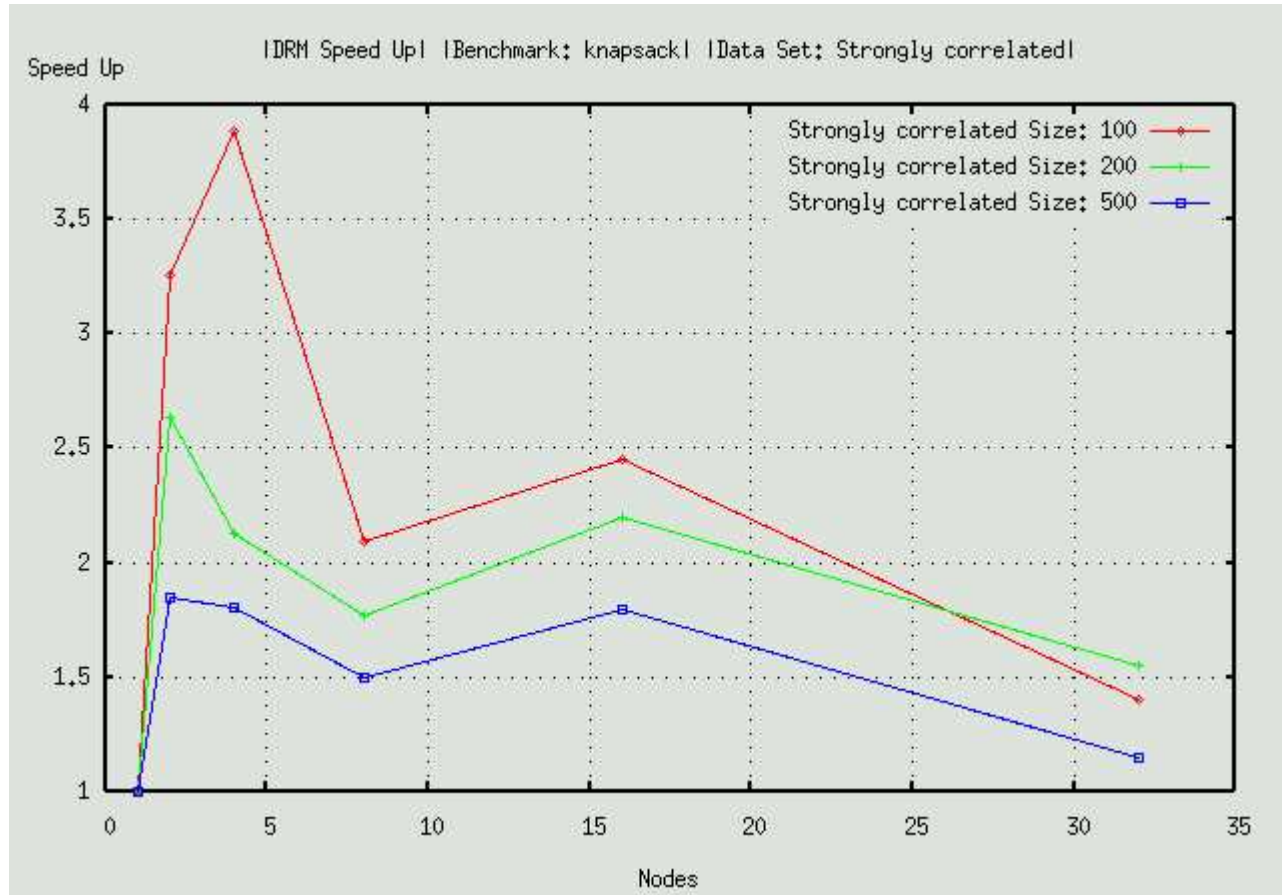


Figure 1: DRM speed-up for the knapsack problem solved using evolutionary computation; nodes are increased along the x axis, while the y axis shows speed up. Speedup for more than 16 machines is clearly lower than expected.

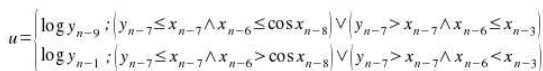


Figure 2: One of the best structures computed using genetic programming

2.3 Memetic algorithms in genetic programming

Genetic programming includes several constants, whose change via mutation takes a very long time; combining this global algorithm with a local algorithm such as simulated annealing can yield better results in less time. The ITI-UPV group is testing such combination on generated signals; it still takes some time, making it a good candidate for parallelization.

2.4 Novelty detection in text streams

In constantly changing text streams, such as weblogs, news sites or Internet forums, it is almost impossible to know what topics are *hot* in a particular moment. Some algorithms have been developed to follow text streams and *fire* when some topic is being mentioned more than usual [Kleinberg, 2002]; however, these algorithms, approached from the point of view of optimization, have not been measured extensively, and, besides, its efficiency, as well as predictive power, could probably be improved using global optimization algorithms such as evolutionary algorithms. The number of times a term appears in a text stream (in this case, a set of weblogs) is shown in figure 3

A model for these streams, which consists of different *states*; (the higher, the *hotter* is this issue at a particular point in time), has been created by comparing different measures of proximity, and finally fitting using an evolutionary algorithm; the combination of both techniques yields a fit that is better than the one found by Kleinberg previously. An example of fits with different distance measures is found in figure .

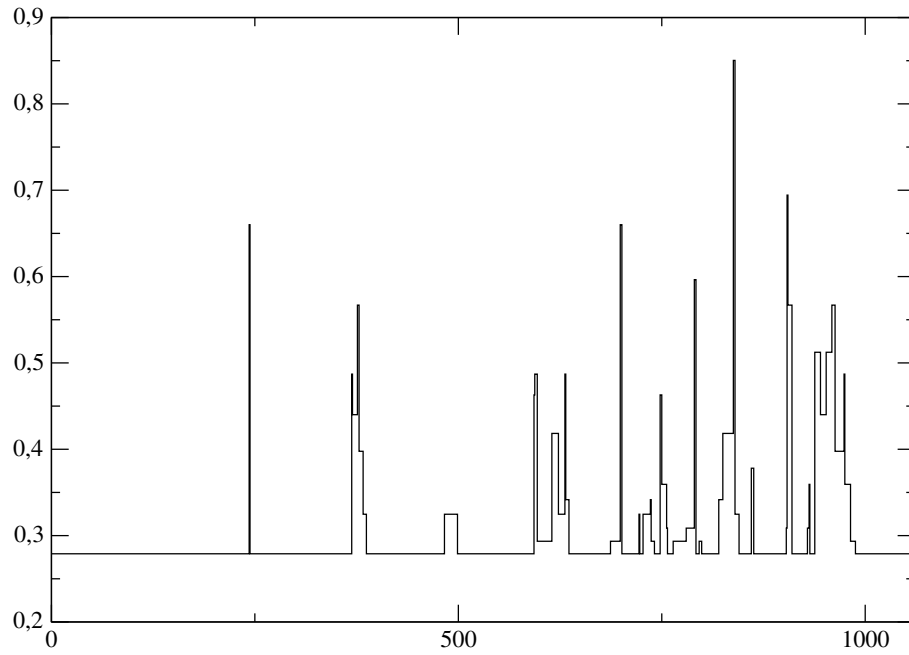


Figure 3: Number of times the word **RSS** appears in a text stream vs. time. This figure shows the bursty structure of this kind of streams.

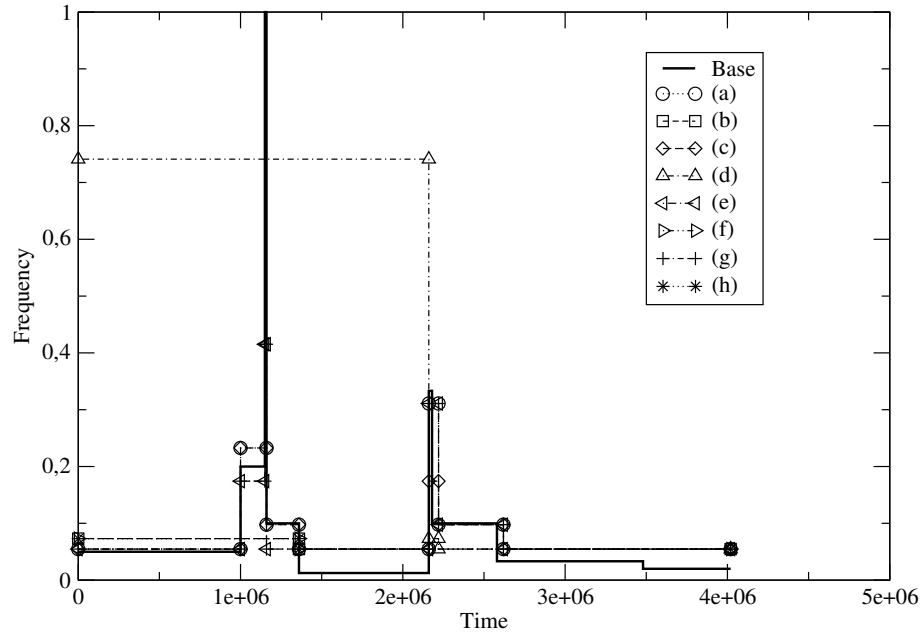


Figure 4: Fitting different measures of proximity to a text stream that uses the same representation as in figure 3. In this case, distance labelled as (g) yields the best result; this will be the one that is finally used.

2.5 Bankruptcy prediction

Some advances have been made in the selection of variables involved in the prediction of bankruptcy of industrial firms. The Granada-EFC group has made some headway in that area, and found out that indebtedness, funding costs and economic return on investment are the best predictors of firm health. For the time being, no adaptive algorithms are being used; in a future stage, these problems will be approached using the DREAM/JEO platform.

3 Results indicators

There are two persons that have been hired for the project, Juan Luis Jiménez Laredo and Eva Alfato Cid; Juan Luis is currently engaged in the PhD program at the University of Granada. Additionally, two persons, Juan José Samper, whose PhD is in an advanced state, and Juan Torrecillas, who has also signed up for the PhD program of the ATC dept. at the University of Granada, have profited from their interaction with members of the project.

A website has been created for diffusion of the project results, , and is maintained by Juan Luis at the Granada-ATC subproject.

Several papers and technical reports have been published by each subproject:

- ITI-UPV has published the following technical reports [Cid, 2005d, Cid, 2005c, Cid, 2005b, Cid, 2005a, Cid, 2004a, Cid, 2004b], mainly on the implementation of a signal classification system using DREAM and JEO; all of them are in Spanish. Besides, a paper has been accepted in a national congress: [Alfaro-Cid et al., 2005].
- The Granada (which includes a professor from Madrid) subproject has published the following papers in a national multidisciplinary congress: [Romero, 2005b, Arenas, 2005, Tricas, 2005, Romero, 2005a]. Two have been published in an international congress: [Gustavo Romero and Mereño, 2005, Castillo et al., 2005]. Several papers have been submitted to international journals, and are in the review phase.
- The EFC group has published a paper in an international congress [Gómez et al., 2005], and another one submitted to a national accounting congress [De la Torre et al., 2005]

Finally, some groups, specially Granada-ATC, maintain links with several other universities: Zaragoza, Málaga, Rey Juan Carlos I in Madrid, but no serious attempt to initiate an international project has been made. We intend to release software as free software as soon as it is well-packaged and we can write a good roadmap for the future development of the platform.

References

- [Alfaro-Cid et al., 2005] Alfaro-Cid, E., Esparcia-Alcázar, A., and Sharman, K. (2005). Clasificación de señales de electroencefalograma usando programación genética. In *Proceedings of the IV Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados MAEB'05*, Granada, Spain.

- [Arenas et al., 2002a] Arenas, Foucart, Schoenauer, and Merelo (2002a). Computación evolutiva en Java: JEO. pages 46–53. Universidad de Extremadura.
- [Arenas et al., 2001] Arenas, M., Foucart, L., Merelo-Guervós, J.-J., and Castillo, P. A. (2001). JEO: a framework for Evolving Objects in Java. In *Actas Jornadas de Paralelismo*. Universidad Politécnica de Valencia.
- [Arenas et al., 2004] Arenas, M. G., Castillo, P. A., Merelo, J. J., Rateb, F., and V. Rivas, G. R. (2004). Un nuevo enfoque evolutivo: Sistema Multi-ACU. pages 396–403.
- [Arenas et al., 2002b] Arenas, M. G., Dolin, B., Merelo-Guervós, J.-J., Castillo, P. A., de Viana, I. F., and Schoenauer, M. (2002b). JEO: Java Evolving Objects. In *Poster Accepted at GECCO 2002*. Available from <http://geneura.ugr.es/pub/papers/MPP104.ps.gz>, CiteSeer context.
- [Arenas, 2005] Arenas, V. M. M. J. J. M. G. J. L. J. L. M. G. (2005). Algoritmos evolutivos en java: resolución del TSP usando DREAM. In *Actas XVI Jornadas de Paralelismo, incluido en CEDI'2005. Granada, septiembre 2005. En prensa, actas editadas por Thompson*.
- [Castillo et al., 2005] Castillo, P., Romero, G., Arenas, M., Merelo, J., and Prieto, A. (2005). Designing a control system for an autonomous robot using an evolutionary algorithm. In [IWANN 2005, 2005].
- [Cid, 2004a] Cid, E. A. (2004a). Aumento de la velocidad de convergencia usando un sistema distribuido (dream). informe técnico iti-sac-004. ITI-SAC-REPORT ITI-SAC-004, Instituto Tecnológico de Informática.
- [Cid, 2004b] Cid, E. A. (2004b). Implementación de un clasificador de señales usando programación genética. informe técnico iti-sac-003. ITI-SAC-REPORT ITI-SAC-003, Instituto Tecnológico de Informática.
- [Cid, 2005a] Cid, E. A. (2005a). Clasificación de señales de electroencefalograma usando programación genética. informe técnico iti-sac-006. ITI-SAC-REPORT ITI-SAC-006, Instituto Tecnológico de Informática.
- [Cid, 2005b] Cid, E. A. (2005b). Implementación de fdas en jeo y su aplicación al problema de clasificación de señales de electroencefalograma. informe técnico iti-sac-007. ITI-SAC-REPORT ITI-SAC-007, Instituto Tecnológico de Informática.
- [Cid, 2005c] Cid, E. A. (2005c). Implementación de una técnica de muestreo aleatorio para reducir la sobreadaptación. informe técnico iti-sac-008. ITI-SAC-REPORT ITI-SAC-008, Instituto Tecnológico de Informática.
- [Cid, 2005d] Cid, E. A. (2005d). Programación genética con nodos de aprendizaje. informe técnico iti-sac-009. ITI-SAC-REPORT ITI-SAC-009, Instituto Tecnológico de Informática.
- [De la Torre et al., 2005] De la Torre, J., Gómez, M., and Román, I. (2005). Análisis de sensibilidad temporal en los modelos de predicción de insolvencia: una aplicación a las pymes industriales. enviado al XIII Congreso de la Asociación Española de Contabilidad y Administración de Empresas, a celebrar en Oviedo en septiembre de 2005 y a la Revista Española de Financiación y Contabilidad.

- [Gustavo Romero and Merelo, 2005] Gustavo Romero, M.G. Arenas, P. C. and Merelo, J. (2005). Evolutionary design of a brain-computer interface. In [IWANN 2005, 2005], page 653.
- [Gómez et al., 2005] Gómez, M. E., Román, I., and De la Torre, J. M. (2005). The analysis of the economic risk in small and medium-sized companies of the trade sector. In *28th Annual European Accounting Association Congress, Goteborg*.
- [IWANN 2005, 2005] IWANN 2005 (2005). *8th International Work-Conference on Artificial and Natural Neural Networks, IWANN 2003*, Lecture Notes in Computer Science, LNCS. Springer-Verlag.
- [JXTA,] JXTA. Project JXTA web site www.jxta.org.
- [Kleinberg, 2002] Kleinberg, J. (2002). Bursty and hierarchical structure in streams. In *Proc. 8th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*, pages 91–101. ACM.
- [Paechter. et al., 2000] Paechter., B., Bäck, T., Schoenauer, M., Sebag, M., Eiben, A. E., Merelo, J. J., and Fogarty, T. C. (2000). A distributed resource evolutionary algorithm machine (DREAM). In *Evolutionary Computation, 2000. Proceedings of the 2000 Congress on*, volume 2, pages 951–958. IEEE, IEEE Press. Available from <http://ieeexplore.ieee.org/iel5/6997/18853/00870746.pdf>.
- [Riolo and Worzel, 2003] Riolo, R. L. and Worzel, B. (2003). *Genetic Programming Theory and Practice*. Genetic Programming Series. Kluwer, Boston, MA, USA. Series Editor - John Koza.
- [Romero, 2005a] Romero, M. G. A. P. C. J. J. M. G. (2005a). Coevolución de perceptrones multicapa y conjuntos de entrenamiento. In *Actas Simposio Metaheurísticas, Algoritmos Evolutivos y Bioinspirados, MAEB2005, incluido en el CEDI2005. En prensa, actas editadas por Thompson*.
- [Romero, 2005b] Romero, M. G. A. P. C. J. J. M. G. (2005b). Implementación de un sistema complejo clásico utilizando una librería evolutiva distribuida. In *Actas Simposio sobre Sistemas Complejos, SISCO, incluido en CEDI'2005. Granada, septiembre 2005. En prensa, actas editadas por Thompson*.
- [Tricas, 2005] Tricas, J. J. M. F. (2005). Integración de wikis en una clase de informática: un paso más en la creación de comunidad. In *Actas VI Congreso Nacional de Informática Educativa, Simposio Nacional de Tecnologías de la Información y las Comunicaciones en la Educación, SINTICE'2005 (ADIE). En prensa, actas editadas por Thompson*.

IDE-E: Technological advances for a Web-based National Spatial Data Infrastructure, convergence with the European initiative INSPIRE

TIC2003-09365-C02

Pedro R. Muro-Medrano¹
Depto. de Informática e Ing. de Sistemas
Universidad de Zaragoza

Miguel Angel Bernabé Poveda²
Depto. Ing. Topográfica y Cartografía
Universidad Politécnica de Madrid

Michael Gould Carlson³
Depto. de Lenguajes y Sistemas Informáticos
Universitat Jaume I

Abstract

INSPIRE is an initiative for a European directive to create a Spatial Data Infrastructure (SDI) at the European level, its objective is to facilitate the access to spatial data, information which is recognized of vital importance mainly in the environmental sector. The development of such SDI involves many technological problems mainly related with interoperable geoprocessing and related areas. This project, which is the continuation of a previous project, joins together the efforts of the research groups working in this area to provide inside and gain experience. The result indicators show that the activities of the R&D groups involved in the project have contributed in a significant way to the advance of the technology and the implantation of the Spatial Data Infrastructure of Spain.

Keywords: OGC, Java, Catalogs, Catalog Services, Metadata, Thesaurus, Geographic Information, GIS, Web Services, Geoprocessing Services, Web Map Servers, Interoperability, Distributed Objects, Spatial Data Infrastructures, Open Standards

¹ Email: prmuro@unizar.es

² Email: m.a.bernabe@terra.es

³ Email: gould@lsi.uji.es

1. **Project objectives**

The project has a goal of technological advancement toward the construction of a Spatial Data Infrastructure (SDI) in Spain making possible the development of national technology in this field. SDI can be seen as a tool to facilitate the access and exploitation of key spatial information for a long list of applications, mainly in the public sector. From the technological point of view, a SDI consists of reference geographic data, metadata, and diverse web services (standard and interoperable), especially access via catalogs, that allow that users (scientists, public sector technicians, general public) are able to locate and access these data and their value-added products. Many challenging technological problems arise to deal with SDI in areas such as distributed geoprocessing, technological and semantic interoperability, open geoprocessing services and standards, web services and architectures, information recovery, metadata, geodata models, geodata harmonization, ... A previous project served to overcome the previous lack of SDI development at national level and produced technology which has been successfully put into practice in several projects which are operational. The technological position we have reached with the previous project imposes us now to follow the research a variety of aspects which may and must be improved, as much for the inclusion of new ideas, as to solve problems found and to take advantage of new possibilities open by other technologies.

A key development now underscores the need for advancement in this line of R+D: in November 2001 the European Commission launched INSPIRE (INfrastructure for SPatial InfoRmation in Europe), an initiative to create European legislation to guide national and regional SDI development, initiative that is at the co-decision process (Commission/Parliament) and is expected to become directive by 2.006. This initiative, sponsored at the highest levels within the EC, (see Memorandum of Understanding for the cooperation between Commission services Directorate General of Environment, EUROSTAT and Directorate General of the Join Research Center), will mandate how and when each member state should create its national SDI. We must also add at this time, the convergence between different standardization organizations in the field of geographic information (OGC, ISO, CEN, FGDC, ...), institutions (GSDI, EUROGI, AGILE, JRC, ...) and European initiatives (GETIS, GINIE, INSPIRE, ...). Inside the Spanish framework, AENOR is trying to converge geographic standards, and the Consejo Superior Geográfico (Commission of Geomatics) which is in charge of the development of the National Spatial Data Infrastructure in Spain. This convergence of interests allows functionalities which make profit of the new communication technologies (mainly through the web) and demands important advances in the technological and also semantic interoperability in the domain of spatial information. Although it has been a large advance in the software technology for the manipulation of geographic information, the creation of this new infrastructures and the conformity with new standards needs the development of new technology.

Introduction to INSPIRE

Good policy depends on high-quality information and informed public participation. Policymakers have recognized the growing interconnection and complexity of the issues affecting the quality of life today, and this recognition is influencing the way new policies are now being formulated. A new approach is therefore needed to deal with monitoring and reporting and with data management and delivery across the different levels of government. Policies need to be employed to reduce duplicated data collection and to assist and promote the harmonisation, broad dissemination and use of data. Such policies should result in increased efficiency, the benefits of which can be reinvested in improving the availability and quality of information. In turn, the increased availability of information will stimulate innovation among information providers in the commercial sector.

Spatial information can play a special role in this new approach because it allows information to be integrated from a variety of disciplines for a variety of uses. A coherent and widely accessible spatial description of the Community territory would deliver the requisite framework for coordinating information delivery and monitoring across the Community. Spatial information may also be used to produce maps, which are a good way of communicating with the public. Unfortunately, the technical and socio-economic characteristics of spatial information make the problems of coordination, information gaps, undefined quality and barriers to accessing and using the information particularly acute.

The Commission has therefore decided to submit to the European Parliament and the Council of the European Union the INSPIRE proposal to make interoperable spatial information readily available in support of both national and Community policy and to enable the public to access to this information. This initiative derives from the commitment of several Commission services in particular DG Environment, Eurostat and the Joint Research Centre.

The proposed Directive creates a legal framework for the establishment and operation of an Infrastructure for Spatial Information in Europe, for the purpose of formulating, implementing, monitoring and evaluating Community policies at all levels and providing public information. A key objective of INSPIRE is to make more and better spatial data available for Community policy-making and implementation of Community policies in the Member States at all levels. INSPIRE focuses on environmental policy but is open for use by and future extension to other sectors such as agriculture, transport and energy. The proposal focuses specifically on information needed in order to monitor and improve the state of the environment, including air, water, soil and the natural landscape. Much of this information needs to be underpinned by "multi-purpose" spatial data. INSPIRE is designed to optimize the scope for exploiting the data that are already available, by requiring the documentation of existing spatial data, the implementation of services aimed at rendering the spatial data more accessible and interoperable and by dealing with obstacles to the use of the spatial data.

Objectives of the project

For the technology to be developed to be truly useful, it should follow international standards and be developed using the latest innovative methodologies; for this reason this project proposes to continue following the progress of (and participation in) these INSPIRE-related initiatives, especially in areas such as metadata, information retrieval, multilingual access, interoperability and the semantic web. Within the realm of technology creation the project will follow and develop innovative architectural frameworks that are optimal for web-based systems (web services) and will advance in the area of metadata creation and management. Certain key components (services) for SDIs will be implemented or improved (OGC specifications: Catalog, WFS, WCS, WTS, SLD y Gazetteers) with a focus on their interoperability with 3rd party services. Also to be explored are new manners of implementing service chaining of heterogeneous web services, access to geographic information from wireless terminals, and e-Commerce aspects of geographic information delivery. The development will focus on two areas: 1) in the area of SDI, specific functionality will be built to support needs of the administration, with emphasis on access via catalogs, and 2) in the area of the environment, the initial focus area of INSPIRE, a wide range of software tools will be created to support the needs of the collaborating public organizations, mostly from the environmental sector. This technology will be linked to information content and demonstrated within a pilot project, allowing access to key geographic information by participating administrations and businesses. Finally, the project continues on-going labors of dissemination and technology and knowledge transfer to administrations and businesses, to raise awareness of the new possibilities afforded by SDIs.

To reach that we propose to split in the following subobjectives:

1. Tracking of standardization processes and technology watching. For the technology to be useful, it is necessary its conformity with international standards and make use of the new technological possibilities, to do that a tracking, including the participation, of the standardization processes and technology watching in topics such as metadata, modeling, advanced access and information recovery services, multilingual access, web services, interoperability or semantic web.
2. Technological improvements in the architecture and basic components of metadata and catalogue. Specialized architectural patterns will be proposed for this type of computation in web and advances will be made in the software infrastructures to facilitate the creation and management of metadata and catalogue services of data sets and services catalogue.
3. Tecnolgogical improvements in software components of visualization and access services. Technological and functional improvements will be made, or will be developed if needed, in middleware software components in the SDI sector to support web services for WMS, WFS, WCS, SLD and Gazetteers.
4. Tentative exploitation of new possibilities such as service chaining, wireless equipment and e-commerce of geographic information services.
5. Creation of specific application software in two sectors: public catalogue services at the national level and environmental information

6. Put in practice the technological and functional improvements in a pilot project to prove and demonstrate the technology to interested institutions and companies.
7. Diffusion, awareness and technology transfer.

The means needed to get these objectives are mainly related with:

- human resources, in addition to the research personal technical people is needed to be able to deal with important technical work
- computer infrastructures (servers and personal computers, and networks)
- software, some commercial (oracle, ArcGis, ...) some self made

The work to achieve the previously mentioned objectives is structured into seven activities according to the following timetable:

Activity	First year	Second year	Third year
1. Project management and coordination			
2. Standards and technology tracking			
3. Metadata and catalogue improvement			
4. Web services development			
5. Technological and functional exploitation			
6. Pilot Project			
7. Diffusion and technology transfer			

2. Success level reached in the project

So far we are satisfied with the success level reached in the project.

- Metadata. Advances have been done in aggregation and derivation concepts, crosswalks for interoperability between metadata standards, the use of disambiguated thesauri to improve information retrieval. A book recently published by Springer and written by members of the projects is mostly devoted to these topics. A new ontology service has been proposed and developed. Some work has also been done with multilingual thesaurus and supporting tools. Specialized metadata profiles have been also proposed for Dublin core, inspire and the water framework directive.
- Metadata tool. The tool has been improved with several new features and provided as open source. The tool has been selected as the reference metadata tool by the Spanish SDI, Poland SDI and have been contracted by EUROSTAT. At this moment we have collaboration agreements for multilingual support in 9 languages.

- Catalogue services. Much work has been done in at different areas such as indexing (textual and spatial), incorporation of aggregations, support for OGC standards, multilingual support, multi metadata profile support (using DC as a kernel) for the incorporation of other types of resources.
- Other web geodata services. Improvements has been done in several web servers: Map Server, Gazetteer server, route server
- Pilot project. The pilot project that has been established consists in the development of a Spatial Data Infrastructure (SDI) to support access to geographic information resources concerned with the Water Framework Directive within an inter-administration and cross-border scenario that involves: two countries, France and Spain; and, the two main river basin districts at both sides of the border, the Adour-Garonne basin district, managed by the Water Agency for the Adour-Garonne River Basins (“L’Agence de l’Eau Adour-Garonne”) and the Ebro river basin district, managed by the Ebro River Basin Authority (“Confederación Hidrográfica del Ebro”). This pilot is more challenging than the one originally planned, and is co financed by the European Commission through EUROSTAT.
- Standardization. We continue our participation on standardization processes and as a result of this experience we have been working with the OGC (Open Geospatial Consortium) and we have been collaborating in the definition of the metadata profile which has been adopted for the Spanish spatial data infrastructure as well as with the Spanish gazetteer model recently adopted.
- Technology transfer. This is a very important part of our research strategy and it is our feedback mechanism and a objective demonstration of our results and, as it can be seen in section 3, we are already having clear success on that.
- Diffusion and knowledge transfer. We have already been participating in many actions as conferences or web portal. Several related courses has been designed at the general, bachelor and PhD levels (see section 3 for details). We have organized in Zaragoza the first workshop of the Spanish Spatial Data Infrastructure (280 participants) and we will be also organizing the second by the Madrid subproject. We also have an important presence in the Working Group for the Development of the Spanish Spatial Data Infrastructure organized by the Comisión de Geomática of the Consejo Superior Geográfico.

Just one, but important, difficulty has arisen in the development of the project which has been the low budget awarded. We have received 6 times less than the budget in the proposal (several demands to improve this situation were not accepted, neither renunciations). As a way of comparison, a previous subproject with 5 ejc (equivalent to full time persons) was awarded with about 132.000 € whereas this one with 23 (13,5 researchers and 10 others) ejc has been awarded with 72.000 € for the 3 years. As it can be easily understood this limitation could have produced a significant drawback for the project. The solution so far has been to use money from other sources.

3. Result indicators

The research work undertaken in this project became reality in a variety of ways as techniques, processes, methodologies, models, architectures, concepts, software components, systems, papers, projects, courses, ... We will focus below in some objective indicators:

3.1. PhD Thesis

Three PhD thesis have been finished during this project, two of them awarded with the PhD extraordinary prize, the other one pending (just defended last July 21th).

1. *Contributions on the problem of collections, interoperability and information retrieval of geographic information metadata for Spatial Data Infrastructures*. PhD student: Javier Nogueras Iso. PhD advisors: Pedro R. Muro Medrano y F. Javier Zarazaga Soria. PhD program: Systems Engineering and Computer Science. Title: PhD in Computer Science, University: University of Zaragoza, Calification: Pass Cum Laude and PhD extraordinary prize of the University of Zaragoza. Date: September 1, 2004
2. *Arquitecturas y patrones de diseño para el modelado e integración de sistemas de información heterogéneos. Una aproximación aplicada al contexto de los servicios basados en la localización*. PhD student: Pedro Javier Álvarez Pérez-Aradros. PhD advisors: Pedro R. Muro Medrano y Jose A. Bañares Bañares. PhD program: Systems Engineering and Computer Science. Title: PhD in Computer Science . University: University of Zaragoza. Calification: Pass Cum Laude and PhD extraordinary prize of the University of Zaragoza. Date: March 8, 2004
3. *Contributions to the Generation and Representation of Digital Terrain Models*. PhD student: José Poveda Albadalejo. PhD advisor: Michael Gould Carlson. PhD program: Advanced Informatic Methods. Title: PhD in Computer Science. University: University Jaume I. Calification: Pass Cum Laude. Date: July 21, 2005

3.2. Publications

The following list of books and papers present results directly related with this project (see bibliography for full references).

3.2.1. Books

[1]

3.2.2. **Journals and book chapters**

[2],[3],[4],[5],[6],[7],[8],[9],[10],[11]

3.2.3. **Conference contributions**

[12],[13],[14],[15],[16],[17],[18],[19],[20],[21],[22],[23],[24],[25],[26],[27],[28],[29],[30],
[31],[32],[33],[34],[35],[36],[37],[38],[39],[40],[41],[42],[43],[44],[45],[46],[47],[48],[49],[50]

3.3. **Projects**

All the research and development projects we are involved are related with the topics in project TIC2003-09365-C02, we list above live projects during the life time of project TIC2003-09365-C02 so far.

3.3.1. **European R&D Projects**

- 1) AWARE: A tool for monitoring and forecasting available water resources in mountain environment (FP6-Space-1, proposal 12257). Sponsored by DG H. Transport and Space. European Union. University Jaume I. 2005-2007.
- 2) SDIGER: A cross-border inter-administration Spatial Data Infrastructure to support WFD information access for Adour-Garonne and Ebro River Basins. Consulting and Technical Assistance Contract. Sponsored by EUROSTAT, Invitation to Tender (2004/S 111-092104/EN), Pilot projects on the implementation of the Infrastructure for Spatial Information in Europe (INSPIRE). Participant entities: IGN (France), IGN FI, University of Zaragoza (coordination of Spanish part), CNIG. 2004-2006.
- 3) ACE-GIS: Adaptable and Composable E-commerce and Geographic Information Services. (Ref. IST-2001-37724). Sponsored by IST Program, European Union. Norwegian Mapping Agency and University Jaume I 2002-2004.
- 4) TOWNTOLGY: Urban Ontologies for an improved communication in Urban Civil Engineering Projects. European Cooperation in the field of Scientific and Technical Research - COST – Secretariat. European Concerted Research Action COST Action C21. 2005-2008
- 5) EUSC-IDSS Symbol schema concept and symbol design and test of symbol schema implementation. Consulting and Technical Assistance Contract. Sponsored by European Union Satellite Centre (EUSC) subcontracted by Open GIS Consortium (Europe) Limited (OGCE) (Ref. EUSC-RF-Ph1+2-PMM). 2004.
- 6) ESA Harmony. Consultancy Agreement. Sponsored by ESA (European Space Agency). University of Zaragoza subcontracted by OGCE (Open Geospatial Consortium Europe) (Nr. ESA Harmony-PMM-2005-01, OGCE Job Code “ESA Harmony 20013”). 2005.

- 7) Collaboration Agreement in the area of the Spatial Data Infrastructures (SDI). Collaboration Agreement. Sponsored by Environment and Sustainability Institute of the Joint Research Centre (JRC) of the European Community and University of Zaragoza. 2005-2010.

3.3.2. National projects and R&D activities sponsored by public calls

- 1) On line access to the Resources Catalogue of Geographic Information of the University of Zaragoza. Sponsored by Aragón Government, Department of Science, Technology and Universities, Directorate General of Technologies for the Information Society. Ref. 52/2004. 2004.
- 2) Development of digital contents of historic artistic buildings of Zaragoza and publication in Internet. Sponsored by Aragón Government, Department of Science, Technology and Universities, Directorate General of Technologies for the Information Society. Ref. 158/2004. 2004.
- 3) Mobil access to field service module with geographic localization. Sponsored by Ministry of Industry, Commerce and Tourism, project Profit FIT-350100-2004-367 (National Program of Technologies of Services for the Information Society) Participant entities: Insynergy Consulting España, S.L. and University of Zaragoza. 2004.
- 4) IDE-E: Technological advances for a National Spatial Data Infrastructure based on Web, convergence with the European initiative INSPIRE. Sponsored by Ministry of Science and Technology, coordinate project, ref.: TIC2003-09365-C02-01 Participant entities: Polytechnic University of Madrid, coordination: University of Zaragoza. 2003-2005.
- 5) Implementation of Environmental data infrastructures (IMD): Prototype of application of geographical information systems to hydric resources management. Sponsored by Fundació Bancaixa Castelló. UJI. 2002-2005.
- 6) Virtual interaction technologies with set of stocked objects of high cardinal. Application to container depot and storage centres. Sponsored by MCYT. University Jaume I. 2002-2005.

3.3.3. Projects for management of R&D activities sponsored by public calls

- 1) Consolidated research group – Advanced Information Systems Lab. Sponsored by Aragón Government, Department of Education and Science. Directorate General of High Education (attendance to support Operative Units of Research of the Autonomous Community of Aragón ref T56) 2005.

- 2) Creación de un Laboratorio de Tecnologías de la Información Espacial. Sponsored by Dirección General del Instituto Geográfico Nacional, Ministerio de Fomento, UPM. 2005 (renovable anualmente). 144.000 Euro/año.
- 3) Thematic Network of Geographic Information Technologies. Sponsored by Ministry of Education and Science, Complementary Actions in the framework of some National Programs of the National Plan of Scientific Research, Development and Technological Innovation 2004-2007 (ref. TSI2004-20019-E). Participant entities: National Geographic Institute; and the Universities Castilla la Mancha, A Coruña, Jaume I, Polytechnics of Madrid, Santiago de Compostela, Sevilla and Zaragoza. Coordinated by the University of Zaragoza. 2005
- 4) Emergent research group – Advanced Information Systems Lab. Sponsored by Aragón Government, Department of Education and Science. Directorate General of High Education (attendance to support Operative Units of Research of the Autonomous Community of Aragón) 2004.

3.3.4. Projects and R&D activities with companies and/or administrations

- 1) Trophy state evaluation of main water reservoir of the Ebro basin by spatial teledetection techniques. Technical Assistance Contract. Sponsored by Ebro River Basin Authority (Ref. 2005-PH-18-I). 2005
- 2) Report server system improvement and extension. Technical Assistance Contract. Sponsored by Ebro River Basin Authority (Ref. CHE 2004-PH-40-I). 2005
- 3) IPA Data Base migration, improvement and exploitation. Technical Assistance Contract. Sponsored by Ebro River Basin Authority (Ref. CHE 2004-PH-40-I). 2005.
- 4) Optimization and extension of the Spanish Spatial Data Infrastructure Portal and the National Geographic Institute clearing house of data and geographic services. Collaboration Agreement. National Geographic Institute, Ministry of Public Works. 2005.
- 5) Complementary development of technology and methodology of metadata capture of geographic information of the Spatial Data Infrastructure of Spain and the node of the National Geographic Institute clearing house of data and geographic services. Collaboration Agreement. National Geographic Institute, Ministry of Public Works. University of Zaragoza y Polytechnic University of Madrid. 2005.
- 6) Armonización planimétrica de ortofotos con su cartografía catastral. Sponsored by Dirección General del Catastro, Ministerio de Hacienda, UPM. 2004-2005 (prorrogable por años). 68.000 €.
- 7) Planimetric harmonization of orthophotos with their cadastral cartography. Sponsored by Dirección General del Catastro, Ministerio de Hacienda. University Polytechnic of Madrid. 2004-2005. 68.000 €/year

- 8) Open software business strategic study. Sponsored by IVER Tecnologías de la Información, S.A. (Valencia). University Jaume I. 2004-2005.
- 9) Participation and support software for the Project 04/160 Implementation services of an open server of catalogues. Consulting and Technical Assistance Contract. Sponsored by Generalitat Valenciana. University of Zaragoza subcontracted by the Jaume I University of Castellón. 2004-2005.
- 10) Services towards the implementation of a Catalog Open Server. Sponsored by Conselleria de Infraestructuras y Transporte, Generalitat Valenciana. University Jaume I. 2004-2005.
- 11) Definition and implementation of a protocol of creation of metadata as support of a spatial data infrastructure for the Conselleria of Infrastructure and transport. Sponsored by Conselleria de Infraestructuras y Transporte, Generalitat Valenciana (Ref.:CONMEN/2004/01). University Jaume I. 2004.
- 12) Two conferences held in the International Geocycle of Conferences. Sponsored by AENA. University Polytechnic of Madrid. 2004. 6000€.
- 13) Technical assistance for the preparation of an analysis for the establishment of a tourism model based on new technologies. Consulting and Technical Assistance Contract. Sponsored by Aragón Government, Directorate General of Technologies for the Information Society. Digital Cities Program. 2004.
- 14) Viability study for the establishment of an SDI node for the Ebro River Basin Authority. Consulting and Technical Assistance Contract. Sponsored by Ebro River Basin Authority, (Ref. CHE 2004-PH-40-I) 2004.
- 15) Creation of a Spatial Information Technology Laboratory. National Geographic Institute, Ministry of Public Works. University Polytechnic of Madrid. 2004. 114.000€/year.
- 16) Implementation of query, quality control and data visualization functionalities in the Quimipa software application. Consulting and Technical Assistance Contract. Sponsored by Ebro River Basin Authority, (Ref. 2004-PH-35-I). 2004-2004.
- 17) Architectural, interoperability, modelling aspects and other R&D topics of information technologies in the software development for the integral management of hospitals and geriatric centres. Consulting and Technical Assistance Contract. Sponsored by Savac Consultores S.L. 2004-2006.
- 18) Integration of geographic information for the management of financial aids of the Communitarian Agrarian Policy. Consulting and Technical Assistance Contract. Sponsored by Sociedad de Infraestructuras Rurales Aragonesas, S.A., SIRASA, for the Diputación General de Aragón, Department of Agriculture. 2004-2005.
- 19) Storage, verification and assurance of the geometric correspondence of aerial images of Aragón, Phase II. Consulting and Technical Assistance Contract. Diputación General de Aragón, Department of Agriculture. 2004.

- 20) Support Works to attend the requirements derived from the Water Framework Directive in the topics of Hydrologic Planning (Ebro River Basin). Consulting and Technical Assistance Contract. Sponsored by Ministerio de Medio Ambiente, Ebro River Basin Authority. University of Zaragoza subcontracted by Zeta-Amaltea. 2004-2006.
- 21) Search applications with NEM metadata for geographic information catalogue services for the SDI-GV. Consulting and Technical Assistance Contract. Sponsored by Generalitat Valenciana, Consellería of Infrastructures and Transport and Valencia Cartographic Institute. 2004.
- 22) Cartographic information visualization tool for aqueduct. Consulting and Technical Assistance Contract. Sponsored by Berta Bescós Roy. 2004.
- 23) Geographic Information Catalogue Services for the SDI-GV. Consulting and Technical Assistance Contract. Sponsored by Generalitat Valenciana, Consellería of Infrastructures and Transport and Valencia Cartographic Institute. 2004.
- 24) Spatial Calculus application for windmills. R&D contract. Sponsored by Iberinco. 2004.
- 25) Analysis of the information systems of the Area of Water Quality and performance proposal. Consulting and Technical Assistance Contract. Sponsored by CHE, Comisaría, Area of Water Quality. 2004.
- 26) Evaluation of the modellization of the Integral Management of a Hydrographic Basin, HIDROSIS. Consulting Contract. Sponsored by AENOR. 2004.
- 27) Cartographic and plan information visualization tool for the Nerja Almuñecar highway. Consulting and Technical Assistance Contract. Sponsored by Berta Bescós Roy. 2004.
- 28) Interoperable software components for a route and map server in a web and PDA environment. R&D contract. Sponsored by RACC, subcontracted by Net2U. 2004.
- 29) Design of the evolution towards a Spatial Data Infrastructure in the Zaragoza local area, IDEZar. Collaboration Agreement. Sponsored by Zaragoza Council, Concejalía of Science and Technology. 2003-2004.
- 30) Software environment for the storage of Hydrogeochemistry information. The Quimipa application. Consulting and Technical Assistance Contract. Sponsored by Ebro River Basin Authority, Hydrographic Planning Office. 2003-2004
- 31) Actualization and improvements in the IPA information system. Consulting and Technical Assistance Contract. Sponsored by Ebro River Basin Authority. 2003-2004.
- 32) Definition and implementation of mechanisms for the Internet remote access to IPA. Consulting and Technical Assistance Contract. Sponsored by Ebro River Basin Authority. 2003-2004.
- 33) Creation of the web portal of the Spanish Spatial Data Infrastructure and the node of the National Geographic Institute for distribution and intermediation of geographic data and

services. Collaboration Agreement. Sponsored by National Geographic Institute, Ministry of Infrastructures. 2003-2005.

- 34) Creation of web services for route optimization based in open standards for the RACC portal. Consulting and Technical Assistance Contract. Sponsored by Net2U. 2003-2004.
- 35) Development of the Environment and Geographic Information Network of Galicia. Cooperation Agreement. Xunta de Galicia, Conselleria of Environment. Entities participants: Teide Consortium (University of Zaragoza, Jaume I University, Polytechnic University of Madrid), responsible University of Zaragoza. 2003-2004.

3.4. Education and knowledge transfer

Only activities sponsored by public calls.

- 1) eduGILA. Sponsored by ALFA program, European Union. University Jaume I. 2003-2004.
- 2) Spatial Data Infrastructure of Spain (IDE-E). Course for the installation of Web Map Servers (WMS) and Object Servers (WFS), in conformance with OGC specifications using Open Source. Sponsored by Ministerio de Fomento. 2004. 8.000 €
- 3) VideoConference infrastructure to give classes of Doctorate courses: Geoinformation for Government and Society. Sponsored by UPM. 2005. 4.500 €
- 4) Establishment of a PhD join program on Geoinformation for Government and Society. Sponsored by AECL. 2005. 12.000 €
- 5) LIFE-TIERMES. Support to IDE creation and courses about archeologic topography. Sponsored by LIFE project, Tiermes Museum Friends Association. University Polytechnic of Madrid. 2005. 12.000€.

3.5. Diffusion

Only activities sponsored by public calls.

- 1) Workshop of the Spatial Data Infrastructure of Spain. Sponsored by the University of Zaragoza with financial support for the organization of Scientific Congress and Meetings. 2005.
- 2) Workshop of the Spatial Data Infrastructure of Spain. Sponsored by Ministry of Infrastructures (Nº Exp. 33/04. 2004). 2004

- 3) JTIDEE04. Workshop of the Spatial Data Infrastructure of Spain. Sponsored by Aragón Government, Department of Science, Technology and Universities, Directorate General of Technologies for the Information Society. Ref. 157/2004. 2004.
- 4) Producción y distribución de un Newsletter sobre Infraestructuras de Datos Espaciales (IDE) en español complementado con difusión y entrega de materiales sobre IDE a las instituciones participantes. Sponsored by Global Spatial Data Infrastructure (GDSI). Small Grants Program, participants UNPSJB y UPM. 2005. 2.500 US\$
- 5) Organización de dos Conferencias dentro del Geo-Ciclo Internacional de Conferencias. Sponsored by AENA (Aeropuertos Españoles y Navegación Aérea). 2004. 6.000 €

Groups of the project have organized

- 1st Workshop of the Spatial Data Infrastructure of Spain hold in Zaragoza las November and is organizing the 2nd one that it will take place in Madrid next November.
- 1st Workshop on Geographic Information Management (Zaragoza, sept. 2004) and is organizing the 2nd that it will take place in Denmark next September.

3.6. Experts

3.6.1. INSPIREexperts

Another relevant indication of the expert knowledge got by the members of the project is that two of them have been selected as experts in the call for experts for the development of INSPIRE draft Implementing Rules (5 drafting teams of 6 members each have been selected from a total of 180 all along Europe -see <http://inspire.jrc.it/>-). An additional expert has also been selected in a support expert group (only 4 experts have been selected from Spain).

3.6.2. WG-IDEE

We are also participating as experts (the only in the working group for the Spatial Data Infrastructure of Spain (IDEE standing for Infraestructura de Datos Espaciales de España), a working group (of about 100 persons in representation of all institutions involved in the IDEE) established by the Comisión de Geomática del Consejo Superior Geográfico which is in charge of the development of the national SDI. Additionally we are part of the Comisión de Geomática in representation of Universities.

4. References

- [1] Javier Nogueras-Iso, F.J. Zarazaga-Soria, P.R. Muro-Medrano. *Geographic Information Metadata For Spatial Data Infrastructures - Resources, Interoperability, Information Retrieval*. Springer Verlag, Germany, ISBN: 3-540-24464-6, 264 pages, March 2005.
- [2] R. Béjar, J. Aboal, P. Vila, M. Gould y P.R. Muro-Medrano. *Las infraestructuras de datos espaciales en la gestión forestal. Una iniciativa de la Xunta de Galicia*. Montes volume 79. pages 26-34, 2005.
- [3] J.Nogueras-Iso, F.J.Zarazaga-Soria, R.Béjar, P.J. Álvarez, P.R.Muro-Medrano. *OGC Catalog Services: a key element for the development of Spatial Data Infrastructures*. Computers and Geosciences volume 31/2. pages 199-209, 2005.
- [4] M. Á. Latre, R. Miguel, P. R. Muro-Medrano. *Inventario de Puntos de Agua de la Ebro River Basin Authority*. Hidrogeología y Recursos Hidráulicos, chapter XXVII, pages 587-596. Asociación Española de Hidrogeólogos, (Editor) Antonio Fernández Uría, Zaragoza, 2005.
- [5] J. Nogueras-Iso, F.J. Zarazaga-Soria, J. Lacasta, R. Béjar, P.R. Muro-Medrano. *Metadata Standard Interoperability: Application in the Geographic Information Domain*. Computers, Environment and Urban Systems, volume 28/6, pages 611-634, November 2004.
- [6] F. J. Zarazaga-Soria, J. Nogueras-Iso, R. Béjar and P. R. Muro-Medrano. *Political aspects of Spatial Data Infrastructures*. Electronic Government, Lecture Notes in Computer Science, LNCS, volume 3183, pages 392-395, Springer-Verlag. Roland Traummüller, Germany, 2004.
- [7] J. Nogueras-Iso, M. Á. Latre, P. R. Muro-Medrano, F. J. Zarazaga-Soria. *Building e-Government services over Spatial Data Infrastructures*. Electronic Government, Lecture Notes in Computer Science, LNCS, volume 3183, pages 387-391, Springer-Verlag. Roland Traummüller, Germany, 2004.
- [8] J. Nogueras-Iso, J. Lacasta, J.A. Bañares, P.R. Muro-Medrano, F.J. Zarazaga-Soria. *Exploiting disambiguated thesauri for information retrieval in metadata catalogs*. Current Topics in Artificial Intelligence. Lecture Notes on Artificial Intelligence, LNAI, volume 3040, pages 322-333, Springer-Verlag. Roland Traummüller, Germany, 2004.
- [9] Poveda, J. y M. Gould. *Multidimensional binary indexing for neighborhood calculations in spatial partition trees*. Computers & Geosciences 31(1), pages 87-97. 2005
- [10] M. Gould (guest editor). *Geographic information science research in Europe: contributions from AGILE (monographic edition)* Computers & Geosciences 31(2), 2005
- [11] Poveda, J. and M. Gould. *The BQV algorithm for point-in-polygon determination*. GESTS International Transactions Computer Science and Engineering, 6(1), pages 207-218. 2005

- [12] Oscar Cantán, F. Javier Zarazaga-Soria. *Multi-Source Framework For Seamlessly Exploiting and Leveraging Disparate Spatial Data Catalogues*. Proceedings of the 11th EC GI & GIS Workshop, ESDI Setting the Framework , pages 91-95, Alghero, Sardinia (Italia), June 2005
- [13] D. Portolés-Rodríguez, P. Álvarez, P.R. Muro-Medrano. *IDEZar: an example of user needs, technological aspects and the institutional framework of a local SDI*. Proceedings of the 11th EC GI & GIS Workshop, ESDI Setting the Framework , pages 56-58, Alghero, Sardinia (Italia), June 2005
- [14] Joanna Nowak, Stephen Peedell, J. Nogueras-Iso. *Issues Of Multilinguality In Creating A European SDI – The Perspective For Spatial Data Interoperability*. Proceedings of the 11th EC GI & GIS Workshop, ESDI Setting the Framework , pages 47-48, Alghero, Sardinia (Italia), June 2005
- [15] Javier Lacasta, Pedro R. Muro-Medrano, Javier Nogueras-Iso, F. Javier Zarazaga-Soria. *Web Ontology Service, a key component of a Spatial Data Infrastructure*. Proceedings of the 11th EC GI & GIS Workshop, ESDI Setting the Framework , pages 19-22, Alghero, Sardinia (Italia), June 2005
- [16] M.A.Latre, F.J.Zarazaga-Soria, J.Nogueras-Iso, R. Béjar, P.R.Muro-Medrano. *SDIGER: A cross-border inter-administration SDI to support WFD information access for Adour-Garonne and Ebro River Basins*. Proceedings of the 11th EC GI & GIS Workshop, ESDI Setting the Framework , pages 5-7, Alghero, Sardinia (Italia), June 2005
- [17] José R. R. Viqueira, Pedro Álvarez, José Varela, Pedro J. Saco. *Architecture of a Natural Disasters Management Framework and its application to Risk Assessment*. Proceedings of the 8th AGILE Conference on Geographic Information Science (Estoril, Portugal). ISBN 972-8093-13-6, pages 529-538, May 2005
- [18] R. Tolosana-Calasan, D. Portolés-Rodríguez, J. Nogueras-Iso, P. R. Muro-Medrano, F.J. Zarazaga-Soria. *CatServer: a server of GATOS*. Proceedings of the 8th AGILE Conference on Geographic Information Science (Estoril, Portugal). ISBN 972-8093-13-6, pages 359-366, May 2005
- [19] R. Tolosana-Calasan, D. Portolés-Rodríguez, J. Nogueras-Iso, P. R. Muro-Medrano, F.J. Zarazaga-Soria. *CatServer: a server of GATOS*. Proceedings of the 8th AGILE Conference on Geographic Information Science (Estoril, Portugal). ISBN 972-8093-13-6, pages 359-366, May 2005
- [20] Lars Bernard, Alessandro Annoni, Ioannis Kanellopoulos, Michel Millot, Javier Nogueras Iso, Joanna Nowak, Katalin Toth. *What Technology does INSPIRE need? – Development and Research Requirements*. Proceedings of the 8th AGILE Conference on Geographic Information Science (Estoril, Portugal). ISBN 972-8093-13-6, pages 333-340, May 2005
- [21] Pedro Álvarez, José R. R. Viqueira, José Varela, Pedro J. Saco. *A Sensor Web based Framework for Natural Disasters Management: Application to Risk Assessment in*

Galicia. Actas de Jornadas Técnicas de la Infraestructura de Datos Espaciales de España (JIDEE'04), Zaragoza, Spain, November 2004

- [22] Jesús Barrera, Natalia Olmos, Rubén Béjar, David Portolés-Rodríguez. *Búsqueda de rutas de carretera en Web, una aproximación tecnológica basada en servicios de IDE*. Actas de Jornadas Técnicas de la Infraestructura de Datos Espaciales de España (JIDEE'04), Zaragoza, Spain, November 2004
- [23] P.Gallardo Konczanin, R.Béjar Hernández. *Adaptación de información geográfica extensiva en formato CAD para su explotación por WMS en IDE*. Actas de Jornadas Técnicas de la Infraestructura de Datos Espaciales de España (JIDEE'04), Zaragoza, Spain, November 2004
- [24] Javier Lacasta, Javier Nogueras-Iso, David Portoles, F. Javier Zarazaga-Soria. *La gestión de ontologías en el ámbito de las Infraestructuras de Datos Espaciales*. Actas de Jornadas Técnicas de la Infraestructura de Datos Espaciales de España (JIDEE'04), Zaragoza, Spain, November 2004
- [25] Silvia Laiglesia, Rubén Béjar, David Portolés-Rodríguez, P.R. Muro-Medrano. *Tecnología de IDE en agricultura, un ejemplo de servicios de información distribuidos para la gestión de ayudas a la política agraria comunitaria*. Actas de Jornadas Técnicas de la Infraestructura de Datos Espaciales de España (JIDEE'04), Zaragoza, Spain, November 2004
- [26] R. Miguel, M.Á. Latre, J. López-Pellicer, P.R. Muro-Medrano. *IDE en los sistemas de información de la CHE para la implantación de la Directiva Marco del Agua*. Actas de Jornadas Técnicas de la Infraestructura de Datos Espaciales de España (JIDEE'04), Zaragoza, Spain, November 2004
- [27] Rafael Tolosana-Calasanz, F. Javier Zarazaga-Soria, Javier Nogueras-Iso. *Optimización de los servicios de búsqueda de un catálogo de metadatos geográficos*. Actas de Jornadas Técnicas de la Infraestructura de Datos Espaciales de España (JIDEE'04), Zaragoza, Spain, November 2004
- [28] J. Nogueras-Iso, F. J. Zarazaga-Soria, P.R. Muro-Medrano. *Management of nested collections of resources in Spatial Data Infrastructures*. Proc. of the 1st International Workshop on Geographic Information Management (GIM'04), Zaragoza, Spain, pages 878-882, September 2004
- [29] Rubén Béjar, Pablo Gallardo, Michael Gould, Pedro R. Muro-Medrano, Javier Nogueras-Iso, F. J. Zarazaga-Soria. *A High Level Architecture for National SDI: the Spanish Case*. Proc. of the 10th EC-GI & GIS Workshop ESDI: The State of the Art, Poland, Warsaw, June 2004
- [30] M. Á. Latre, P. J. Álvarez, R. Béjar, P. R. Muro-Medrano. *Integrating legacy systems and the GIS elements of the Water Framework Directive into an Spatial Data Infrastructure*. Proc. of the 10th EC-GI & GIS Workshop ESDI: The State of the Art, Poland, Warsaw, June 2004

- [31] J. Nogueras-Iso, F.J. Zarazaga-Soria, J. Lacasta, R. Tolosana, P.R. Muro-Medrano. *Improving multilingual catalog search services by means of multilingual thesaurus disambiguation*. Proc. of the 10th EC-GI & GIS Workshop ESDI: The State of the Art, Poland, Warsaw, June 2004
- [32] M.A. Manso, J.Nogueras-Iso, M.A. Bernabé, F.J.Zarazaga-Soria. *Automatic Metadata Extraction from Geographic Information*. Acts of AGILE 2004, 7th AGILE conference on Geographic Information Science, Heraklion,Greece, May 2004
- [33] C. Granell, J. Poveda, M. Gould. *Incremental Composition of Geographic Web Services: An Emergency Management Context..* Proc. of the 3rd Intl. Conf. on Electronic Government (EGOV'04). Zaragoza, Sep. 2004. Springer LNCS 3184, pp. 562-563 (ISBN: 3-540-22916-7)
- [34] C. Granell, J. Poveda, M. Gould. *Incremental Composition of Geographic Web Services: An Emergency Management Context..* F. Toppen, P. Pastracos (eds.): Proc. of the 7th Conference on Geographic Information Science (AGILE 2004) Heraklion, Creta (Grecia). Abril 2004 Crete University Press, pp. 343-348 (ISBN: 960-524-176-5).
- [35] J. Poveda, M. Gould, C. Granell. *ACE GIS Project Overview: Adaptable and Composable E-commerce and Geographic Information Services..* F. Toppen, P. Pastracos (eds.): Proc. of the 7th Conference on Geographic Information Science (AGILE 2004) Heraklion, Creta (Grecia). Abril 2004 Crete University Press, pp. 795-799 (ISBN: 960-524-176-5).
- [36] Lars Bernard, Max Craglia, Michael Gould, Werner Kuhn. *Towards an SDI Research Agenda*. Proceedings of the 11th EC-GIS Workshop, Sardinia, 28 June - 1 July 2005
- [37] C. Granell , M. Gould. *Reutilizando composiciones de servicios web mediante la creación de componentes integrados*. Actas de las I Jornadas Científico-Técnicas en Servicios Web (JSWEB 2005), Granada, Spain, September 2005. Editorial Thomson-Paraninfo (in Spanish)
- [38] C. Granell , M. Gould, R. Gronmo, D. Skogan. *Improving Reuse of Web Service Compositions*. Proceedings of EC-Web 2005 Conference (Lecture Notes in Computer Science), Copenhagen, Denmark, 23-26 August 2005.
- [39] C. Granell , M. Gould, and F. Ramos. *Service Composition for SDIs: integrated components creation*. Proceedings DEXA 2005 Workshops (IEEE CS Press), Copenhagen, Denmark, 23-26 August 2005.
- [40] M.A. Manso, M.A. Bernabé. *Prototipo de portal de mapas antiguos basado en tecnología WMS*. ctas de Jornadas Técnicas de la Infraestructura de Datos Espaciales de España (JIDEE'04), Zaragoza, Spain, November 2004.
http://idee.unizar.es/jidee/documentos_en_linea/papers/PrototipoGeoPortalDeMapasAntiguos.pdf
- [41] M.A. Manso, M.A. Bernabé. *Los principios del diseño de interfaces aplicados a los servidores de mapas*. Actas de Jornadas Técnicas de la Infraestructura de Datos Espaciales

de España (JIDEE'04), Zaragoza, Spain, November 2004.

http://redgeomatica.rediris.es/JIDEE_2004_Zaragoza/Princip_diseño.pdf

- [42] D. Ballari, M.A. Manso, M.A. Bernabé. *MapServer OpenSource: Aplicación con los datos públicos del Gobierno de La Rioja (España)*. Actas de Jornadas Técnicas de la Infraestructura de Datos Espaciales de España (JIDEE'04), Zaragoza, Spain, November 2004. http://redgeomatica.rediris.es/JIDEE_2004_Zaragoza/MapServer_rioja.pdf
- [43] M.A. Bernabé, M.A. Manso. *Las Infraestructuras de Datos Espaciales y su aplicación a los deslizamientos de El Salvador de 2001*. Jornadas Técnicas de Modelización de Riesgos. 2004., Universidad Politécnica de Madrid, November 2004 http://www.cedex.es/publicador/formacion/2004/jor_mod_ries_2004.pdf
- [44] A. Barral, D. Balari, M.A. Bernabé, M. Castejón, M.A. Manso, J. Owens. *Spanish Christian Sacred Sites: from library to interoperability Web Services (WMS-WFS)*. GIS Planet 2005, Estoril (Portugal), June 2005
- [45] M.A. Manso, M.A. Bernabé, A. Staller. *Dinamic CRS WebService Catalog & Parser approach*. GIS Planet 2005, Estoril (Portugal), June 2005
- [46] M.A. Bernabé, M.A. Manso, , D. Ballari. *Correction of relief inversion in images served by a web map server*. International Cartographic Conference (ICC 2005), A Coruña, Spain, July 2005
- [47] M.A. Bernabé, M.A. Manso, *Open Source componets for geospatial portal*. International Cartographic Conference (ICC 2005), A Coruña, Spain, July 2005
- [48] D. Ballari, M.A. Manso, M.A. Bernabé, *Spatial Data Infrastructures, Web Map Server and historical heritage: Project LIFE Valle del Tiermes – Caracena*. International Cartographic Conference (ICC 2005), A Coruña, Spain, July 2005
- [49] M.A. Manso, M.A. Bernabé, *New Advances in the Automatic Metadata Retrieval from Geographic Information*. 11th EC GI & GIS Workshop, Alghero, Sardinia (Italia), June 2005
- [50] M.A. Manso, M.A. Bernabé, *Open Source components to build a GeoPortal*. 11th EC GI & GIS

Connectionist and Rapid Prototyping for Analyzing and Forecasting Data Series TIC2003-08807-C02

Miguel Delgado Calvo-Flores^{*}
Main Researcher, Co-ordinator
TIC2003-08807-C02-01
University of Granada

Luis Jiménez Linares^{**}
Main Researcher
TIC2003-08807-C02-02
University of Castilla-La Mancha

Abstract

Data series are multidimensional variable sequences which are indexed by means of the values of a one-dimensional variable. Classic methods for analyzing these data series impose restrictions which are rarely satisfied in real problems. As a result, research began several years ago in order to adapt pattern learning and recognition techniques developed in the field of artificial intelligence to data series analysis, and this project is set within this line. We shall examine not only dynamic recurrent neural networks, decision trees and hybrid methods but also numeric learning and quick prototyping tools for analyzing and predicting data series. In addition, the obtaining of the equivalent rule-based systems allows implicit knowledge to be described in the series in terms which are acceptable for the users. Theoretical developments shall be contrasted through experimentation on real data. We shall focus on two types of problems: the analysis of data from the management of both public and private economic units, paying special attention to problems such as predicting economic-financial risks, value models, etc.; and the study of data series modeling in biomechanics, and in particular the series from the sampling of the characteristic variables of human arm and leg movement.

Keywords: Data Series, Dynamic Recurrent Neural Networks, Time Series, Evolutionary algorithms, Multiobjective Optimization, Temporal Fuzzy Chains, Biomechanical modeling.

1 Project aims

1.1 Introduction

Data series are sequences of values of a multi-dimensional variable which are indexed by means of another, usually one-dimensional, variable and which generalize the time series idea. Classic methods for analyzing and forecasting these data series are basically statistical, and generally impose restrictions which are rarely satisfied in real problems. For this reason, a fruitful line of research began several years ago into adapting pattern learning and recognition techniques

^{*}Email: mdelgado@ugr.es

^{**}Email: luis.jimenez@uclm.es

developed in the field of artificial intelligence to data series analysis. As the hypotheses of these adapted techniques are much more lenient than those imposed by classic methods, they therefore have greater applicability. If we refer to specialist literature, we can see that research has been carried out into various tools which have become known as “intelligent computation” (artificial neural networks, flexible computation, evolutionary algorithms, regression trees, etc.), revealing the appearance in recent years of ad hoc-developed techniques within the emerging approach known as “data mining”.

It is evident that connectionist models (neural networks and associative memories), and also certain fuzzy system models, which have been designed “ad hoc” due to the difficulty of identifying them in the case of time dependent data, have so far generally produced the best results. The employment of dynamic recurrent neural networks has also been analyzed in order to identify an automaton which symbolically describes the overall behavior of a series. We shall pay special attention to this matter by using the results of some of our previous research and by combining dynamic recurrent neural networks with evolutionary algorithms.

Although decision and regression trees are pattern recognition tools which are not widely used in the field of time series, they have been successfully applied in order to identify extremely efficient description rule-based systems, which are both predictive and causal of a specific time series. For this reason, we shall explore in greater depth the possibility of analyzing and predicting data series by means of regression trees.

In this project, we therefore propose to examine the use of dynamic recurrent neural networks and hybrid methods as numerical learning tools for analyzing and predicting data series and the obtaining of equivalent rule-based systems which allow the results (the implicit knowledge in the series) to be described in terms which users can accept and understand. Our general aim is to enrich the panorama of tools for analyzing and predicting data series, which is of interest in many fields, and we shall focus on two of these: firstly, the description of the behavior of an economic unit, and secondly, the description of human movements.

The experiments of the first subproject shall focus on predicting the behavior of an economic unit, while those of the second shall focus on describing human movements; the models obtained shall then be shared and compared. The key to the advantages obtained from coordination lies in the fact that both the description of the behavior of an economic unit and the temporal space description of a human movement may be formulated in similar way to a set of evolving variables. The final result shall be to have a wider battery of contrasted prediction tools in problems of a very diverse nature and origin, although all within the analysis and prediction of data series.

In the following sections, in addition to the general objectives which we have previously described, we shall break down the objectives for the different subprojects, including in each case the work which must be carried out within the scheduled coordination.

1.2. Subproject 1 Objectives: Connectionist Models and Hybrid Methods for Analyzing and Predicting Data Series (MONADA)

General objectives

- To research specific design and learning methods based on evolutionary algorithms for recurrent dynamic networks which are more efficient than current ones.

- To study the problem of analyzing and predicting data series, mainly in the sphere of economic-financial risk prevention, assessment, corporate trajectory (using the models developed in O.1 for this purpose). Similarly, the results obtained shall be compared with those provided by other classic and non-classic techniques.
- To compare the techniques and models developed in this subproject with those obtained in Subproject 2. We shall place special emphasis on the data series prediction experiments of the economic-financial risk of both private and public economic units.
- To develop (if required) the necessary economic models for interpreting the results of the previous experimental studies.
- To develop the necessary software tools for implementing the obtained results.

Specific objectives

- To develop methods to determine the optimum topology of a dynamic recurrent network.
- To develop training methods which are more efficient than current ones (mainly based on error minimization by means of decreasing gradient).

For the development of specific objectives, we propose that evolutionary algorithms be used, in particular classic genetic algorithms with real coding (generational model, stationary model), multimodal evolutionary algorithms, and multiobjective evolutionary algorithms. Beginning with data series about the evolution and viability of economic units, we shall establish a segmentation criterion by means of dynamic recurrent networks which have been constructed and trained to enable the economic-financial risk of economic units to be predicted, and we shall study assessment models which can be applied to these. The results obtained will be contrasted with those generated by classic and non-classic techniques, particularly with other connectionist models such as associative memories.

1.3. Subproject 2 Objectives: Soft-Computing Methods for Analyzing and Categorizing Data Series (DIMOCLUST)

General objectives

- To research the improvement of the methods studied (i.e. ADRI and Temporal Fuzzy Chains) which enable their inference quality to be increased in order to reduce their estimated error.
- To study the possibility of establishing metrics which enable us to compare structural models such as ADRO or CDT.
- To focus the field of study on the data series of human movements for biomechanical analysis and study.
- To compare the techniques and models studied in this subproject with those obtained in Subproject 1.

- To develop the necessary software tools to implement the obtained results.

Specific objectives

- To incorporate mechanisms to minimize error on a local level. To research the effectiveness of evolutionary methods such as classic genetic algorithms.
- To study the improvement of the models studied, when expert information is incorporated a priori in the form of linguistic labels.
- To research methods and algorithms in order to equip our models with the facility of identifying cycles in the data series.
- To quantify the models by means of measurements on the number of elements, structural complexity, etc.
- To see whether it is possible to establish a segmentation of the models studied by using the suggested measurements.

1.4. Time diagram

In order to achieve these objectives, we have subdivided the work into the following subtasks:

- T1.1 Review of the techniques for designing and learning dynamic recurrent neural networks (9 months).
- T1.2 Review of the techniques for designing and learning dynamic recurrent neural networks based on the use of evolutionary algorithms (21 months).
- T1.3 Software specification and construction (18 months).
- T1.4 Review of the techniques for analyzing and predicting the behavior of an economic unit (9 months).
- T1.5 Adaptation and use of the methodologies developed in this subproject for analyzing and predicting the behavior of an economic unit (21 months).
- T1.6 Analysis and prediction of movement series (18 months).
- T1.7 Setting up, running and maintenance of an information and development server through the creation of a small Intranet which would enable us to use collaborative tools; development of centralized software; and publication of all developments and results (30 months).
- T2.1.1 Training in motion capture instrumentation to be used (3 months).
- T2.1.2 Capture and digitalize motion images of people (both top-class athletes and sportspeople and also amateurs) doing sport (27 months).
- T2.2.1 Study and updating of local model adjustment techniques (9 months).
- T2.2.2 Selection of alternative techniques and their incorporation into current software (9 months).
- T2.2.3 Running of model induction using constructed software (12 months).
- T2.3.1 Study and updating of techniques for parametric and non-parametric segmentation (clustering) (9 months).

TIC2003-08807-C02

- T2.3.2 Study and updating of segmentation and identification techniques based on expert knowledge (9 months).
- T2.3.3 Study of hybrid segmentation mechanisms based on data and expert knowledge (12 months).
- T2.4 Analysis and prediction of economic time series (18 months).
- T2.5.1 Creation and setting up of a group corporative server (18 months).
- T2.5.2 Development of an application to standardize all the achievements of the subproject (12 months).
- C1 Exchange of models and techniques. Study and monitoring of the development of the combined models (30 months).
- C2 Specification and description of the software to be developed (24 months).
- C3 Standardization of results (6 months).
- C.4 During project execution, at regular intervals, meetings will be held as necessary between the two main researchers and team components to monitor the development of the work and the landmarks reached.

Table 1 below shows the time diagram for the coordination of the different tasks.

Table 1: Time diagram and task distribution

Tasks	First year	Second year	Third year
T1.1	x x x x x x x x		
T1.2	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
T1.3		x x x x x x x x x x x	x x x x x x x x x x x
T1.4	x x x x x x x x x x x		
T1.5	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
T1.6		x x x x x x x x x x x	x x x x x x x x x x x
T1.7	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
T2.1.1	X x x x x x x x x x x		
T2.1.2	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
T2.2.1	X x x x x x x x x x x		
T2.2.2	x x x	x x x x x x x x x x x	
T2.2.3		x x x x x x x	x x x x x x x x x x x
T2.3.1	x x x x x x x x x x x		
T2.3.2	x x x	x x x x x x x x x x x	
T2.3.3	x x x	x x x x x x x	x x x x x x x x x x x
T2.4		x x x x x x x x x x x	x x x x x x x x x x x
T2.5	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
Coordination			
C1	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
C2		x x x x x x x x x x x	x x x x x x x x x x x
C.3			x x x x x x x
C4	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

2 Project progress and achievements

According to the work schedule proposed in Table 1, after slightly over a year and a half of work, the state of the evolution of the project is on course. In accordance with the proposed actions (see Section 1.4), the relevant results obtained are outlined in the following section.

2.1. Summary of the scientific results and techniques obtained for Subproject 1.

The proposal of this subproject includes both the theoretical development and the construction of software which enables it to be applied in data series prediction problems. Research is currently quite well advanced, borne out by publications in national and international congresses and journals. The experiments carried out have shown the viability of the project, producing satisfactory results in benchmark problems which have been widely used in the bibliography and also in certain real applications.

The theoretical models are currently being modified so that full advantage may be taken of their potential. In parallel, the smooth running of the models is being checked in real problems for predicting data series of a financial-environmental nature and social indicators, and also the comparison of results with the different models of feedforward networks previously used in the chosen field of application. In terms of the theoretical contribution, this has so far consisted in studying and developing new evolutionary techniques to improve the training of dynamic recurrent neural networks.

Evolutionary algorithms are search, optimization and learning algorithms which have been applied to a large number of problems, and have obtained suitable results. For example, in the research field of grammatical inference and time series prediction, the use of genetic algorithms for training recurrent neural networks has produced better results than those obtained by classic gradient-based training algorithms. Due to the good behavior of the genetic algorithms applied in this field, a research line is presented which promises encouraging results.

Some classic techniques, however, which improve the power of gradient-based algorithms such as certain Quasi-Newton methods and other non-linear programming techniques have proved to be suitable tools for training dynamic recurrent neural networks in time series prediction problems. While evolutionary algorithms carry out a heuristic search of the solution space and attempt to find optimum solutions, the Quasi-Newton algorithm search is based on the information provided by the gradient. In order to exploit the advantages of each methodology and so as to achieve better results, we have developed hybrid techniques which combine the characteristics of evolutionary algorithms and Quasi-Newton algorithms, and which improve the training of dynamic recurrent neural networks.

Finally, the development of multi-objective optimization techniques has enabled us to obtain both the optimum structure of the network and the trained network itself. The proposal has the following advantages:

- Easy adaptation to different models of neural networks.
- Capacity to solve the problem of obtaining local solutions which appears in classic gradient-based algorithms and other heuristic techniques (e.g. taboo search and simulated annealing).

- Integration of the optimization stages of the structure and training of a DRNN, thereby reducing the number of experiments which need to be performed.
- Generality for its application in different types of problems.

On a practical level, the methodology developed presents certain innovative aspects for solving data series prediction problems:

- Capacity for modeling systems where the temporal relations between data vary with time (adaptation to dynamic systems).
- Reduction in the number of experiments in the neural network input and output structure selection stage.
- Capacity for processing data patterns of variable length.
- Capacity for learning complex temporal relations in the long and short term.

The proposed models are being validated with financial data about the indebtedness of town councils in the Autonomous Community of Andalusia in collaboration with these.

Our current work focuses on the theoretical improvement of hybrid and multi-objective models. Another line of research focuses on the construction of a web system which enables remote access to the data of the results using either PCs with Internet connection or mobile devices. Some prototypes about abstract architectures and multiagent systems to implement that kind of web systems with remote access have been already developed and checked. Some papers about that have been also published.

Due to the good results obtained so far, it is hoped that project completion may culminate in an important contribution in the research field of neural network training and optimization.

2.2. Summary of the scientific results and techniques obtained for Subproject 2.

The achievements of this subproject may be divided into two blocks: scientifically-endorsed theoretical achievements and practical results where they may be incorporated.

Among the theoretical achievements obtained, it is worth mentioning the definition of Temporal Fuzzy Chains (TFCs) as a mechanism for representing systems which evolve in time, which we shall call “Dynamic Systems”. This formalism is based on the specification of fuzzy states (which represent the stable system), and the identification of the transitions between states. This representation enables a linguistic transcription of the model, obtaining a qualitative tool for the description of this type of system. The use of intervals and predetermined linguistic variables increases the descriptive character of the model.

For TFC construction, a set of algorithms have been proposed which are directly induced from a dataset monitoring the system or which are obtained by means of the temporal ordering and grouping of a set of fuzzy rules which describe the dynamic system. When we begin with a model of fuzzy rules which define the system, these rules are ordered in a way which reflects their sequentiality. This ordered type of rule sequence has been called a temporal fuzzy model.

All these models and methods have been accepted in different national and international scientific fields, where they have been presented and discussed (as demonstrated by the series of publications which endorse them).

Another of the most outstanding aspects of the theoretical results of the project is the

creation of a line of research into handling information on Internet. From the initial work for the use of deformable prototypes in the sphere of temporal complex systems, new techniques have been formed where the treatment of term imprecision, similarity and concordance and also the introduction of new ontologies enable a greater exploitation of large databases or enormous amounts of information, as in the case of Internet.

From the practical point of view, the theoretical results of TFCs have been used to describe simple human movements, such as jumps or steps, with good results; meanwhile, search and information handling techniques have provided systems for e-mail management and Internet search engines (FIS-CRM, GUMsen).

3 Performance indicators

3.1. Trainees

- Mr. Manuel Pegalajar Cuellar, FPU grant-holder, is working on the subject from project TIC2000-1362-02. He shall defend his doctoral thesis in approximately one year's time.
- Mr. Oscar Pérez Piñeiro, FPI grant-holder in the grant associated with this project joined the team in June 2004. He shall defend his doctoral thesis in approximately three years' time.
- Mr. Carlos Gonzalez Morcillo, part-time teacher at the University of Castilla la Mancha, shall defend his doctoral thesis in approximately three years' time.
- Mr. Luis Rodríguez Benítez, Professor at the University School at the University of Castilla la Mancha shall defend his doctoral thesis in approximately three years' time.

3.2. Summary of publications

Table 2: Summary of publications

Type	Number	References
<i>Papers in SCI journals</i>	6	[6] [7][8][9][10][11]
	2	[20][36]
<i>Papers in other international journals</i>	1	[4]
<i>International Conference papers</i>	5	[1][2][13][14][15]
	12	[20][21][24][25][28][29] [31]-[33][35][37][39]
<i>National Conferences papers</i>	4	[3][5][12][16]
	8	[17][18][19][22][23][27] [34][38]

3.3. Technological Transfer (patents, contact with other projects, etc.)

As we have emntioned before he proposed models are being validated with financial data about the indebtedness of town councils in the Autonomous Community of Andalusia in collaboration with these.

On the other hand, the execution of this project has enabled strong links to be forged with important biomechanical research groups both on the level of data analysis and of common structures. This multidisciplinary connection is apparent in the laboratories and establishments, and also in the personal contact between those at the Escuela Superior de Informática in Ciudad Real and the Facultad de Deporte in Toledo (University of Castilla-La Mancha), as revealed by the composition and participation of various thesis committees.

Similarly, the results obtained in previous projects, and also the exploration in greater depth of these within the project, have favored the concentration around J. Angel Olivas of a *National Thematic Network on soft-computing-based systems for accessing web information: SCIA*, sponsored by the Ministry of Science and Technology.

3.4. Teaching

The theoretical results have been used for teaching on the two doctoral courses:

- “*Advanced models for modeling and predicting chronological series*” on the doctoral program “*Design, analysis and applications of intelligent sysems*”, and
- “*Soft computing techniques*” on the doctoral program “*Architecture and management of information and knowledge in networked systems*,”

both doctoral programs awarded the Mention of Quality.

3.5. Dissemination of results

According to the task schedule, web pages have been constructed with information about the evolution and results of the project and this can be accessed at the following addresses: <http://oreto.inf-cr.uclm.es> and <http://drimys.ugr.es/proyectos/PM2003>.

4 References

- [1] Cuéllar, M. P., Delgado, M., Pegalajar, M. C., “An application of non-linear programming algorithms to train recurrent neural networks in time series prediction problems”, Proceedings of ICEIS05, Vol. 2, pp. 35-43, Miami, 2005.
- [2] Cuéllar, M. P., Delgado, M., Fajardo, W., Pérez-Pérez, R. “An information system to perform services remotely from a web browser”, Proceedings of ICEIS05, Vol. 2, pp. 91-100, Miami, 2005.

- [3] Cuéllar M. P., Delgado M., Pegalajar M.C., "Redes neuronales recurrentes para predicción de series temporales" in Proceedings of SICO'05 (Accepted).
- [4] Cuéllar M. P., Delgado M., Pegalajar M. C., "Multiobjective Evolutionary Optimization for Elman Recurrent Neural Networks, Applied to Time Series Prediction", in Fuzzy Economical Review, vol. X, no. 1, pp. 17-34, May 2005.
- [5] Cuéllar M. P., Delgado M., Pegalajar M.C., "Algoritmos evolutivos híbridos para entrenamiento de redes neuronales recurrentes", in MAEB'05 (Accepted).
- [6] Delgado, M., Fajardo W., Gibaja, E., Pérez-Pérez, R. "BioMen: an information system to Herbarium", Expert System with Applications. Vol. 28/3 pp 507-518.
- [7] Delgado, M., Olavarrieta, P., Vergara, "Fuzzy Sets Tools in Process Quality Control", International Journal of Uncertainty, Fuzziness and Knowledge-based Systems (Submitted)
- [8] Cuéllar, M. P. , Delgado, M. , Pegalajar, M. C. , "Training Fully Recurrent Neural Networks with Different Evolutionary Techniques", Neural Networks (Submitted).
- [9] Cuéllar, M. P. , Delgado, M. , Pegalajar, M. C. , "A Multiobjective Evolutionary Algorithm to find optimal Fully Recurrent Neural Networks", International Journal of Approximate Reasoning, (Submitted).
- [10] Cuéllar, M. P., Delgado, M., Pegalajar, M. C., "Memetic Evolutionary Training for Recurrent Neural Networks. An application to Time Series Prediction", Expert Systems, (Submitted).
- [11] Cuéllar, M. P., Delgado, M., Pegalajar, M. C., "Extraction of Fuzzy Rules from Recurrent Neural Networks", IEEE trans. On Neural Networks, (Submitted).
- [12] Cuéllar, M. P., Delgado, M., Pegalajar, M. C., "Extracción de reglas difusas para describir la dinámica de redes neuronales recurrentes", Proc. ESTYLF'04, Jaén 2004 (Spain), pp. 483-489
- [13] Cuéllar, M. P., Delgado, M., Pegalajar, M. C., R. Perez-Perez, "A comparative study of evolutionary algorithms for training Elman recurrent neural networks to predict autonomous indebtedness", Proc. ICEIS'04, Porto 2004 (Portugal), pp. 461-465
- [14] Cuéllar, M. P., Delgado, M., Pegalajar, M. C., R. Perez-Perez, "An adaptable time delay neural network to predict the Spanish economic indebtedness", Proc. ICEIS'04, Porto 2004 (Portugal), pp. 457-460
- [15] Cuéllar, M. P., Delgado, M., Pegalajar, M. C., "Multiobjective evolutionary optimization for Elman recurrent neural networks, applied to time series prediction", Proc SIGEF04, Reggio-Calabria (Italia), 2004
- [16] Cuéllar, M. P., Delgado, M., Pegalajar, M. C., Pérez-Pérez R., "Predicción de series temporales mediante redes neuronales recurrentes y algoritmos evolutivos", Proc MAEB04, Córdoba, 2004, pp. 593-600.
- [17] Abián Vicén J, Alegre Durán L M, Fernández Rodríguez JM, Jiménez Linares L, Sánchez Amador L, Aguado Jodar X. "Variabilidad en la Capacidad de Amortiguación de Caídas Después de Ejercicios Intensos", Archivos de Medicina del Deporte: I Jornadas Internacionales de Medicina del Deporte, Zaragoza, 26-27 November 2004.
- [18] Abián Vicén J, Alegre Durán L M, Jiménez Linares L, Sánchez Amador L, Aguado Jodar X. "Tests de Salto con Plataforma de Fuerzas en el Ingreso de una Facultad de Ciencias del Deporte". Archivos de Medicina del Deporte: I Jornadas Internacionales de Medicina del Deporte, Zaragoza, 26- 27 November 2004.
- [19] Amador J. Sánchez, Javier Abián Vicén, Luis M. Alegre Durán, Luis Jiménez Linares, Xavier Aguado Jodar. "Medición Directa de la Potencia con Tests de Salto en Voleibol Femenino". Archivos de Medicina del Deporte, Volumen XXII (2), March-April, 2004, p 111-120.

- [20] Francisco P. Romero, José A. Olivas, Pablo J. Garcés, "Using Soft-Computing for e-mail management", Proceedings of the Tenth International Conference on Information Processing and Management Uncertainty in Knowledge-Based Systems, IPMU 2004, Volume 3, p 1783-1790, Perugia, Italy, 4-9 July 2004.
- [21] Francisco P. Romero, José A. Olivas, Pablo J. Garcés. "A Soft-Computing based Tool for Intelligent e-mail Management", Proceedings of the International Conference on Artificial Intelligence, IC-AI-04, Volume I, p 56-61, Las Vegas, Nevada, USA, 2004.
- [22] Francisco P. Romero, José A. Olivas, Pablo J. Garcés. "Metodología de Tratamiento Inteligente del Correo Electrónico Mediante Técnicas de Soft Computing", Actas del XII Congreso Español Sobre Tecnologías y Lógica Difusa, p 503-508, Jaén, 15-17 September 2004.
- [23] J. Moreno-García, L. Jiménez, J.J. Castro-Sánchez, L. Rodríguez. "Conversión de Modelos Difusos Temporales a Cadenas Difusas Temporales". Actas del XII Congreso Español Sobre Tecnologías y Lógica Difusa, p 621-626, Jaén, 15-17 September 2004.
- [24] J. Moreno-García, L. Jiménez, J. J. Castro-Sánchez, C. González, R. Manjavacas, S. García, X. Aguado. "Fuzzy Modelling of Sport Movements", 1st International Working Conference IT and Sport & 5th Conference dvs-Section Computer Science in Sport, p 21-22, Cologne, Germany, 15-17 September 2004.
- [25] J. J. Castro-Sánchez, R. Manjavacas, J.L. Castro, J. Moreno. "Design of a merchant broker agent based on fuzzy logic as mediator in e-commerce", Proceedings of the Tenth International Conference on Information Processing and Management Uncertainty in Knowledge-Based Systems, IPMU 2004, Volume 3, p 1833-1840, Perugia, Italy, 4-9 July 2004.
- [26] J. J. Castro-Sánchez, L. Jiménez, J. Moreno, L. Rodríguez "Using fuzzy repertory table-based technique for decision support", Decision Support System 39 , 293-307.2005
- [27] Javier de la Mata, José A. Olivas, Jesús Serrano-Guerrero. "Desambiguación de Significados en GUMSe basada en el refinamiento del sentido". Actas del XII Congreso Español Sobre Tecnologías y Lógica Difusa, p 315-320, Jaén, 15-17 September 2004.
- [28] Javier de la Mata, José A. Olivas, Jesús Serrano-Guerrero. "Improving Web Search Using Query Customization with GUMSe", Current Issues in Data and Knowledge Engineering, p 325-333, Warszawa, Poland, 2004.
- [29] Javier de la Mata, José A. Olivas, Jesús Serrano-Guerrero. "Overview of an Agent Based Search Engine Architecture", Proceedings of the International Conference on Artificial Intelligence, IC-AI-04, Volume I, p 62-67, Las Vegas, Nevada, USA, 2004.
- [30] Jesús Serrano-Guerrero, José A. Olivas, Javier de la Mata. "Relaciones Conceptuales para la Construcción de Ontologías con GUMSe", Actas del XII Congreso Español Sobre Tecnologías y Lógica Difusa, p 669-674, Jaén, 15-17 September 2004
- [31] Jesús Serrano-Guerrero, José A. Olivas. "Discovery of conceptual relations for ontology construction in GUMSe", Annual Meeting of the North American Fuzzy Information Processing Society, NAFIPS 2004, p 647-651, Banff, Alberta, Canada, 27 - 30 June 2004.
- [32] José A. Olivas, Javier de la Mata, Jesús Serrano-Guerrero. "Ontology Constructor Agent for improving Web Search with GUMSe", Proceedings of the Tenth International Conference on Information Processing and Management Uncertainty in Knowledge-Based Systems, IPMU 2004, Volume 2, p 1341-1348, Perugia, Italy, 4-9 July 2004.

- [33] José Jesús Castro-Sánchez, Juan Moreno-García, Ramón Manjavacas, Carlos Gonzalez, "Assisting Negotiations and Supporting Decisions in Small E-Business", Proceedings of 2004 IEEE International Conference on Fuzzy Systems, p 855-860, Budapest, Hungary, 25-29 July 2004.
- [34] Juan Moreno-García, Carlos González, Ramón Manjavacas, "Inducción Directa de Modelos difusos temporales", Actas del XII Congreso Español Sobre Tecnologías y Lógica Difusa, p 615-620, Jaén, 15-17 September 2004.
- [35] Juan Moreno-García, Luis Jiménez, José Jesús Castro Sánchez. "A Direct Induction Algorithm of Temporal Fuzzy Models", Proceedings of 2004 IEEE International Conference on Fuzzy Systems, p 635-640, Budapest, Hungary, 25-29 July 2004.
- [36] Juan Moreno-García, Jose Jesus Castro-Sánchez, Luis Jiménez "A fuzzy inductive algorithm to model dynamical systems in a comprehensible way ", IEEE Transactions on fuzzy system. (Submitted).
- [37] Pablo J. Garcés, José J. Castro-Sánchez, Ramón Manjavacas, José A. Olivas. "A suppliers searcher agent based on FIS-CRM used in E-commerce websites", Proceedings of the International Conference on Artificial Intelligence, IC-AI-04, Volume I, p 50-55, Las Vegas, Nevada, USA, 2004.
- [38] Ramón Manjavacas, José Jesús Castro Sánchez, Juan Moreno García. "Agente vendedor difuso para entornos de negociación automática en comercio electrónico", Actas del XII Congreso Español Sobre Tecnologías y Lógica Difusa, p 645-650, Jaén, 15-17 September 2004.
- [39] Romero, F. P.; Olivas, J. A.; Garcés, P.: "A Methodology for Intelligent e-mail Management", Proceedings of the Seventh International Conference on Enterprise Information Systems, ICEIS 2005, Miami, USA, ISBN: 972-8865-19-8, pp. 11-16.

Rules as a tool for the design and management of relational databases: algebraic foundations, deduction methods and efficient tools.

TIC2003-08687-C02

Inma P. de Guzmán *
Univ. Málaga

Olga Pons †
Univ. Granada

Summary

Data connections are becoming an outstanding area in databases, involving dependencies (of several kinds) and associations. These data relations may be conceive as integrity rules which provide information about the database structure and, therefore, may help the design and management tasks. In this proposal, we intend to give a uniform and efficient solution to the problem of how to specify and to manipulate these data relations.

Normally, the work in this area is centered on the design of methods to discover new data relations directly from the database itself (for example, using data mining techniques). Another direction is to provide an uniform treatment of the different kinds of rules and to develop deduction methods to discover new rules from existing rules. Both approaches are complementary, because inferred rules allow us to prune the search in data mining techniques and rules discovered with these data mining methods provide new rules to the deduction method to carry out new inference tasks.

We propose the use of Logic to represent several types of data dependencies (concretely, we will focus on functional dependencies, multivalued dependencies, inclusion dependencies, fuzzy dependencies and temporal dependencies) and we will develop efficient deduction methods to manage these database rules. This approach allows a uniform and formal way to represent different levels of knowledge about the information system data and to prepare them to be manipulated in an automated way. The use of logic opens the doors to develop automated deduction methods which can be extended at the same time the research development is increasing. In this way, the theoretical result can be incorporated to the tools generated with lower cost.

*Email: guzman@ctima.uma.es

†Email: opc@decsai.ugr.es

1 Objectives of the project

The main aim of the project is to develop a general framework that allows us the specification and manipulation of database constraint, more concretely rules and dependencies. In the first stage of the project we have focussed on the development of a formal basis which allows to integrate different result in a efficient way.

The aim of the first sub-project is to develop deduction methods for Functional Dependencies logics. The kinds of logics may be used to design relational databases. Thus, the result have been applied to the field of Intelligent Information System, concretely, some CASE tools have been developed. These tools are based on a formal basis and they use the automated deduction techniques developed in this proposal to manage the database constraint in an efficient way.

The objectives in this sub-project are the following:

- From the starting point of the Functional Dependencies methateoretical results concerning Funtional Dependencies Logics developed by this subgroup in the past, we design efficient algorithms directly based on the logic deduction system. These algorithm face on with the most outstanding problems in this area: Canonical and Minimal Clousure, Normalization algorithms, the Key problem, etc.
- Extension of Functional Dependencies logic to manage other kinds of dependencies (Fuzzy, Temporal, etc). This objective looks for a uniform way to treat enriched database constraints.
- We propose the use of TAS, a methodology to develop Automated Deduction Method, to get an efficient treatment of database dependencies with a deduction method style. We intend to incorporate the notion of FD-Tree, a new representation of a set on Functional Dependencies, to the TAS methods.
- These works will be integrated around the FD³ methodology which allows the design of relational databases in a cooperative environment. This new methodology are based in a new data model strongly based on the notion of Funcional Dependencies.

The aim of the second sub-project is to obtain sound theoretical results directed to the definition of a complete unique model for fuzzy relational deductive databases able to collect all the necessary elements. It is also important the development of the tools that permit to implement and take advantage of the mentioned model with all its capacities. The final point is to validate and prove the whole system using a big amount of data both precise and imprecise as well as a great variety of types of rule. To carry out this last task it is very important the collaboration of an expert in the field, not only for the rules definition but also for the objective criticism of the obtained results.

More specifically, this general objective include:

- To define the concept of generalized deductive rule with certainty degree. In the current state, the fuzzy relational model supports the definition of non-fuzzy rules in the sense that the rule is not affected by a degree.
- To define a deductive relational syntax built specifically for the definition of fuzzy rules.
- To study the different meanings for the accomplishment degree in a rule.

- To investigate algorithms for the optimization of rules application making more efficient all data handling, specially the queries.
- To study the inclusion of integrity rules inside the model. Integrity rules will be used to avoid inconsistency in the stored data and to hold important constraints among them without the used of external applications.
- To carry out the creation of a new type of hybrid relation called intensional-extensional in which part of the content is given in an explicit way and part of it is given by means of a set of imprecise rules.

Research tasks

In this section we describe the planed task in order to cover the above objectives:

1. *Formal study of associations:*

The project expects to extend the initial results of this project by the sub-group of the University of Málaga to other types of dependencies, like the multivalued dependencies and, outstandingly, to fuzzy dependencies to relate it with the work made with the coordinated sub-group belonging to the investigation group of IdBIS (Intelligent Database and Information Systems) of the University of Granada.

This generalization requires a strong work of formalization in which we are obtaining good results and that we try to go deep. For it we have worked in the design of a formal frame within the theory of coalgebras. This theory has consolidated like the ideal framework for the formalization in a great variety of fields of Computer Science (Rutten, 2000). Like examples in which coalgebraic formalizations have given rise to important advances we can emphasize the Kripke's structures, the labelled transition systems, diverse types of automata, in particular the non-determinist ones (Rutten, 2000), the programs and classes in oriented to objects languages (Reichel, 1995; Jacobs and Rutten, 1997). We try, in summary, to give a coalgebraic formalization for the different types of dependencies that allow us to unify the different theories.

In addition in this study is included not only the dependency notion but even more flexible concepts like the association, that recently is seeing grow its interest in the scope of the data bases. So we will have a formal and uniform frame to deal with several types dependencies jointly and we will be able to take advantage of the results for the development of a new frame of logical models for data bases.

2. *Methods for manipulation of dependencies.*

The actual works have allowed the design of algorithms that manipulate efficiently a non structured set of dependencies. The formal framework has eased the introduction of a new logic to specify and manage Functional Dependencies. The main novelty of our logic is the introduction of a new rule, the substitution rule, which facilitates the design of automated deduction methods to manage Functional Dependencies.

These deduction methods are the basis to develop a set of executable modules that transform set of Functional Dependencies. These transformations have been integrated into some Information System intelligent tools successfully.

In a first stage we have developed algorithms that reduce efficiently the dependencies using only the substitution rule, the new inference rule developed by the research group of this project.

In our work with Functional Dependencies we have in mind a central goal: to provide a accurate formalization and to facilitate the design of new methods and tools that manipulate database constraints in an efficient way. More specifically, we have traversed the following way:

- First we have designed a preprocessing transformation that establishes an efficient preprocessing pruning based mainly on the substitution rules, and allows for the application of any other depuration method over the new Functional Dependencies set, which is shorter than the original one.
 - We have illustrated the difficulties to face the implication problem with a method directly based on logics of Functional Dependencies. We have introduced the notion of atomic-minimality, which guides the treatment of sets of Functional Dependencies in a rewriting logic. $\mathbf{SL}_{FD}$ axiomatic system is easily translated to rewriting logic and Maude 2.
 - Recently, we have developed a novel method to solve the attribute closure problem. We have designed an algorithm which is based directly on the $\mathbf{SL}_{FD}$ inference system, having an automated deduction style. We have implemented the algorithm and the most efficient classical algorithm, obtaining very promising results. The complexity in the worst case is the same as the classical linear closure algorithms.
 - Finally, we show how the formal techniques are used in heterogeneous and cooperative environments. We develop a tool to integrate relational schemes in an automatic way.
3. *Tools*: In the objectives we have mentioned the integration of the techniques concerning the manipulation of dependencies into a new architecture to design databases in a cooperative way. Our intention is to provide a set of intelligent tools with deduction capabilities which allows the user to manipulate a great number of information in a simple manner. Thus, we are developing a set of CASE tools that covers all the stages involved with all the database design process and introduces the deduction methods as the core of the information treatment step. The deduction methods allows a formal and efficient manipulation of the data structure at the same time that reduce the problem removing the redundancy that different user may introduce when all of them specify the same system separately.

Besides that, these tools may be used in heterogeneous environment and allows the integration of several data sources in a unique system. To do that, the tools are extended with reverse engineering capabilities to capture the dependencies specified by the user with the notion of key. The information are stored in a Functional Dependencies data dictionary (FDD^3) which may be enriched by the database administrator in the integration process. The FDD^3 is the center element of this methodology. It may be considered as a non-structured data source which usually stores a great amount of very simple elements (the dependencies). Thus, the key of this methodology is the deduction methods

which eases the task of manipulate this great amount of simple information and give an intelligent analysis of them.

Moreover we have developed a new data model named Functional Dependencies Data Model which states the specification and manipulation of information using Functional Dependencies. This data model includes a high level representation of the information using the Attribute/Dependence Diagram and a method to transform a Functional Dependencies specification into a relational database.

More concretely, we have developed two case tools which covers separately the two problems cited above.

- We have developed a CASE tool named A/D CASE (Attribute/Dependence CASE) that illustrates the practical benefits of the *FD3* architecture. We propose the use of logics based on the notion of Functional Dependencies (FD) to allows formal specification of the objects of a data model and to conceive future automated treatment. The existence of a FD logic provides a formal language suitable to carry out integration tasks and eases the design of an automatic integration process based in the axiomatic system of the FD logic.

To communicate the information collected in the global model, we introduce a *High Level Functional Dependencies* (HLFD) Data Model which is used in a similar way as the Entity/Relationship Model. The HLFD data model can be deduced automatically from a logic specification. Furthermore, it is possible to translate automatically the HLFD data model into a relational database.

- We have also used the logic algorithms to develop a new tool which integrates several ORACLE database schemes. When the user connects with the tool he introduces the identifier of the user account must be integrated and the tool, using the reverse engineering task, deduce the Functional Dependencies included in the different schemes generating a *FD³*. The remove redundancy step allows the elimination of replicate and superfluous information, generating a unique and global view of the union of schemes.

To protect the applications from this changes, a set of views are generated. These views play the role of the original tables and redirect the old queries to the new global scheme and the new tables.

Moreover, the tool carries out an integration of the schemes and an integration of the data itself, providing a new database with a common structure and containing all the information provided in the original subschemes.

4. *Formal Techniques for Software Engineering:*

The general framework of this project is the use of formal methods in software development. In this framework, the formalization of the static system models (data or class models) by means of dependencies it is essential. Moreover, we also consider very important to approach the formalization of the dynamic (behavior) models to obtain a complete system specification.

In this sense, we have consider adequate to provide formalization mechanisms that can be (optionally) hidden to the software developer. In our approach, the developer may

maintain the use of standard behavior models. In particular, we have considered that UML state machines are the basic behavior model. As it is well-known, UML (Unified Modeling Language) is the 'de facto' standard in object oriented software modeling.

Among the activities we have carried out in this project we have developed a formalization of the UML state machines using interval temporal logic. This approach, as far as we know, have nor been done before.

5. On the other hand, the group of Grananda University has the following important tasks to carry out:

- To decide which is the deductive fuzzy model that better fits with our problems. In this sense we have studied all fuzzy operators and fuzzy implications in the literature. Once the previous choice is made, the next task is to extend the classical syntax for rules to the fuzzy case, adding an accomplishment degree. This part is already completed, but an important point now is to determine the different meanings that both the rule and the degree can have depending on the type: integrity rule, deductive rule, definition rule, dependency rule ... and to study their properties, the inclusion inside the theoretical model, and the way to store and encode them into the system.
- An important part of the work is being devoted also to the implementation of these new elements in the system, basically extending the catalog and creating new modules to create and store the rules defined. The language used to do this work is FSQL (Fuzzy SQL) developed by members of the research group IDbIS. The extension obtained for the inclusion of fuzzy rules is called DFSQL.
- Once the rules can be inserted in the system, we have developed the procedures that permit to use the rule to fill in a table, that is to compute the result of the application of a rule against the existing tables. The most interesting result of this part of the project is the possibility to define recursive rules which is a new feature inside the relational world.

Gantt chart

Tasks	1st year	2nd year	3rd year
Task 1: Theoretical Foundations of Dependencies			
1.1 To formalize the concept of minimal closure	■		
1.2 To extend schema. To formalize keys and antikeys	■		
1.3 Formal definition of normal forms	■		
Task 2: To extend the algebraic study to:			
2.1 Multivalued	■		
2.2 Fuzzy	■		
2.3 Rules	■		
Task 3: To study new order structures			
Task 4: Automatic Reasoning over dependencies			
Task 5: Development of new methodology for the design of DBs			
5.1 Implementation of diagrams of dependencies			■
5.2 Tools. Integration and Purification			■
5.3 Construction of the data base			■
5.4 Construction of a new Model of High Level			■

Tasks	1st year	2nd year	3rd year
Task 1. To define a new concept of generalized deductive rule with certainty degree.	■		
Task 2: To define a relational deductive syntax.	■		
Task 3: To study the different meanings of certainty degrees in rules	■		
Task 4: To investigate the algorithms for optimization of rules execution	■		
Task 5: To study the inclusion of integrity rules	■		
Task 6: Creation of a new type of intensional-extensional relation	■		
Task 7: To develop the algorithms for the analysis of functional dependencies	■		

2 Level of success achieved to date

Most of the proposed tasks are being fulfilled as expected. In the rest of the section we will describe the major achievements obtained in each of the tasks.

The publication [6] is specifically oriented to task 1. In this article, we introduce the novel algebraic formalization for the concept of schema that we propose. This formalization has the aim to be the framework to generalize the functional dependency concept and is based on the concepts of non-deterministic operators and ideals in generalizations of the lattice structure.

Task 2 and 3 are covered in [4, 9, 5, 2]. In these papers we study in deep the concepts in which our theoretical formalization is based. The underlying idea is to provide a more

general context in which several kinds of dependencies can be treated in the same way. These generalizations are finally given in terms of coalgebraic theory.

The number of articles around Task 3 is higher perhaps because it is the more close to the applications one. Concretely, in [18] we show the advantages of the novel $\mathbf{SL}_{FD}$ logic presented in [3] for the treatment of Dependencies.

In [13, 14, 17] we introduce a pre-processing transformation which removes redundancy of a given set of Functional Dependencies and allows a more efficient treatment by other well known indirect algorithms. We study in the framework of lattice theory the minimal generator of a set of functional dependencies (it mean a colateral contribution to Task 1) and we carry out an empirical study to prove the practical benefits of our approach. We develop in Prolog a automated deduction method to prune a set of Dependencies.

In [1] and in [12] we use rewriting systems to transform automatically dependencies. These articles have also colateral contributions to Task 1. In they, we consider a complexity criterion which allows us to introduce a new minimality property for sets of Functional Dependencies named *atomic-minimality*. The $\mathbf{SL}_{FD}$ logic has allowed us to develop the heart of the work, which is the use of Rewriting Logic and Maude 2 as a logical framework to search for atomic-minimality.

In [19] we study the closure operators in automatic reasoning. In [10] and [11] we present a new method to compute the closure operator for a set of attributes. The main novelty of this method is the fact that it is directly based on a logic inference system. Our method has all the advantages of using logic (solid base, flexible deduction methods, metatheory development, etc) and is the base of a novel algorithm having linear complexity in the worst case. Finally, we carry out an empirical study of the execution of our algorithm and classical closure algorithms. This study shows the best behavior of our method.

Furthermore, we have carried out an empirical study about the execution time of the Diederich and Milton closure algorithm and our closure algorithm. $\mathbf{SL}_{FD}$ Closure has had a significantly better behavior than Linear Closure. In 594 cases from 600, that is in the 99 per cent of the cases, $\mathbf{SL}_{FD}$ Closure has calculated the closure spending less time than Linear Closure. The average of time using $\mathbf{SL}_{FD}$ Closure is 110,86 and using Linear Closure is 942,05.

In [20], we present LNint-e, a novel modal temporal logic with topological semantics. LNint-e has an expressive power that is unusual in modal temporal logics, since it allows point and interval expressions, as well as the absolute and relative approaches to time. This expressive power has allowed an exhaustive formalization of both the different kinds of operations of the state machines and all the transition types. LNint-e has also allowed the formalization of concurrent and non-concurrent composite states.

In [16] we have presented the logic Lni1, a first order extension of LNint-e logic. With Lni1 we could complete the formalization of our previous work, allowing the use of parameters in operations and events. Parameters are obviously an essential element in software modeling and has not been considered in LNint-e since this is a propositional logic.

In [8] we present results concerning the normalization of LNint-e expressions. As it is well known, normalization is a previous step in the development of deduction methods. We must point out that the definition of our normal form includes transformations that remove binary connective nestings. In this way we considerably reduce the formula management complexity.

Finally, we have began to work in Task 5 with [15] in which we develop a CASE tool that allows the integration of local subschemes designed using dependencies and in [7] we

have also used the $\mathbf{SL}_{FD}$ algorithms to develop a new tool which integrates several ORACLE database schemes. It carries out an integration of the schemes and an integration of the data itself, providing a new database with a common structure and containing all the information provided in the original subschemes. The tool propose an uniform web-interface for information integration and manipulation. This tool obtains from local sub-schemes a integrated scheme using the $\mathbf{SL}_{FD}$. We use the $\mathbf{SL}_{FD}$ language as canonic model and $\mathbf{SL}_{FD}$ inference system as integration and simplification method. Moreover, this tool allows apply direct and inverse engineering techniques with the relational model.

With respect to the Grananda sub-group the main results obtained can be summarized as follow:

We have organized a special session about features related to the project inside the International Congress of Information Processing and Management of Uncertainty, celebrated in Italy in July of 2004. We have already completed the algorithm used to trigger a rule and insert the results into a table. We have solved one of the possible cases to be treated in relation to the problem of rules definition in a fuzzy relational database. The algorithm calculates an intensional table (defined by means of a rule) from the fuzzy data stored in the DB [23]. We have also studied the same problem from the object-oriented perspective as can be seen in the paper [21]. The first formal results together with other results obtained previously by the group have been included in the paper [24] where real data about real state have been used to prove our system. Other results obtained have been published in the paper [22].

3 Progress indicators

In this moment two PhD students associated to the first sub-project are actively participating in different research tasks. This collaboration has produced a workshop contribution and several draft papers submitted to technical journals. We estimate that this collaboration is the beginning of theirs PhD Thesis.

The progress of the project has fruitful results that we have published in several communications.

Journal Publications

- Journals in Science Citation Index: Annals of Mathematics and Artificial Intelligence [4], Kybernetika [9], Discrete Mathematics [5]
- Other journals: Software and Systems Modeling [20]¹, Revista Iberoamericana de Inteligencia Artificial [8]

International Conferences.

- Conferences with proceedings published in Lecture Notes series: IBERAMIA: Ibero-American Conference on Artificial Intelligence [16], SBIA: XVII Brazilian Sympo-

¹This journal edited by Springer Verlag since 2001 and it focuses on theoretical and practical issues pertaining to the development and application of software and system modeling languages and techniques. Its editorial board includes well-known researchers in the area such as G. Booch or D. Harel.

sium on Artificial Intelligence [1], Conferencia de la Asociación Española para la Inteligencia Artificial CAEPIA [13]

- Other conferences: Conference of Computational Methods in Sciences and Engineering - ICCMSE [14], 6th International Conference on Enterprise Information System - ICEIS [15], International Conference on Computational and Mathematical Methods in Science and Engineering - CMMSE [17],

International Workshops

Theory and Applications of Relational Structures as Knowledge Instruments. Meeting of the COST Action 274 [18]

National Conferences

Primer Congreso Conjunto de Matemáticas RSME-SCM-SEIO-SEMA [12], XV Congreso Nacional Usuarios de Oracle QUORE [7],

The main researcher of the sub-group of Granada is tutoring a PhD Thesis with the aim to study and develop all the types of rules presented in the project. The PhD student is a researcher of the Instituto Andaluz de Astrofísica (ESA) and the research project he is included in is Exploration of the Solar System.

References

- [1] Gabriel Aguilera, Pablo Cordero, Manuel Enciso, Angel Mora, and Inmaculada P. de Guzmán. A non-explosive treatment of Functional dependencies using rewriting logic. *Lecture Notes in Artificial Intelligence, Springer-Verlag*, 3171, 2004, pp. 31–40.
- [2] Pablo Cordero, Inmaculada de las Peñas, Gloria Gutiérrez and Javier Martínez. Non determinism, coalgebras and multiset lattices. Submitted to *Theoretical Computer Science*.
- [3] Pablo Cordero, Manuel Enciso, Inmaculada P. de Guzmán, and Angel Mora. SLFD logic: Elimination of data redundancy in knowledge representation. *Lecture Notes in Artificial Intelligence 2527, Springer-Verlag*, pages 141–150, 2002.
- [4] Pablo Cordero, Gloria Gutiérrez, Javier Martínez, and Inmaculada P. de Guzmán. A new algebraic tool for automatic theorem provers. *Annals of Mathematics and Artificial Intelligence*, 42(4):369-398, 2004.
- [5] Pablo Cordero, Javier Martínez, Gloria Gutiérrez, and Inmaculada P. de Guzmán. Generalizations of lattices looking at computation. *Discrete Mathematics*, 295(1-3):107-141, 2005.
- [6] Pablo Cordero, Angel Mora, Inmaculada P. de Guzmán and Manuel Enciso. Non deterministic ideal operators: An adequate tool for formalization in Data Bases. *Discrete Applied Mathematics*, in review.

- [7] Manuel Enciso, Angel Mora, J. Marco, José Millán. Herramienta web para integrar la información de diferentes esquemas Oracle. *XV Congreso Nacional Usuarios de Oracle. Octubre de 2005.*
- [8] Manuel Enciso, Inmaculada P. de Guzmán, and Carlos Rossi. Una forma normal temporal independiente del método de deducción. *Inteligencia Artificial*, 8, 23 (2004) pp. 27–46.
- [9] J. Martínez, P. Cordero, G. Gutiérrez, and Inmaculada P. de Guzmán. Restricted ideals and the groupability property: tools for temporal reasoning. *Kybernetika*, 39(5):521-546, 2003.
- [10] Angel Mora, Gabriel Aguilera, Manuel Enciso, Pablo Cordero, and Inmaculada P. de Guzmán. Closure via Functional Dependence Simplification. *Sometido a Journal Acta Informatica. Elsevier. JUNE 2005,*
- [11] Angel Mora, Gabriel Aguilera, Manuel Enciso, Pablo Cordero, and Inmaculada P. de Guzmán. A new closure algorithm based in logic: SLfd-Closure versus classical closures. *Sometido a Conferencia de la Asociación Española para la Inteligencia Artificial (CAEPIA 2005). Lecture Notes in Computer Science. JUNE 2005,*
- [12] Angel Mora, Gabriel Aguilera, Manuel Enciso, Pablo Cordero, and Inmaculada P. de Guzmán. SLfd-logic to obtain atomic-minimality in Functional Dependencies. *Sesión especial de Herramientas matemáticas para la computación. Primer Congreso Conjunto de Matemáticas MAT.ES (RSME-SCM-SEIO-SEMA). Valencia 2005.*
- [13] Angel Mora, Manuel Enciso, Pablo Cordero, and Inmaculada P. de Guzmán. An efficient preprocessing transformation for functional dependencies sets based on the substitution paradigm. *Lecture Notes in Artificial Intelligence, Springer-Verlag*, 3040, 2004, pp. 136–146.
- [14] Angel Mora, Manuel Enciso, Pablo Cordero, and Inmaculada P. de Guzmán. Minimal generator of a non-deterministic operator.. *International Conference of Computational Methods in Sciences and Engineering - ICCMSE 2004. Lecture Series on Computer and Computational Sciences. VSP international Science Publishers. The Netherlands. Series ISSN 1573-4196 ISBN: 90-6764-418-8. , Volume 1, pp. 396-399.*
- [15] Angel Mora, Manuel Enciso, Pablo Cordero, Inmaculada P. de Guzmán and José Guerrero. A/D CASE: A new heart for FD3. *6th International Conference on Enterprise Information System. ICEIS 2004 ICEIS PRESS. Proceedings of the 6th International Conference on Enterprise Information System. ISBN: 972-8865-00-7 PORTO - PORTUGAL, 2004.*
- [16] Manuel Enciso, Carlos Rossi, Angel Mora. A first order temporal logic for behavior representation IBERAMIA *Ibero-American Conference on Artificial Intelligence. Lecture Notes in Artificial intelligence, vol 3315 pp 408–418 Mexico, 2004.*
- [17] Angel Mora, Manuel Enciso, Pablo Cordero, and Inmaculada P. de Guzmán. SLfd logic in Prolog to remove redundancy in database. *International Conference on Computational and Mathematical Methods in Science and Engineering. V - Proceedings of the CMMSE 2005. ISBN: 84-609-4844-7, pp. 390–395.*

- [18] Angel Mora, Manuel Enciso, Pablo Cordero, and Inmaculada P. de Guzmán. Logic treatment of Functional Dependencies in the Relational Model, yet another novelty. *Theory and Applications of Relational Structures as Knowledge Instruments. Meeting of the COST Action 274 Málaga. March 31 - April 2, 2005.*
- [19] Francisco Rodríguez, Pablo Cordero, and Inmaculada P. de Guzmán. Operadores de cierre y razonamiento automático. *In Actas del III Encuentro Andaluz de Matemática Discreta 2003*, pp. 83–86.
- [20] Carlos Rossi, Manuel Enciso, and Inmaculada P. de Guzmán. Formalization of UML state machines using temporal logic. *Software and Systems Modeling*, 3, 1 (2004) 31-54.
- [21] F. Berzal, N.Marin, O.Pons, M.A. Vila A framework to build fuzzy object-oriented capabilities over an existing database using classical object-oriented concepts *Advances in fuzzy object oriented databases: modelling and applications*, chapter 7 . 2004
- [22] F. Berzal, N.Marin, O.Pons, M.A. Vila Managing fuzziness on conventional object oriented platforms *Advances in fuzzy database technology. Int. Journ. of Intelligent Systems.* 2005
- [23] I.Blanco, M.J. Martin-Bautista, O. Pons, M.A. Vila A Tuple-Oriented Algorithm for Deduction in a Fuzzy Relational Database *Int.Journ. of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 11. 2003
- [24] C.D. Barranco, J. Campaña, J.M. Medina, O. Pons InmoSoftWeb: A web based fuzzy application for real state *Lecture Notes in Computer Science*, vol. 3034. 2004

Resolución de Problemas Combinatorios con Restricciones Suaves y Fuzzy Utilizando Algoritmos de Satisfactibilidad TIC2003-00950

Teresa Alsinet *
Universidad de Lleida

Abstract

The main goal of our project is to solve, using SAT technology, combinatorial problems that have both soft and hard constraints in its definition. At the same time such constraints can be crisp or fuzzy. In our approach we will use the language (extended with weights and preference degrees when necessary) of Boolean and many-valued formulas to model the problems. Additionally, we will use (maximum) satisfiability algorithms to find the solutions. In other words, the goal of our project is to design and implement a generic problem solving method for OCSs (over-constrained constraint satisfaction problem) using (maximum) Boolean and many-valued algorithms. To experimentally evaluate our algorithms we will build a benchmark repository with realistic OCS instances and design and implement random instance generators.

Keywords: Satisfiability algorithms, soft and fuzzy constraints.

1 Context and objectives

A constraint satisfaction problem (CSP) is defined as a tuple (V, D, C) , where $X = \{X_1, \dots, X_n\}$ is a set of variables, $D = \{D_1, \dots, D_n\}$ is a set of domains in which D_i denotes the finite set of possible values that can be assigned to the variable X_i and $C = \{C_1, \dots, C_r\}$ is a set of constraints where each constraint denotes a subset of the Cartesian product of the domains of the variables associated with the constraint. A solution to the CSP is an assignment of values to the variables such that all the constraints are satisfied. We have different problems: the decision problem (check if at least one solution exists), finding all solutions and finding the best solution under some quality criteria. For finding solutions of a CSP, systematic and local search algorithms have been developed [17]. A particular class of CSPs, widely studied, is the Satisfiability problem of Boolean CNF formulas (SAT). In this problem variables are boolean, and the constraints are always disjunctions of boolean literals. Different algorithms have been developed for solving SAT: Chaff, RelSAT, WalkSAT and Satz [12, 13, 14] (among others).

*Email: tracy@eps.udl.es

These algorithms, except WalSAT that is a local search algorithm, are systematic algorithms based on the Davis-Putnam algorithm. The differences between both techniques are the data structures used for fast logical unit-propagation, the intelligent branching heuristics used and the kind of backtracking used (chronological versus conflict directed back-jumping). One problem that is between CSP and SAT is the Many-valued SAT problem. This problem is like SAT, but now variables are not boolean because we consider a truth-value set of any cardinality and we have many-valued literals. A many-valued literal is an expression of the form $S : p$ where p is a propositional variable and S denotes the subset of the truth-value set that satisfies the literal. This SAT problem for Many-valued formulas has been widely studied by our research group.

Solving combinatorial problems using CSPs as a constraint language, has one main drawback: a solution of a CSP has to satisfy all the constraints. However, in many real situations, we cannot find such a solution, but there may be some assignments that satisfy *almost* all the constraints and can be considered as acceptable solutions because the constraints not satisfied may not be as *important* as the satisfied by the assignment. These unsolvable CSPs are known as over-constrained CSPs (OCSPs). A solution for a OCSP is an assignment that satisfies in the *best* possible way the constraints of the problem. In this context, we have to consider different kinds of constraints: (i) crisp: constraints involving precise knowledge or information, (ii) fuzzy: constraints involving vague knowledge, (iii) hard: constraints that have to be satisfied in every solution of the OCSP and (iv) soft: constraints that are not needed to be satisfied in every solution of the CSP. In the context of OCSP, there are two relevant problems related to the classical SAT problem: Max-SAT and weighted Max-SAT [15, 16]. Max-SAT is the problem of finding a boolean assignment to the variables of a CNF formula such that the maximum possible number of clauses is satisfied. The clauses are equally important, so we do not have here hard or soft constraints. In contrast, in the weighted Max-SAT, we associate to every clause a "weight" that in some way defines the importance of the clause. Now the goal is to find an assignment with the maximum value of the sum of the weight of the satisfied clauses. The current existing implemented algorithms for Max-SAT and Weighted Max-SAT are not as efficient as the ones for SAT. One reason could be that they are harder to solve than SAT. For example, 2SAT is polynomially solvable, but the corresponding Max-2SAT and Weighted Max-2SAT are NP-hard problems.

The main goal of our project is to solve, using SAT technology, combinatorial problems that have both soft and hard constraints in its definition. At the same time such constraints can be crisp or fuzzy. In our approach we will use the language (extended with weights and preference degrees when necessary) of Boolean and many-valued formulas to model the problems. Additionally, we will use (maximum) satisfiability algorithms to find the solutions. In other words, the goal of our project is to design and implement a generic problem solving method for OCSPs using (maximum) Boolean and many-valued algorithms. To experimentally evaluate our algorithms we will build a benchmark repository with realistic OCSP instances and design and implement random instance generators. Finally, in order to model constraints dealing with fuzzy information and uncertainty (preferences) we have to formalize a language being able to deal with both kinds of human indeterminism; define realistic problems for which such approach will be a suitable representation model; and develop exact satisfiability branch-and-bound algorithms.

To be precise the objectives of the project can be summarized as follows:

- The design and implementation of efficient Max-SAT algorithms for many-valued clausal formulas, and the improvement of exact algorithms for Boolean formulas by means of the use of some new data structures for modeling formulas, new search strategies, non chronological backtracking, and some learning techniques for clauses. The aim is to reduce the search space that should be compute in order to find the optimal solution.
- The design and implementation of a generic method for solving OSCP, based on satisfiability algorithms for Boolean and many-valued formulas. At the same time such constraints can be both hard or soft, and crisp or fuzzy.
- The development of a benchmark repository build from realistic combinatorial problems. The benchmark repository will allow us to evaluate the generic method for solving OSCP.
- The analysis, from logical and computational points of view, of different propositional languages for representing hard, soft, crisp and fuzzy constraints.

2 Contributions of the project

2.1 Max-SAT solvers

When we started the project, there were a wide variety of extremely efficient SAT solvers, but very few exact Max-SAT solvers implementing the Davis-Putnam procedure with a branch and bound scheme [3, 8, 10]. Nowadays, there are a considerable number of fast Max-SAT solvers [5, 6, 9, 19, 20] and, roughly speaking, a Max-2SAT instance with 100 variables is currently solved 1000 times faster than five years ago. Basically, the best existing solvers improve previous solvers in the following aspects: (i) efficient data structures for representing and manipulating formulas, (ii) lower bound computation methods that provide lower bounds of better quality, (iii) clever variable selection heuristics, and (iv) powerful inference techniques.

Our first contribution was to show that Borchers and Furman algorithm (BF) could be improved by incorporating a new version of the Jeroslow-Wang rule [11], and considering inconsistency counts in the computation of the lower bound [4, 5]. We refer to that solver as AMP.

We then focused our research on developing a faster solver than AMP. Our first steps was to define lazy data structures inspired by the two-watched literals data structures of zChaff [21]. We defined very simple data structures that allow one to efficiently traverse the search space and, at the same time, to apply inference techniques and compute lower bounds very quickly. The only constraint of our data structures is that we have to apply a static variable selection heuristic. We refer to this solver as Lazy (Publications [1] and [2]). Besides the data structures, Lazy implements an original lower bound of better quality and applies neighborhood resolution as a preprocessing technique. We have developed both a weighted (Publication [1]) and an unweighted version (Publication [2]).

In our last contribution (Publication [18]), we have designed and implemented a Max-SAT solver that makes a novel use of unit propagation: it is used to compute lower bounds. Our approach subsumes existing lower bound computation methods like those based on inconsistency counts. It is worth to point out that our solver is the best performing state-of-the-art solver, at least, on randomly generated Max-2SAT and Max-3SAT instances.

We are currently working in two directions: On the one hand, we are incorporating more powerful inference techniques into our solvers. In particular, we have defined a unit resolution rule that ensures optimality (the standard unit resolution rule discards optimal solutions). On the other hand, we are designing and implementing Max-SAT solvers for multiple-valued clausal forms.

2.2 Solving over-constrained solvers with SAT technology

We developed a new generic problem solving approach for over-constrained problems based on Max-SAT. To this end, we defined a Boolean clausal form formalism, called *soft CNF formulas*, that deals with blocks of clauses instead of individual clauses, and that allows one to declare each block either as *hard* (i.e., must be satisfied by any solution) or *soft* (i.e., can be violated by some solution). An optimal solution for a given soft CNF formula consists of finding a truth assignment that satisfies all the hard blocks and the maximum number of soft blocks.

We designed and implemented two solvers for Soft CNF formulas (Publications [5] and [6]). Our solvers are branch and bound algorithms equipped with original lazy data structures, powerful inference techniques, lower bounds of good quality, and original variable selection heuristics. We tested our solvers on a representative sample of instances (random 2-SAT, Max-CSP, graph coloring, quasigroup completion, and pigeon hole). Since we exploit more the structure of the problem than Max-SAT approaches, both our inference techniques and our lower bound computation method give raise to better performance profiles. The gains, compared with Max-SAT, are up to two orders of magnitude. Besides, we compared our solvers with Max-CSP solvers, and observed that our solvers are very competitive (an in many cases superior) to Max-CSP solvers.

We are currently working on the extension of the formalism of Soft CNF formulas in order to include fuzzy constraint; only crisp constraints are treated in the existing solvers of Soft CNF formulas.

2.3 SAT and CSP problems typical-case complexity

Usually, one defines the complexity of a Satisfiability problem over clausal forms (CNFs) with Many-valued (signed) literals in which all the literals have the same syntactic form [7]. However, when using these signed CNF formulas as a constraint language to encode some real-world combinatorial problem (frequency assignment, scheduling, timetabling, ...) not all the clauses of the resulting CNF have literals with the same form.

So, beyond understanding the worst-case complexity of signed CNF formulas with a rigid syntactic structure, in our work (Publication [4]) we start the study of the effect of the presence of different fractions of clauses in a Signed CNF with a particular structure in their signed literals.

We consider two different kinds of signed literals: signed literals whose classical Many-valued satisfiability problem (the problem in which all the literals have the same form) is tractable (polynomially solvable) and signed literals for which the same problem is NP-complete. Then, we consider the Satisfiability problem for CNFs in which a fraction p of the clauses contain "NP-complete literals" (from the second kind) and a fraction $(1-p)$ of the clauses contain "tractable literals" (from the first kind). In the two combinations we have tested in the paper, what we find is that even if the problems are NP-complete for any $p > 0$, the typical-case

complexity (average complexity) of solving a test-set of instances is not the same for any value of p . For one of the problems, there is a critical value of p that separates two very different regions: one in which all the problems are solved with polynomial average time and the other in which all the problems are solved with exponential average time.

In conclusion, when facing the question of which constraint language one should use to encode a combinatorial problem one needs to solve, looking only at worst-case complexity results is not a sufficient information to take into account.

2.4 A benchmark for distributed CSPs

A distributed CSP problem is a Constraint satisfaction problem which is divided in different subproblems. Every such subproblem belongs to an unique “agent” that is the only one allowed to change the values of the variables of the subproblem. We have two different kinds of constraints: intra-agent constraints (constraints between variables of a same subproblem) and inter-agent constraints (constraints between variables of different subproblems).

In our work (Publication [8]) we consider the worst-case and typical-case complexity of solving the distributed constraint satisfaction problem (DisCSP). We perform the study by first introducing two realistic benchmark problems: SensorDCSP and GSensorDCSP, and studying their worst-case complexity. These problems are based on a real-world application that arises in the context of networked distributed resource allocation systems for the tracking of mobile objects. To perform a realistic analysis of the typical-complexity, we have used a discrete-event network simulator, which allows us to model the impact of different network traffic conditions of the performance of the (distributed) algorithms. We use two well known DisCSP algorithms, ABT and AWC, although as the result of one of our empirical results we propose an improvement for the ABT algorithm.

One important result obtained in the typical-case complexity analysis we have performed is that random delays (due to network traffic or in some cases actively introduced by the agents) can affect the performance of the algorithms in very different ways, because we have found that AWC is considerably more robust to the variance of the delays than ABT.

As a consequence, when designing DisCSP algorithms one wants to use in real communication networks, one should take into account the sensibility of the algorithms to the networks conditions that the algorithm is going to find.

2.5 On the formalization of a propositional language for representing fuzziness and uncertainty

In the last years defeasible argumentation frameworks have proven to be a successful approach to formalizing qualitative, commonsense reasoning from incomplete and potentially inconsistent knowledge. Defeasible Logic Programming (or DeLP) [18] is one of such formalisms, combining results from defeasible argumentation theory and logic programming. Although DeLP has proven to be a suitable framework for building real-world applications that deal with incomplete and contradictory information in dynamic domains, it cannot deal with explicit uncertainty, nor with vague knowledge, as defeasible information is encoded in the object language using “defeasible rules”.

In the framework of the project we have formalized P-DeLP (Publications [14] and [16]), a new logic programming language that extends original DeLP capabilities for qualitative

reasoning by incorporating the treatment of possibilistic uncertainty and fuzzy knowledge. Such features are formalized on the basis of PGL [1, 2], a possibilistic logic based on the Horn-rule fragment of Gödel fuzzy logic. In PGL formulas are built over fuzzy propositional variables and the certainty degree of formulas is expressed by means of a necessity measure. In a logic programming setting, the proof method for PGL is based on a complete calculus for determining the maximum degree of possibilistic entailment of a fuzzy goal.

In the context of complex logic-programming frameworks (like the one provided by extended logic programming), PGL lacks of an adequate mechanism to handle contradictory information, as conflicting derivations can be found. In P-DeLP such conflicts are solved using an argument-based inference engine. Formulas are supported by arguments, which have an attached necessity measure associated with the supported conclusion. The ultimate answer to queries is given in terms of warranted arguments, computed through a dialectical analysis. One particularly interesting feature of P-DeLP is the possibility of defining aggregated preference criteria by combining the necessity measures associated with arguments with other syntax-based criteria.

Finally, Defeasible argumentation in general and P-DeLP in particular provide a way of modeling non-monotonic inference. From a logical viewpoint, capturing defeasible inference relationships for modeling argument and warrant is particularly important, as well as the study of their logical properties. In our work (Publication [15]) we have analyzed two non-monotonic operators for P-DeLP which model the expansion of a knowledge base by adding new weighed facts associated with argument conclusions and warranted literals, resp. Different logical properties for the proposed expansion operators have been studied and contrasted with a traditional SLD-based Horn logic. One particularly interesting feature is that this analysis provides useful comparison criteria that can be extended and applied to other argumentation frameworks.

3 Publications and collaborations

The results obtained in the framework of the project have already been published as journal articles or in conference proceedings:

- [1] T. Alsinet, F. Manyà and J. Planes. *Improved Exact Solver for Weighted Max-SAT*. Proceedings of the Eighth International Conference on the Theory and Applications of Satisfiability Testing, SAT-2005, St. Andrews, Scotland, LNCS 3569, pp. 371-378, 2005.
- [2] T. Alsinet, F. Manyà and J. Planes. *A Max-SAT Solver with Lazy Data Structures*. Proceedings of the IX Ibero-American Conference on Artificial Intelligence, IBERAMIA-2004, Puebla, México, Springer LNCS 3315, pp. 334-342, 2004.
- [3] C. Ansótegui and F. Manyà. *Mapping Problems with Finite-Domain Variables to Problems with Boolean Variables*. Proceedings of the Seventh International Conference on the Theory and Applications of Satisfiability Testing, SAT-2004, Vancouver, Canada, Springer LNCS 3542, pp. 111-119, 2004.
- [4] C. Ansótegui, R. Béjar, A. Cabiscol and F. Manyà. *The Interface between P and NP in Signed CNF Formulas*. Proceedings of the 34th International Symposium on Multiple-Valued Logics (ISMVL), Toronto, Canada. IEEE CS Press, pp. 251-256, 2004.
- [5] J. Argelich and F. Manyà. *Solving Over-constrained Problems with SAT Technology*. Proceedings of the Eighth International Conference on the Theory and Applications of satisfiability Testing, SAT-2005, St. Andrews, Scotland, LNCS 3569, pp.1-15, 2005.

- [6] J. Argelich and F. Manyà. *Solving Over-Constrained Problems with Max-SAT Algorithms*. Proceedings of the Workshop on Modelling and Solving Problems with Constraints, 16th European Conference on Artificial Intelligence, ECAI-2004, Valencia, Spain, pp. 116-124, 2004.
- [7] J. Argelich and F. Manyà. *An Exact Max-SAT Solver for Over-Constrained Problems*. Proceedings of the Workshop on Preferences and Soft Constraints, 10th International Conference on Principles and Practice of Constraint Programming, CP-2004, Toronto, Canada, 2004
- [8] R. Béjar, C. Domsلاك, C. Fernández, C. Gomes, B. Krishnamachari, B. Selman and M. Valls. *Sensor Networks and Distributed CSP: Communication, Computation and Complexity*. Artificial Intelligence, 161: 117-147, 2005.
- [9] M. Capobianco, C.I. Chesñevar and G. Simari. *An Argument-based Framework to Model an Agent's Beliefs in a Dynamic Environment*. First Internacional Workshop on Argumentation in Multiagent Systems. Springer LNCS, Vol. 3366, pp.95-110, 2005.
- [10] C. Chesñevar, R. Brena and J. Aguirre. *Knowledge Distribution in Large Organizations Using Defeasible Logic Programming*. 18th Canadian Conference in Artificial Intelligence, Victoria, British Columbia, Canada. Springer LNAI 3501, pp. 244-256, 2005.
- [11] C.I. Chesñevar and A.G. Maguitman. *An Argumentative Approach to Assessing Natural Language Usage based on the Web Corpus*. European Conference on Artificial Intelligence (ECAI 2004), pp. 581-585. Valencia, Spain, 2004.
- [12] C.I. Chesñevar and A.G. Maguitman. *Combining Argumentation and Web Search Technology: Towards a Qualitative Approach for Ranking Results*. Intl. Journal of Advanced Computational Intelligence & Intelligent Informatics, Vol. 9, No. 1, pp. 53-60, 2005.
- [13] C.I. Chesñevar, A.G. Maguitman and G. Simari. *A first Approach to Argument-Based Recommender Systems based on Defeasible Logic Programming*. International Workshop on Non Monotonic Reasoning (NMR 2004), pp. 109-117 Whistler, Canada, 2004.
- [14] C. Chesñevar, G. Simari, T. Alsinet and L. Godo. *A Logic Programming Framework for Possibilistic Argumentation with Vague Knowledge*. Uncertainty in Artificial Intelligence (UAI-2004), pp. 76-84, ISBN 0-9749039-0-6. Banff, Canada, 2004.
- [15] C. Chesñevar, G. Simari, L. Godo and T. Alsinet. *Argument-based Expansion Operators in Possibilistic Defeasible Logic Programming: Characterization and Logical Properties*. 8th European Conference on Symbolic and Qualitative Approaches to Reasoning with Uncertainty (ECSQARU-2005), Barcelona, Spain. LNAI 3571, pp. 353-365, 2005.
- [16] C. Chesñevar, G. Simari, T. Alsinet and L. Godo. *Modelling Agent Reasoning in a Logic Programming Framework for Possibilistic Argumentation*. 2nd European Workshop on Multiagent Systems (EUMAS 2004), pp. 135-142. Barcelona, Spain, 2004.
- [17] S.A. Gómez and C.I. Chesñevar. *Integrating Defeasible Argumentation with Fuzzy ART Neural Networks for Pattern Classification*. Journal of Computer Science and Technology, Vol. 4, No. 1, pp.45-57, 2004.
- [18] C. M. Li, F. Manyà, and J. Planes. *Exploiting unit propagation to compute lower bounds in branch and bound max-sat solvers*. In 11th International Conference on Principles and Practice of Constraint Programming, CP-2005, Sitges, Spain. Springer LNCS, 2005.

Some of the above publications are the result of collaborations with national and international researchers. To design and implement Max-SAT algorithms we have collaborated with Chu Min Li of the University of Picardie. To study the heavy-tails phenomena and distributed CSPs we have collaborated with Carla Gomes and Bart Selman of the University of Cornell. Finally, to formalize a language for representing fuzzy and uncertain information in the framework of argumentative systems we have collaborated with Lluís Godo of the IIIA-CSIC and Guillermo Simari of the University Nacional del Sur.

Finally, we would remark that Jordi Planes and Josep Argelich are developing their PhD. Thesis in the framework of the project.

References

- [1] Alsinet, T. and Godo, L. A complete calculus for possibilistic logic programming with fuzzy propositional variables. In *Proc. of the UAI-2000 Conference*, pages 1–10, 2000.
- [2] Alsinet, T. and Godo, L. A proof procedure for possibilistic logic programming with fuzzy constants. In *Proc. of the ECSQARU-2001 Conference*, pages 760–771, 2001.
- [3] J. Alber, J. Gramm, and R. Niedermeier. Faster exact algorithms for hard problems: A parameterized point of view. In *25th Conf. on Current Trends in Theory and Practice of Informatics*, LNCS, pages 168–185. Springer-Verlag, November 1998.
- [4] T. Alsinet, F. Manyà, and J. Planes. Improved branch and bound algorithms for Max-2-SAT and weighted Max-2-SAT. In *Proceedings of the Catalanian Conference on Artificial Intelligence*, 2003.
- [5] T. Alsinet, F. Manyà, and J. Planes. Improved branch and bound algorithms for Max-SAT. In *Proceedings of the 6th International Conference on the Theory and Applications of Satisfiability Testing, SAT-2003, Portofino, Italy*, 2003.
- [6] T. Alsinet, F. Manyà, and J. Planes. Improved exact solvers for weighted Max-SAT. In *Proceedings of the 8th International Conference on the Theory and Applications of Satisfiability Testing, SAT-2005, St. Andrews, Scotland, UK*, pages 371–367. Springer LNCS 3569, 2005.
- [7] R. Bejar, A. Cabiscol, C. Fernandez, F. Manyà and C. Gomes. Capturing Structure with Satisfiability. *7th International Conference of Constraint Programming, CP-2001*. Paphos. Cyprus. LNCS 2239 pp. 137-152, 2001.
- [8] B. Borchers and J. Furman. A two-phase exact algorithm for MAX-SAT and weighted MAX-SAT problems. *Journal of Combinatorial Optimization*, 2:299–306, 1999.
- [9] S. de Givry, J. Larrosa, P. Meseguer, and T. Schiex. Solving Max-SAT as weighted CSP. In *9th International Conference on Principles and Practice of Constraint Programming, CP-2003, Kinsale, Ireland*, pages 363–376. Springer LNCS 2833, 2003.
- [10] E. Freuder and R. Wallace. Partial constraint satisfaction. *Artificial Intelligence*, 58:21–71, 1992.
- [11] R. G. Jeroslow and J. Wang. Solving propositional satisfiability problems. *Annals of Mathematics and Artificial Intelligence*, 1:167–187, 1990.
- [12] C.M. Li and Anbulagan. Look-ahead versus Look-back for Satisfiability Problems. *International Conference on Principles and Practice of Constraints Programming, CP-1997*, 1997.
- [13] M. Moskewicz, C. Madigan, Y. Zhao, L. Zhang and S. Malik. Chaff: Engineering an Efficient SAT Solver. 39th Design Automation Conference, 2001.
- [14] B. Selman, H. Kautz and B. Cohen. Noise Strategies for Improving Local Search. *National Conference on Artificial Intelligence, AAAI-96*, 1996.
- [15] S. Joy, J. Mitchell and B. Borchers Solving MAX-SAT and Weighted MAX-SAT problems using Branch-and-Cut. *Journal of Combinatorial Optimization*, 2003.
- [16] B. Borchers and J. Furman. A two-phase exact algorithm for MAX-SAT and weighted MAX-SAT problems. *Journal of Combinatorial Optimization*, 2:299–306, 1999.
- [17] I. Miguel and Q. Shen (2001). *Solution Techniques for Constraint Satisfaction Problems: Foundations*. Artificial Intelligence Review 15, pp. 243-267.
- [18] García, A. and Simari, G. Defeasible Logic Programming: An Argumentative Approach. *Theory and Practice of Logic Programming*, 4(1):95–138, 2004.
- [19] Z. Xing and W. Zhang. Efficient strategies for (weighted) maximum satisfiability. In *Proceedings of CP-2004*, pages 690–705, 2004.
- [20] H. Zhang, H. Shen, and F. Manyà. Exact algorithms for MAX-SAT. In *4th Int. Workshop on First order Theorem Proving*, June 2003.
- [21] L. Zhang, C. Madigan, M. Moskewicz, and S. Malik. Efficient conflict driven learning in a Boolean satisfiability solver. In *International Conference on Computer Aided Design, ICCAD-2001, San Jose/CA, USA*, pages 279–285, 2001.

Study of the Systems with Uncertainty in Presence of a Fuzzy Similarity TIC2003-04564

J. Recasens*
Sec. Matemàtiques i Informàtica
ETSAV
Universitat Politècnica de Catalunya

Abstract

In real problems, there are basically two kind of uncertainty: uncertainty due to the lack of knowledge with respect to well defined objects, but for which there are different hypothetical situations, and uncertainty due to the intrinsic vagueness of the objects. Traditionally, the first type of uncertainty has been studied with the help of Probability Theory while the second one has used Fuzzy Set Theory and Possibility Theory. In situations where the two kind of uncertainty are present, theories capable of handling the probability and the possibility of fuzzy events together are needed. On the other hand, many times the vagueness of the objects is due to the existence of a fuzzy equality that makes indistinguishable different objects of the universe. This can happen, for example, because of the imprecision of the measurement tools. Indistinguishability operators or similarities model this kind of equalities in a satisfactory way.

The main target of this project is to formulate a general framework that allows us to deal with different types of uncertainty in universes where a similarity between their objects is defined in a coherent way.

This general target can be divided in the following three partial ones:

- a) Study of probability distributions compatible with a given similarity relation.
- b) Generation of similarity relations on a universe starting from a similarity defined on a set of prototypes of this universe.
- c) Extending the way normalized fuzzy sets generate possibility distributions to the non-normalized case.

Keywords: Fuzzy relation, Similarity, Uncertainty, Fuzzy sets, Random sets, Possibility Theory, Probability Theory.

1 Objectives of the project

For many years, members of our department belonging to this research project have focused their research on Fuzzy logic, especially fuzzy relations and indistinguishability operators. After a long period of theoretical studies and results, there have been some attempts to apply them to a number of branches of AI such as Approximate Reasoning or Data Mining obtaining interesting results that have encouraged the team to find new applications of these relations and to generate new models that allow the treatment of real situations where there are different kinds of uncertainty. Since many times part of the uncertainty on a system is due to the existence of a vague equality that

* Email: j.recasens@upc.edu

does not allow to distinguish clearly some elements of the universe of discourse, indistinguishability operators seem a good tool to study these systems. Hence, the main goal of this project is the study of systems in which there are different types of uncertainty and to define a general framework for their study.

The project has been divided in three realistic subprojects with the idea of spending roughly a year to each one of them:

1. Study of probability distributions compatible with an indistinguishability operator.
 - 1.1. The study of compatibility relations between a probability distribution and an indistinguishability operator defined on the same universe of discourse.
 - 1.2. Generating probability distributions starting from an indistinguishability operator in a coherent way both from the theoretical and applied points of view.
 - 1.3. Some toy examples.
2. Generation of an indistinguishability operator on a universe starting from an indistinguishability operator defined on a set of prototypes.
 - 2.1. Study of the idea of prototype.
 - 2.2. Searching and characterization of Fuzzy sets that can be considered as prototypes of a concept when there is an indistinguishability operator on a universe.
 - 2.3. Finding methods to generate an indistinguishability operator on a universe when there is a family of prototypes with an indistinguishability operator on it and some matching degrees between the elements of the universe and the prototypes modelled by fuzzy subsets.
 - 2.4. Applicability to real problems.
3. Generalization of the possibility measures to non-normalized fuzzy sets.
 - 3.1. Relation between non-normalized fuzzy sets and (normalized) possibility distributions.
 - 3.2. Generation of normalized fuzzy sets and possibility distributions from non-normalized fuzzy sets.
 - 3.3. Normalizations of belief measures when the empty set has non-zero mass assignment.
 - 3.4. Generation of indistinguishability operators from belief and plausibility functions.

2 Success of the project

The main goals for the two first years were

- a) The study of the probability distributions compatible with an indistinguishability operator.
- b) Generation of an indistinguishability operator on a universe starting from an indistinguishability operator defined on a set of prototypes.

There have been advances in both directions:

The study of indistinguishability operators in relation with the Theory of Evidence of Dempster-Shafer has allowed us to define indistinguishability operators from belief and plausibility measures and in particular from probability and possibility distributions.

A new way to normalize probability distributions obtained by combining two sources of evidence has been developed. Namely, finding the normalized probability distribution most similar to the obtained one. The results have allowed us to justify a number of well known methods like Demster's rule or based on the cross entropy and to find new ones.

Starting on from prototypes, we have found indistinguishability operators that take the algebraic and the ordering structure of the universe of discourse into account. This is especially interesting in applications where the universes are intervals of the real line or Cartesian products of them.

Another promising point consisting of the generation of aggregation operators from indistinguishability operators has been developed. The basic idea is that in many situations if we want to aggregate the objects a and b we are looking for another object as similar to a than to b, or in a more logical language, with the same degree of equivalence or equality to a than to b.

A more theoretical task has been the study of the structure of indistinguishability operators from their isometry group.

3 Indicators

There have been published three journal articles, 9 conference papers and three more articles are being reviewed. They are referenced bellow.

The team of the project has been in touch with the AI Group of the University of Bristol. Dr Jonathan Lawry, a member of this group, has been with us in a sabbatical for three months last year.

Articles:

Demirci, M., Recasens, J. Fuzzy Groups, Fuzzy Functions and Fuzzy Equivalence Relations. Fuzzy Sets and Systems 144 (2004) 441-458.

Jacas, J., Recasens, J. The Group of Isometries of an Indistinguishability Operator. Fuzzy Sets and Systems 146 (2004) 27-41.

Hernández, E., Recasens, J. Indistinguishability Relations in Dempster-Shafer Theory of Evidence. *International Journal of Approximate Reasoning* 37 (2004) 145-187.

Lawry, J., Recasens, J., González Rodríguez, I. A Similarity Approach to Evidence Combination in Dempster-Shafer Theory. *Sent to Artificial Intelligence*.

Jacas, J., Recasens, J. Aggregation Operators based on Indistinguishability Operators. *Sent to Int. Journal of Intelligent Systems*.

Boixader, D., Recasens, J. A Map Characterizing the Fuzzy Points and Columns of a T-indistinguishability operator. *Sent to International Journal of Intelligent Systems*

Conferences

Jacas, J., Recasens, J. Aggregation of T-transitive Relations using OWA Operators. *Proceedings IFSA 2003 Congress. Istanbul. (2003)*

Jacas, J., Recasens, J. Aggregating non-finite families of t-transitive relations *Proceedings EUSFLAT 2003 Congress. Zittau. (2003). 429-432.*

Jacas, J., Recasens, J. Indistinguishability Operators Generated by Prototypes. *IPMU2004. Perugia. (2004).*

Jacas, J., Recasens, J. Indistinguishability Operators Generated by Fuzzy Numbers. *FUZZ'IEEE2004. Budapest. (2004).*

Jacas, J., Recasens, J. Preórdenes Duales, Operadores de Clausura y Contextos Borrosos. *ESTYLF2004. Jaén. (2004).*

Jacas, J., Recasens, J. A Similarity-based Approach to Aggregation. *FUZZ'IEEE2005. Reno. (2005).*

Tomás, María Santos, Pseudometrics and 3-positive semidefinite similarities, *43 ISFE, Nantes. (2005)*

Jacas, J., Recasens, J. Aggregating with Indistinguishability operators. *EUSFLAT2005. Barcelona. (2005).*

Hernández, E., J., Recasens, J. Indistinguishability in Cooperative Games. *EUSFLAT2005. Barcelona. (2005).*

Development and Evaluation of Metaheuristics for Constrain Satisfaction Problems

TIC2003-04153

Camino R. Vela¹, Jorge Puente, Cesar Alonso, Inés González, María Sierra and
Ramiro Varela

Grupo de Tecnologías de la Computación.
Centro de Inteligencia Artificial. Departamento de Informática.
Universidad de Oviedo, Campus de Viesques.
33271 Gijón, Spain

<http://www.aic.uniovi.es/Tc>

1 Introduction

The main purpose of this project is to develop advanced techniques and tools for solving constraint satisfaction problems (CSPs), with a special interest in scheduling problems. It is well known that many real problems belong to this family and that scheduling problems are in fact a paradigm of constraint satisfaction problems. Consequently developing general purpose strategies for these problems is an interesting issue.

In this project, we plan to exploit a number of artificial intelligence techniques such as heuristic, evolutionary algorithms, state space search, parallel processing and models to cope with uncertain knowledge. We will try to combine them in order to reach a threshold of efficiency that allows us to obtain effective solutions to scheduling problems. The development of systematic methods to combine heuristics for complex problems is expected to be one of the main contributions from the methodological point of view.

In addition to the formal techniques, another relevant contribution of the proposed project will be the implementation of tools incorporating these techniques, which will allow us to work on massive sets of experiments. In order to do that, these tools must be enhanced with both a user interface for experiments setup and experimental results management, as well as an execution environment that distributes the processes on a cluster of heterogeneous processors.

For experimental studies, we will use problem instances taken from conventional repositories, as well as problems with particular characteristics. In the case when problems of this class are not available, we will make our own proposal.

¹ Email: camino@aic.uniovi.es

2 Main Objectives

The main objectives of this research project can be summarized as follows

2.1 New strategies for Scheduling problems with Genetic Algorithms and State Space Search

- New seeding methods for solving the Job Shop Scheduling (JSS) problem with GAs
- Combining Local Search, Structured Populations and Heuristic Seeding in GAs
- Solving Scheduling problems other than JSS, such as Open Shop and Flow Shop
- Solving the JSS problem with Branch and Bound and A* like algorithms. Heuristics development for these search paradigms, and study of alternatives that require less computational resources

2.2 Implementation of a tool for massive experimentation with GAs in a distributed hardware

- Here our objective is incorporating the standard models of parallelism in GAs: farm and islands, into our logical model of GAs. We extend our GA object model in this task by using standard parallel environments such as MPI or PVM.

2.3 Parallel Implementation of GAs

- Here our objective is adapting and exploiting a parallel virtual machine that we have developed in a previous project (to implement a parallel interpreter of logic programs) for parallel execution of GAs.

2.4 Study of some formal aspects of GAs for Scheduling

- Here the main goal is to compare various codification schemas suitable for Scheduling and related problems, trying to explain the different be-

haviour between them, and if possible design new and better codification schemas

2.5 Design and Implementation of a prototype tool for production scheduling

- The objective is to develop a general purpose scheduler that could be adapted to standard real manufacturing industries or workshops, similar to some current commercial tools

2.6 GAs for problems other than Scheduling

- Here we aim at applying the same or similar techniques used for Scheduling to other problems of combinatorial nature such as, for example, the Knapsack Problem, String Equations or Vehicle Routing

2.7 CSP formulation is the framework of Possibility Theory

- First we will try to formulate some heuristics, in particular the variable and value ordering heuristics for JSS, within this theory in order to compute them more efficiently
- Moreover we will confront CSPs problems with uncertainty or flexible constraints, for example JSS problems with uncertain durations and flexible due dates. Here we will try to define a clear semantics for the solutions to the problems, and generalize the methods used with crisp problems

3 Main Changes with Respect to the Initial Objectives

From some suggestions of the project referees we have done a change in the previous objectives in order to confront some truly real problems in addition to the academic ones, so making this project more useful in practice. We have opted for eliminating objective 2.5, and look for some real problems in our industrial environment instead. The result is that we have contacted two different companies that proposed two interesting problems for us to solve. First, the Spanish rail company (RENFE) has proposed to us a challenging problem that consists in organizing the timetable of train drivers. The second company manufactures rolls of plastic film. In this case the problem consists in optimizing the cut plans in order to minimize the amount of plastic that has to be thrown away as well as the cutting machine arrangements between different cuts. In both cases we have a multiobjective CSP.

4 Summary of main results up to date

As many of the project tasks require much programming effort, we have offered a great part of this work to undergraduate students on Computer Science, so that they have the opportunity of doing their graduation project at the same time that, in many cases, provide us a valuable collaboration and gain some experience in research activities. We also participate in doctoral program of the Department of Computer Science of the University of Oviedo, hence we also supervise doctoral students that collaborate with us in doing their research work and Phd. In the three following subsections we indicate the set of undergraduate student projects, the papers we have published related to this research project, and the doctoral students that collaborate with us.

4.1 Student projects

The following graduation projects were carried out by the indicated undergraduate students within the scope of this research project

- Resolución de Problemas de Scheduling con el Algoritmo A*. *María R. Sierra Sánchez*
- Un Algoritmo Genético para el problema de planificación de proyectos con tareas multimodales. *Lorenzo García Suárez*
- Planificación de turnos de trabajo de personal de conducción de trenes de mercancías. *Juan Fernández Rodríguez*
- Resolución de Problemas de Planificación con búsqueda tipo Backtracking restringida al espacio de planificaciones activas. *Jorge Sánchez Rodríguez*
- Un modelo probabilista para incorporar conocimiento heurístico en un algoritmo genético. *Javier Blanco Rodríguez*
- Representaciones extendidas en Algoritmos Genéticos. *David Serrano Riestra*
- Diseño de un Algoritmo Memético para Problemas de Scheduling con Representación Gráfica Interactiva. *David Moreno López*
- Un algoritmo evolutivo para el problema Field Technician Scheduling Problem. *Andrés Cifuentes Pla*
- Análisis, Monitorización y Síntesis de Poblaciones en Algoritmos Genéticos. *Daniel González Fernández*
- Combinación del entorno LEKIN con un Algoritmo Genético para problemas de Scheduling. *Eduardo García Pérez*
- Monitorización de Algoritmos Evolutivos. *Elias Alvarez Granda*

- Algoritmos evolutivos para el problema de Asignación de Frecuencias. *Francisco González Bulnes*
- Desarrollo de una librería de clases para aprendizaje con sistemas borrosos. *Manuel Espina Álvarez*
- Desarrollo de una librería de clases para aprendizaje de descripciones Lingüísticas. *Miguel Ángel Pastor Olivar*
- Cálculo de soluciones aproximadas y cotas inferiores para el problema de secuenciamiento de una máquina. *Miguel Angel González Fernández*
- Editor Web de planificaciones de Tareas. *Hernán Díaz Rodríguez*
- Entorno Web de ejecución distribuida de algoritmos genéticos. *Antonio Espín Ballesta*
- Un Algoritmo de Ramificación y Poda para Problemas de Scheduling. *Omar Enrique Somoano Fernández*
- Algoritmos Genéticos y Búsqueda en Espacios de Estados para el problema DCMST. *Beatriz Cueto*
- Estudio de Operadores Genéticos e Inicialización Heurística para el problema Flow Shop Scheduling. *David Fernández*
- Estudio de la Búsqueda Local para el problema Job Shop Scheduling. *Rubén García*
- Sistema de Ayuda a la Enseñanza de los Algoritmos Genéticos. *Covadonga Flórez*
- Algoritmos evolutivos para la resolución de sistemas de “Word Equations. *Fátima Drubi Vega*.
- Resolución de sistemas de ecuaciones de palabras mediante algoritmos evolutivos. *Davinia Jurado Santiago y Leticia Reigada Suárez*.

4.2 Published papers

In addition to some papers that are under review process in some journals and conferences, we have published the following papers in the last two years which are related to the objectives of this research project

A new Chromosome Codification for Scheduling Problems

Varela, R., Vela, C. R, Puente, J., Serrano, D., Suárez, A.

Information Processing with Evolutionary Algorithms: From Industrial Applications to Academic Speculations. Ed. Springer, 2004.

Induction of Fuzzy Prototypes with Feature Selection

González Rodríguez, I., Lawry, J.

Soft Methodology and Random Information Systems (Advances in Soft Computing Series). López-Díaz, M., Gil, M.A., Grzegorzewski, P., Hryniewicz, O., Lawry, J. (eds.), Springer-Verlag, pp. 471-478, 2004.

Heuristic Rules and Genetic Algorithms for Open Shop Scheduling Problems

Puente J.; Díez, H.R.; Varela, R.; Vela, C. R.; Hidalgo; L.P.

Current Topics in Artificial Intelligence, LNAI-3040. Springer-Verlag, pp. 394-403, 2004.

Some Issues in Chromosome Codification for Scheduling with Genetic Algorithms.

Varela, R., Puente, J. and Vela, C. R.

Workshop on Constraint Satisfaction Techniques for Planning and Scheduling Problems. ECAI 2004. Valencia, 24-27 Agosto 2004.

Word Equation Systems: The Heuristic Approach

Alonso, C. L.; Drubi, F.; Gómez García, J.; Montaña, J. L.

Advances in Artificial Intelligence- SBIA 2004 LNCS (Serie LNAI 3171) Eds. Ana L. C. Bazzan and Sofiane Labidi. Springer. pp. 83-92. 2004

An Evolutionary Strategy for the Multidimensional 0-1 Knapsack Problem Based on Genetic Computation of Surrogate Multipliers

Alonso, C.L., Caro, F., Montaña, J.L.

Edited by: J. Mira and J.R. Álvarez. IWINAC 2005, LNCS 3562 (ISBN: 3-540-26319-5), Springer-Verlag, pp. 63-73, 2005.

New Codification Schemas for Scheduling with Genetic Algorithms

Varela, R., Serrano, D., Sierra, M.

Edited by: J. Mira and J.R. Álvarez. IWINAC 2005, LNCS 3562 (ISBN: 3-540-26319-5), Springer-Verlag, pp. 11-20, 2005.

An Evolutionary Approach to Designing and Solving Fuzzy Job-Shop Problems

González Rodríguez, I., Vela, C. R, Puente, J.

Edited by: J. Mira and J.R. Álvarez. IWINAC 2005, LNCS 3562 (ISBN: 3-540-26319-5), Springer-Verlag, pp. 74-83, 2005.

Búsqueda Heurística para Problemas de Scheduling

Sierra, M.; Varela, R.

IX Congreso de Ingeniería de Organización, Gijón, 8 y 9 de Septiembre, 2005.

Optimal Scheduling with Heuristic Best First Search

Sierra, M., Varela, R.

9th congress of the Italian Association of Artificial Intelligence (LNCS), Sep. 23-25, Milano, 2005

Some Issues in Chromosome Codification for Scheduling with Genetic Algorithms

Varela, R., Puente, J., Vela, C. R

Planning, Scheduling and Constraint Satisfaction: From Theory to Practice. Volume 117 Frontiers in Artificial Intelligence and Applications

Edited by: L. Castillo, D. Borrajo, M.A. Salido and A. Oddi. (ISBN: 1586034847) IOS Press, pp. 1-10, 2005.

Un algoritmo meta-heurístico para la resolución de problemas $P//\Sigma Ci$

Noelia González, Camino R. Vela

IV Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados, MAEB'2005, Granada, 2005

Un algoritmo evolutivo para el Problema de la Mochila Multidimensional basado en el uso de un esquema de aproximación polinomial para el problema unidimensional

Alonso, C.L., Caro, F., Montaña, J.L.

IV Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados, MAEB'2005, Granada, 2005

4.3 Doctoral students

Currently we are supervising two doctoral students: Ms. María Sierra and Ms. Noelia Sánchez. Both of them have completed the 2-year doctoral program and presented their research works

Búsqueda Heurística para problemas de Scheduling: Cálculo de cotas inferiores y superiores. *María Sierra*

Resolución de problemas Flowshop Scheduling mediante estrategias evolutivas y heurísticas. *Noelia González*

Now they are working on their Phd. María Sierra is funded by a grant from the regional government (FICYT BP04-021) during the period 2004-2008 and Noelia Sánchez works in a private company and is also half-time assistant lecturer in the Dept. of Computer Science.

5 Future Work

During the next year, the last one of the research project, we plan to complete all the project objectives. Our main interest will be focused on the two proposed practical problems: the time tabling of train drivers and the optimal cutting of plastic film. At the moment both of these problems are well-formulated and we are currently working on the first prototype implementation.

We hope that some of the papers already submitted for publication both at international conferences and journals will be accepted. We are also confident that the completion of all project objectives will provide us with ideas and results for new publications.

TEFATE: Finite-state technologies applied to specialized translation*

TIC2003-08681-C02

Mikel L. Forcada[†]

Dept. de Llenguatges i Sistemes Informàtics
Universitat d'Alacant

Francisco Casacuberta Nolla[‡]

Dept. de Sistemes Informàtics i Computació
Universitat Politècnica de València

Abstract

In the current European scenario, characterized by the coexistence of communities writing and speaking a great variety of languages, machine translation has become a technology of capital importance. In areas of Spain and of other countries, coofficiality of several languages implies producing several versions of public information. Machine translation between three languages of the Iberian Peninsula and from them into English will allow for a better integration of Iberian linguistic communities among them and inside Europe. This project aims at applying the know-how developed by the participating teams in previous projects dealing with general machine translation between two languages (Spanish and Catalan) and specific machine translation between several languages (Spanish and Italian into English) to develop *finite-state technologies* for translation and to apply them to build tools to efficiently harness these methods for new language pairs (Spanish-Portuguese, Catalan-English) or subjects (weather reports, tourist information, medical information, phone banking, etc.). We are applying the finite-state technologies developed to (a) the creation of *machine translation systems* for texts and *interactive translation assistance systems* in specialized subjects based upon finite-state technologies from existing machine translation dictionaries and information obtained from *bitexts* (parallel bilingual texts) harvested from the web; (b) the construction of easily-maintained, fast, web-based, finite-state *translation memories* from bitexts; (c) the production of *voice-to-voice translation systems* for specific subjects, trained on specialized bitexts; and (d) the creation of databases of aligned, morphosyntactically annotated *bitexts* for future research on translation technologies. We have built *harvesting robots* to collect bitexts from the web, statistical and geometrical *aligners* enriched with preexisting linguistic information and text and bitext *classifiers* to delimit specific translation subjects.

*Spanish title: "Tecnologías de estados finitos aplicadas a la traducción especializada"

[†]Email: mlf@ua.es

[‡]Email:fcn@iti.upv.es

Keywords: finite-state transducers, text-to-text machine translation, speech-to-speech machine translation, statistical machine translation, classifiers, parallel corpora.

1 Project objectives

This project has, as a basic objective, the development of a range of *finite-state technologies* for translation and, as applied objectives,

1. creating *machine translation systems* and *interactive translation assistants* for texts in specialized subjects, based on the translation dictionaries already built by the groups and on information obtained from *bitexts* (bilingual parallel texts) harvested from the Internet;
2. building easily-maintained fast-response *translation memories* which may be used through the internet, also based on finite states, and are fed with bitexts harvested from the Internet;
3. producing *speech-to-speech* machine translation systems for specialized applications, trained on specific bitexts; and
4. building databases of aligned, morphosyntactically annotated bitexts as resources for future research.

The project will tackle the new language pairs Spanish–Portuguese, Spanish–Basque and Catalan–English, and, among other specialized subjects, tourist information, meteorological reports, patient–doctor dialogs, and phone banking.

2 Degree of performance

2.1 Activity and achievements

This section describes the activity for each module (M) and task (T) of the project. The name of the university (UA, UPV) leading each task appears in parenthesis, together with scheduling details.

M1 Task definition and coordination

T1 Project coordination (UA: 12/2003–11/2006): three coordination meetings have been held: one informal kick-off meeting and two workshops (November 2004 and May 2005); a webpage has been set up (<http://transducens.dlsi.ua.es/TEFATE/> to coordinate the project). We are trying to improve the coordination between groups and the content of the project webpage.

T2 Definition of tasks for text-to-text and speech-to-speech translation (UA, UPV: 12/2003–2/2004).

- For the text-to-text tasks, we have chosen the following language pairs: Spanish–Catalan (corpora: *Diari Oficial de la Generalitat Valenciana* y *El Periódico de Catalunya*) and Spanish–English (corpora: Xerox printer manuals, session minutes of the European Union).

- For the speech-to-speech tasks we chose the following language pairs: Catalan–English and Portuguese–English; in both cases, the task chosen was the tourist task (sentences said by a tourist at a hotel reception desk). In another project related to this one, a similar corpus was defined for a Basque–Spanish translation task.

M2 Acquisition of multilingual texts from the Internet.

T1 Acquisition of multilingual texts from the Internet (UA, UPV: 12/2003–11/2004).

- Using ad-hoc robots, the Valencia group has harvested the following corpora: *Diari Oficial de la Generalitat Valenciana* (DOGV, 19.5 million words in Spanish, 19 million in Catalan, vocabulary sizes 189,000 and 197,000 respectively) and *El Periódico de Catalunya* (14.4 million words in Spanish and 15 million words in Catalan, vocabulary size: 126000 words each).
- Development of tools for the automatic harvesting of corpora from the Internet. A general-purpose robot has been developed and is currently being tested, which harvests bitexts in five languages (Spanish, Catalan, Portuguese, Galician and English) starting from a list of seed URIs; the robot is able to identify the languages of pages and relating them to their translations.
- Elaboration of a new multi-reference bitext for the evaluation of the dictionaries and morphosyntactical taggers used.

T2 Bitext validation (UA: 3/2004–2/2005). The general-purpose harvesting robot being tested evaluates a set of different features of harvested bitext candidates, and delivers only those bitexts above some predefined thresholds.

T3 Data filtering (UA, UPV: 6/2004–5/2005).

- Filtering of harvested data (DOGV, *El Periódico*).
- Automatic tagging and manual testing of corpora for use with a semantic disambiguation module.
- Development of an encoding scheme for the semantical annotation of translation units.
- The threshold parameters used in the general-purpose harvesting robot will be defined through human validation of samples of harvested bitexts using a visual tool.

T4 Integration of linguistic data and structural information in statistical aligners (UA: 9/2004–8/2005). The project has developed a class of aligners based on linguistic-independent heuristics (LIHLA, [3, 4, 5]) with interesting results; however, advancement in aligning methods which use finite-state transducers has been insufficient.

M3 Document classification.

T1 Training of plain text document classifiers (UPV: 6/2004–5/2005).

- Development of finite-state-based classifiers, as well as classifiers based in an error-correcting adaptation of the Viterbi method.
- Development of classifiers based on statistical models, such as multinomial mixtures, as well as their specific estimation methods (expectation maximization).

- Application to text classification (in the Tourist task, the classification error has been improved by 2%).

T2 Structural classification of hypertexts (UA: 12/2004–11/2005). This task has not been initiated (waiting for validated data obtained using the general-purpose harvesting strategies). A method will be implemented, based on probabilistic tree grammars [1] to classify XHTML bitexts.

M4 Building of finite-state and statistical translators.

T1 Improvement of dictionary-based finite-state translators (UPV: 12/2003–5/2006)

- Application of statistical disambiguation using naïve Bayes multinomial methods.
- Improvement of the translation speed of the TAVAL/SISHITRA machine translation system (up to 500 words per second) through a complete rewriting.
- Debugging, extension and improvement of the TAVAL/SISHITRA dictionary through the introduction of complex, inflected translation units, frequent errors, specific terminology and prepositional regime.
- Complete rewriting of the Alacant shallow-transfer machine translation engine (used in the interNOSTRUM and Traductor Universia systems) for release under a GPL license (Corbí-Bellot et al. 2005, <http://sourceforge.net/projects/apertium/>), in collaboration with another publicly-funded project.

T2 Inference of specialized finite-state transducers (UPV: 5/2004–11/2005).

- Incorporation of statistical models into the GIATI technique [2].
- Improvement of smoothing techniques.
- Application of transducer composition.
- Development of multiple translation techniques.

T3 Training of general statistical translators (UPV: 12/2004–5/2006).

- Implementation of statistical Spanish–Catalan translators.
- Using TAVAL/SISHITRA dictionaries to improve the statistical translators.
- Development of automatic techniques for the selection of word parts.
- Developing statistical-model combinations using translation error minimization as a criterion.

T4 Translator smoothing (UPV: 12/2005–11/2006): not started yet.

M5 Subsentential finite-state translation memories.

T1 Incremental building finite-state transducers from translation memory files (UA: 3/2004–2/2005).

- Development of a very fast compiler that takes a complete translation memory and generates a finite-state transducer which is read by a module capable of treating tens of thousands of words per second (Ortiz-Rojas et al. 2005). The compiler is so fast that incrementality may not be considered an advantage, although it has to be tested on translation memories harvested from the Internet in Module 2.

- Design of a new algorithm (Recalign) for the recursive bilingual segmentation of bitexts using a greedy strategy and IBM model 1 (UPV, even if task assigned to UA).

T2 Efficient implementation of finite-state translation memories (UA: 6/2004–5/2005, this task has been merged with M5T1, q.v.).

T3 Inference of left-to-right, longest-match translation memories from partially aligned texts (UA: 6/2005–5/2006, just started).

M6 Systems for assisted translation

T1 Finite-state based assisted translation (UPV: 6/2005–11/2006, just started)

T2 Statistical machine translation based assisted translation (UPV: 6/2005–11/2006, just started)

M7 Creation of syntactically-analyzed bitext corpora.

T1 Building of probabilistic analyzers with partial relaxation of independence constraints (UA: 12/2004–11/2005). Task in progress. An implementation has been recently described [8].

T2 Design of systems to assist in the syntactical annotation of bitexts (UA: 12/2005–11/2006, not started yet).

M8 Building of translation servers

T1 Building of speech-to-speech translation prototypes for specific tasks (UPV: 12/2004–5/2006).

- Building the Catalan–English, Portuguese–English and Spanish–Basque translators for the Tourist task.

T2 Implementation of text-to-text translators as a server (UA: 6/2006–11/2006, not started yet).

T3 Implementation of the user interface (UA: 6/2006–11/2006, not started yet).

2.2 Difficulties and proposed solutions

2.2.1 Subproject TIC2003-08681-C02-01: TEFBARNet, UA

There have been some problems related to the internal restructuring of the research group, which has delayed some of the tasks. Adjustment to the programmed schedule will require a redistribution of tasks as well as a special effort of the active researchers.

2.2.2 Subproject TIC2003-08681-C02-02: ITEFTE, UPV

The main problems encountered during the development of the project have been mainly related to the acquisition, filtering, alignment and categorization of corpora, which has ended up having a higher temporal cost than what was initially planned. The multiple morphosyntactical annotation of translation units inside the dictionaries has also proven to be a slow process.

3 Result indicators

3.1 Trainees

3.1.1 Subproject TIC2003-08681-C02-01: TEFBARNet, UA

The project has a single trainee, Felipe Sánchez Martínez, who has a four-year scholarship to work on a thesis related to the project [6, 7].

3.1.2 Subproject TIC2003-08681-C02-02: ITEFTE, UPV

- Jesús Andrés Ferrer, computer engineer, collaborating scholarship holder (assigned to the project since January 2005). Work area: development of new translation models.
- Laura María Eliodoro Furió, philologist, collaborating scholarship holder (until May 2005). Work area: development, annotation and correction of dictionaries.
- Adrián Giménez Pastor, computer engineer, collaborating scholarship holder. Work area: acquisition of corpora.
- Maria Teresa González Cavero, computer engineer, scholarship holder assigned to the project. Work area: development of translation assistance tools.
- Antonio Luis Lagarda Arroyo, computer engineer, scholarship holder assigned to the project. Work area: development of translation assistance tools.

3.2 Employees

3.2.1 Subproject TIC2003-08681-C02-01: TEFBARNet, UA

The following technical staff has been hired as *técnico superior*:

- Sergio Ortiz Rojas (February 2004–February 2005), computer engineer. Work area: transducer compilers, translation memories.
- Enrique Sánchez Villamil (February 2004–), computer engineer. Work area: bitext harvesting robots.
- Susana Santos Antón (February 2005–), computer engineer. Work area: validation tools for harvested bitexts.

3.2.2 Subproject TIC2003-08681-C02-02: ITEFTE, UPV

The following technical staff has been hired as *técnico superior*:

- Laura María Eliodoro Furió, philologist (from May 2005). Work area: development, annotation, and correction of dictionaries.
- José García Hernández, computer engineer. Work area: corpora acquisition.
- Luis Rodríguez Ruiz, computer engineer. Work area: development of tools for assisted translation.

3.3 Technology transfer

3.3.1 Subproject TIC2003-08681-C02-01: TEFBARNet, UA

Many of the basic techniques developed during this project, especially those related to finite-state transducers, have been transferred to Apertium, an open-source shallow-transfer machine translation engine scheduled to be released at the end of July 2005 and developed as part of another project. Apertium will be used by Imaxin, a software company, to market machine translation services for the Spanish-Galician pair.

3.3.2 Subproject TIC2003-08681-C02-02: ITEFTE, UPV

Although in principle the collaboration of external companies in the subproject was not programmed, some companies —with which contacts had already been maintained during past projects or collaborations— have shown their interest in the results of some of the tasks developed. It is the case of ATOS Origin, who specialize in the development of tools for assisted translation; Celer Soluciones, a technical translation company; Xerox Research Center Europe (XRCE), in view of their needs to incorporate tools to translate technical manuals; and ASP (Adur Software Productions), who are interested in improving their translation-memory-based tools.

It is also the case that some of the results of this project have been useful in other technological development projects with enterprises, such as Rumbo Sistemas and ODEC.

3.4 Participation in international projects

3.4.1 Subproject TIC2003-08681-C02-02: ITEFTE, UPV

The European project TT2, in the IST program, in which the UPV group also participates, has benefited from the results obtained in the project, which have been applied in the last period of the TT2 project. The success of the applied techniques and the results obtained have prompted the groups to apply for two more projects in the fifth call of the EU FP6. Both applications have a strong relation to the subject of the present project.

3.5 Publications

3.5.1 Journal papers

- I. García-Varea and F. Casacuberta. Maximum entropy modeling: A suitable framework to learn context-dependent lexicon models for statistical machine translation. *Machine Learning*, 59:1–24, 2005.
- Verdú-Mas, J.L., Carrasco, R.C., Calera-Rubio, J. (2005). Parsing with probabilistic strictly locally testable tree languages. *IEEE Trans. on Pattern analysis and Machine Intelligence*, 27(7):1040–1050.
- E. Vidal, F. Thollard, F. Casacuberta, C. de la Higuera, and R. Carrasco. (2005) Probabilistic finite-state machines - part I. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(7):1013–1025.

- E. Vidal, F. Thollard, F. Casacuberta C. de la Higuera, and R. Carrasco. (2005) Probabilistic finite-state machines - part II. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(7):1025–1039.
- F. Casacuberta, E. Vidal, and D. Picó. (2005) Inference of finite-state transducers from regular languages. *Pattern Recognition*, 38:1431–1443.

3.5.2 Papers at conference proceedings

- J. Andrés, J. Navarro, A. Juan, and F. Casacuberta (2005). Word translation disambiguation using multinomial classifiers. In *Iberian Conference on Pattern Recognition and Image Analysis*, volume 3523 of *Lecture Notes in Computer Science*, pages 622-629. Springer-Verlag, Estoril (Portugal), June 2005.
- Caseli, H.M.; Nunes, M.G.V.; Forcada, M.L. (2005a, aceptado) LIHLA: A lexical aligner based on language-independent heuristics. En *V Encontro Nacional de Inteligência Artificial (ENIA 2005)*. São Leopoldo-RS, Brasil, julio 2005.
- Caseli, H.M.; Nunes, M.G.V.; Forcada, M.L. (2005b, aceptado). Evaluating the LIHLA lexical aligner on Spanish, Brazilian Portuguese and Basque parallel texts. *XXI Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN)*. Granada, setiembre 2005.
- Caseli, H.M.; Nunes, M.G.V.; Forcada, M.L. (2005c, aceptado). LIHLA: Shared task system description. *Workshop on Building and Using Parallel Corpora: Data-driven Machine Translation and Beyond (ACL05)*. Michigan, EE.UU., junio 2005.
- J. Civera, J.M. Vilar, E. Cubel, A.L. Lagarda, S. Barrachina, F. Casacuberta, E. Vidal, D. Picó and J. González (2004). A syntactic pattern recognition approach to computer assisted translation. In A. Fred, T. Caelli, A. Campilho, R. P.W. Duin and D. de Ridder, editors, *Advances in Statistical, Structural and Syntactical Pattern Recognition*, *Lecture Notes in Computer Science*. Vol 3138, pp. 207-215 Springer-Verlag.
- Corbí-Bellot, A.M. Forcada, M.L., Ortiz-Rojas, S. Pérez-Ortiz, J.A., Ramírez-Sánchez, G., Sánchez-Martínez, F., Alegria, I., Mayor, A., Sarasola, K. (2005). "An open-source shallow-transfer machine translation engine for the romance languages of Spain", *Proceedings of the 10th Annual Conference of the European Association for Machine Translation*, p. 79-86, 30-31 mayo 2005, Budapest, Hungría.
- I. García-Varea, A. Sanchis and F. Casacuberta (2004). A decoding algorithm for speech input statistical translation. In *Text, Speech and Dialogue: Proceedings of the 7th International Conference (TSD 2004)*, volume 3206 of *Lecture Notes in Computer Science*, pages 305-314. Springer-Verlag, Brno, Czech Republic, September 2004.
- I. García-Varea, D. Ortiz, F. Nevado, P.A. Gómez, and F. Casacuberta. (2005) Automatic segmentation of bilingual corpora: A comparison of different techniques. In *Iberian Conference on Pattern Recognition and Image Analysis*, volume 3523 of *Lecture Notes in Computer Science*, pages 614-621. Springer-Verlag, Estoril (Portugal), June 2005.
- J. González, D. Ortiz, J. Tomás and F. Casacuberta (2004). A comparison of different statistical machine translation approaches for spanish-to-basque translation. In *Actas de las Terceras Jornadas de Tecnología del Habla*, Valencia, Spain, November 2004.

- C. D. Martínez-Hinarejos, L. Rodríguez, J. A. Sánchez, A. Sanchis and E. Vidal (2004). Comparison of several speaker adaptation alternatives for a spanish speech task. In *Actas de las III Jornadas de Tecnologías del Habla*, Valencia, Spain, November 2004.
- J. R. Navarro, J. González, D. Picó, F. Casacuberta, J. M. de Val, F. Fabregat, F. Pla and J. Tomás (2004). SisHiTra: A hybrid machine translation system from Spanish to Catalan. In J. L. Vicedo, P. Martínez-Barco, R. Muñoz and M. Saiz, editors, *Advances in Natural Language Processing, Lecture Notes in Artificial Intelligence*, pages 349-359. Springer-Verlag, Alicante, Spain.
- F. Nevado and F. Casacuberta (2004). Bilingual corpora segmentation using bilingual recursive alignments. In *Actas de las III Jornadas en Tecnologías del Habla*, 3JTH, Valencia, November 2004.
- F. Nevado, F. Casacuberta and J. Landa (2004). Translation memories enrichment by statistical bilingual segmentation. In *Proceedings of the IV International Conference on Language Resources and Evaluation - LREC2004*, volume 1, pages 335-338, Lisbon, May 2004.
- Ortiz-Rojas, S., Forcada, M.L., Ramírez-Sánchez, G. (2005). Construcción y minimización eficiente de transductores de letras a partir de diccionarios con paradigmas. XXI Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN). Granada, setiembre 2005.
- D. Picó, J. Tomás and F. Casacuberta (2004). GIATI: A general methodology for finite-state translation using alignments. In *Statistical, Structural and Syntactical Pattern Recognition. Proceedings of the Joint IAPR International Workshops SSPR2004 and SPR2004*, volume 3138 of *Lecture Notes in Computer Science*, pages 216-223. Springer-Verlag, Lisboa, Portugal, August 2004.
- Sánchez-Martínez, F., Pérez-Ortiz, J.A., Forcada, M.L. (2004a). Exploring the use of target-language information to train the part-of-speech tagger of machine translation systems. *Lecture Notes in Computer Science Lecture Notes in Artificial Intelligence 3230 (Advances in Natural Language Processing, Proceedings of EsTAL - España for Natural Language Processing)*, p. 137-148, 20-22/10/2004, Alicante, España.
- Sánchez-Martínez, F., Pérez-Ortiz, J.A., Forcada, M.L. (2004b). "Cooperative unsupervised training of the part-of-speech taggers in a bidirectional machine translation system". *Proceedings of TMI 2004, The Tenth Conference on Theoretical and Methodological Issues in Machine Translation*, p. 135-144, 4-6/10/2004, 2004, Baltimore, MD, EE.UU.
- Sánchez-Villamil, E.; Forcada, Mikel L.; Carrasco, Rafael C. "Unsupervised Training of a Finite-State Sliding-Window Part-of-Speech Tagger" (2004). *Lecture Notes in Computer Science - Lecture Notes in Artificial Intelligence 3230 (Advances in Natural Language Processing, Proceedings of EsTAL - España for Natural Language Processing)*, p. 454-463, 20-22/10/2004, Alicante, España.
- J. Tomás and F. Casacuberta (2004). Statistical machine translation decoding using target word reordering. In *Statistical, Structural and Syntactical Pattern Recognition. Proceedings of the Joint IAPR International Workshops SSPR2004 and SPR2004*, vol-

ume 3138 of Lecture Notes in Computer Science, pages 734-743. Springer-Verlag, Lisboa, Portugal, August 2004.

- J. Tomás, J. Lloret and F. Casacuberta (2004). A Spanish-Catalan translator using statistical methods. In Proceedings of the 16th European Conference on Artificial Intelligence (ECAI04), pages 1099-1100, Valencia, Spain, August 2004. IOS Press.
- J. Tomás, J. Lloret and F. Casacuberta (2004). Spanish-catalan translator using statistical methods. In Actas de las Terceras Jornadas de Tecnología del Habla, Valencia, Spain, November 2004.
- J. Tomás, J. Lloret, and F. Casacuberta (2005). Phrase-based alignment models for statistical machine translation. In Iberian Conference on Pattern Recognition and Image Analysis, volume 3523 of Lecture Notes in Computer Science, pages 605-613. Springer-Verlag, Estoril (Portugal), June 2005.
- E. Vidal and F. Casacuberta (2004). Learning finite-state models for machine translation. In Grammatical Inference: Algorithms and Applications. Proceedings of the 7th International Colloquium ICGI 2004, volume 3264 of Lecture Notes in Artificial Intelligence, pages 16-27. Springer, Athens, Greece, October 2004. Invited conference.

3.6 Activities planned until the end of the project

- Text classification:
 - Classifying bilingual texts using mixture models: combination of IBM models and n -grams. (UPV).
 - Developing hypertext classifiers based on their structure, using probabilistic grammars (UA).
- Development of aligners and translators:
 - Transforming statistical models into finite state models (UPV)
 - Combining GIATI [2] with other statistical models (UPV)
 - Using linguistic information in the automatic construction of statistical and finite-state translators (UPV)
 - Deploying language-independent heuristic aligners (LIHLA, [3, 4, 5]) to obtain linguistic data for machine translation systems (UA).
 - Implementing geometrical aligners based on edit costs which incorporate linguistic information through finite-state transducers (UA).
- Developing search algorithms for statistical translators:
 - For mixtures of IBM models (UPV).
 - Multiple statistical models (UPV).
- Development and debugging of dictionaries and linguistic data:
 - Improving of the TAVAL dictionary by introducing semantic information (UPV).

- Automatically inducing the topology (tag groupings) and parameters of part-of-speech taggers using information from the target language (UA)
- Exploring the use of simple, target-language driven statistical choice models to treat polysemic words in rule-based machine translation systems (UA).
- Implementation of machine translation systems:
 - Improving the code of SisHiTra-TAVAL. Augmenting SisHiTra-TAVAL with the capability to translate web pages (UPV [UA]) and LaTeX files.
 - Implementing the Catalan-Spanish version of SisHiTra-TAVAL (UPV).
 - Implementing systems for multiple translation (UPV).
 - Extension of the Apertium (interNOSTRUM-Traductor Universia) architecture to deal with deeper syntactic transfer and multiple lexical equivalents for polysemic words (UA).
- Assisted translation/transcription systems:
 - Building interactive assisted translation systems: interface, translation engines and incorporation of the SisHiTra-TAVAL system(UPV)
 - Exploring the development of systems for assisted transcription (UPV).
- Harvesting of bitexts from the Internet
 - Use of a bitext evaluation tool by humans to tune the parameters (threshold) of the automatic bitext harvesting robot (UA).
 - Exploring conservative techniques to obtain “safe” subsentential units from fully-automated Internet-harvested bitexts (UA).
 - Creation of a finite-state based prototype that will manage the harvested translation units to act as a translation assistance system (UA).
- Creating syntactically-analysed bitext corpora:
 - Developing tools for the simultaneous syntactic annotation of bitexts using online induction of probabilistic grammars (UA).

3.7 Collaboration with other groups

3.7.1 Subproject TIC2003-08681-C02-01: TEFBARNet, UA

As part of the project activity, a collaboration has started with the Núcleo Interinstitucional de Lingüística Computacional (NILC) of the Universidade de São Paulo (USP) en São Carlos (SP), led by professor Maria das Graças Volpe Nunes. As part of this collaboration,

- Helena de Medeiros Caseli, a student from that group, has visited the Universitat d’Alacant for one year, funded by a Brazilian agency. Helena has been developing a class of bitext aligners, and has produced three publications [3, 4, 5].
- The principal investigator has paid a short visit to the Brazilian group related to the work of this student.

Also, a collaboration has been initiated with Prof. Jan Daciuk, of the Department of Knowledge Engineering of the University of Gdańsk, in Poland. The subject of this collaboration is the incremental construction of minimal tree automata (Prof. Daciuk visited the UA in October 2004).

3.7.2 Subproject TIC2003-08681-C02-02: ITEFTE, UPV

The project has produced an important collaboration with the Portuguese group *Laboratório de Sistemas de Língua Falada* of the Instituto Superior Técnico in Lisbon. This collaboration has been organized around an Integrated Action between both groups, which has allowed mutual visits and an intense collaboration in research tasks.

References

- [1] R.C. Carrasco, J.R. Rico-Juan, "A similarity between probabilistic tree languages: application to XML document families", *Pattern Recognition* 36:9 (2002) 2197-2199.
- [2] F. Casacuberta "Inference of finite-state transducers by using regular grammars and morphisms", *Lecture Notes in Computer Science*, vol. 1891, p. 1-14.
- [3] Caseli, H.M.; Nunes, M.G.V.; Forcada, M.L. LIHLA: A lexical aligner based on language-independent heuristics. En V Encontro Nacional de Inteligência Artificial (ENIA 2005). São Leopoldo-RS, Brasil, julio 2005.
- [4] Caseli, H.M.; Nunes, M.G.V.; Forcada, M.L. Evaluating the LIHLA lexical aligner on Spanish, Brazilian Portuguese and Basque parallel texts. XXI Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN). Granada, setiembre 2005.
- [5] Caseli, H.M.; Nunes, M.G.V.; Forcada, M.L. . LIHLA: Shared task system description. Workshop on Building and Using Parallel Corpora: Data-driven Machine Translation and Beyond (ACL05). Michigan, EE.UU., junio 2005.
- [6] Sánchez-Martínez, F., Pérez-Ortiz, J.A., Forcada, M.L. Exploring the use of target-language information to train the part-of-speech tagger of machine translation systems. *Lecture Notes in Computer Science Lecture Notes in Artificial Intelligence* 3230 (Advances in Natural Language Processing, Proceedings of EsTAL - España for Natural Language Processing), p. 137-148, 20-22/10/2004, Alicante, España (2004).
- [7] Sánchez-Martínez, F., Pérez-Ortiz, J.A., Forcada, M.L. "Cooperative unsupervised training of the part-of-speech taggers in a bidirectional machine translation system". *Proceedings of TMI 2004, The Tenth Conference on Theoretical and Methodological Issues in Machine Translation*, p. 135-144, 4-6/10/2004, 2004, Baltimore, MD, EE.UU.
- [8] J.L. Verdú-Mas, R.C. Carrasco, J. Calera-Rubio. Parsing with probabilistic strictly locally testable tree languages. *IEEE Trans. on Pattern analysis and Machine Intelligence*, 27(7):1040-1050 (2005).

DELIA: Deduction and programming in non-classical logics for knowledge technology

TIC2003-9001-C02

Manuel Ojeda Aciego ^{*}
Univ. Málaga

David Pearce [†]
Univ. Rey Juan Carlos

Summary

This project deals with the study of mathematical foundations, as well as deduction and programming, for several types of non-classical logics, especially multiple-valued logics and intermediate logics, with the aim of providing means for intelligent management of information. The classical problem of deduction is, given explicit knowledge expressed in a given formalism together with a set of inference rules, to deduce the implicit knowledge which could be relevant for applications. Specifically, we are interested in those knowledge bases containing uncertain or incomplete information, to develop mechanisms of analysis in which vague or imprecise queries are permitted. A first formal framework chosen is fuzzy logic and different paradigms of fuzzy logic programming. In this project we propose the development of a theoretical framework which generalises different existent approaches to logic programming in vague or imprecise contexts by considering different lattice-like generalised structures as the set of truth-values: bilattices, trilattices, multi-lattices, biresiduated lattices, etc. For this general goal, two additional specific problems will be addressed: firstly, unification in these general contexts, in which categorical techniques are expected to be useful and, secondly, automated deduction in these systems, in which we are planning to use the TAS methodology, which is mechanised deduction framework developed by the Málaga partner which integrates (some of) the benefits of tableaux and resolution; for instance, non-clausal formulas can be analysed directly whereas unit-resolution and the pure literal rule are built-in. From the programming point of view, extensions are sought of the paradigm of Answer Set Programming (ASP) which in turn is already a generalisation of classical logic programming. These extensions of ASP can be implemented either in existing implementations of ASP, or by using systems based on quantified Boolean formulas (e.g. QUIP), or by using TAS. The importance of basic research in these logical systems for Knowledge Technology is driven by the numerous potential applications they offer for modelling and solving problems in areas such as information management, cognitive robotics, Semantic Web, etc.

^{*}Email: aciego@ctima.uma.es

[†]Email: davidandrew.pearce@urjc.es

1 Objectives of the project

The aim of the first sub-project is to develop deduction methods for lattice-valued logics with applications to soft-computing. Specifically, a general mathematical framework for logic programming extended to a fuzzy context is envisaged which embeds the particular features of the residuated and multi-adjoint approaches to logic programming. The main research objectives in this sub-project are the following:

- Investigate many-valued and fuzzy logics over lattices and other generalised structures: Description of logics taking values in different ordered structures, studying the particularities of their semantics and proof theory.

Development of a theory of multi-lattices, analysing its algebraic features, including the generalisation of the notions of filter and ideal. Investigate the possibility of extending existing results to these structures.

- Extension of fuzzy logic programming paradigms: Generalisation of existing frameworks by using lattice-valued logics.

Development of algorithms for (categorical) unification useful for the extended paradigms introduced so far.

- Generalisation of TAS. Development of the mathematical structures needed in order to study the structure of the set of literals of particular logics of interest. Viability of the extension of the concept of Δ -tree or alternative data structures for these logics.

The aim of the second sub-project is the study of new or extended systems of non-monotonic reasoning by using non-classical logics, investigate their meta-theoretical properties and their potential applications to practical problems of reasoning in knowledge technologies. The systems to be studied can be viewed as belonging to the new paradigm of *answer set programming* (ASP) that is currently being used in AI application areas such as planning, diagnosis, information integration and the Semantic Web. More specifically, areas of investigation include:

- Extending the paradigm of answer set programming to deal with aspects such as growth of the knowledge base (induction, abduction, explanations), change in the knowledge base (belief revision, update theory) and approximate reasoning (probabilistic, fuzzy, bayesian).
- Investigate extensions of the language of ASP/equilibrium logic to include modal, causal/action operators and first-order formalisms.
- Study deduction methods and implementation techniques. Investigate potential applications of the resulting systems.

Research tasks

In this section we briefly describe the different tasks scheduled in order to fulfil the objectives stated above:

1. *Investigation of many-valued logics over generalised ordered structures:*

- (a) *Multi-lattices*: Our starting point is the generalisation of lattices in which the concept of *supremum* (resp. *infimum*) is substituted by *minimal* (resp. *maximal*) of the set of upper bounds (resp. lower bounds), which leads us to work with non-deterministic instead of deterministic operators.
- (b) *Bi-lattices, tri-lattices and generalisations*. We study the theoretical and practical implications of working with more than one lattice structure on the set of truth-values with respect to the associated operational semantics. In this direction, it would be worth considering a further generalisation by using an underlying multi-lattice.
- (c) We study the particularities of the different many-valued logics obtained when using the type of ordered structures described above as underlying set of truth-values.

2. *Extension of fuzzy logic programming paradigms.*

A framework which increases expressivity is sought. The different possibilities concerning orderings in the underlying set of truth-values should be evaluated regarding their scope and limits. Moreover, the corresponding extension to first-order requires studying the problem of unification in non-canonical contexts. More specific tasks are the following:

- (a) *Extending existing fuzzy logic programming paradigms*. Investigate the continuity of the fixed-point operators for the proposed logics. Investigate the existence of sound and complete operational semantics for the proposed systems.
- (b) *Developing categorical unification algorithms* to cope with the extension to first-order of the proposed systems. In a fuzzy context, neither the usual unification algorithm nor those based on similarity relations are able to deal with all the extra expressive power of these logics. We believe that an abstract approach to unification at a categorical level could be an interesting solution.

3. *Mechanisation of reasoning.*

- (a) *Development of the mathematical structures* needed in order to study the structure of the set of literals of particular logics of interest. Viability of the extension of the concept of Δ -tree or alternative data structures for these logics.
- (b) *Definition of a TAS prover for the proposed logics*: Study specific simplification rules, both equivalence- and satisfiability-preserving, for the proposed systems.

4. *Extensions of the current version of equilibrium logic.*

- (a) *Generalisations of the concept of strong equivalence and their applications to logic programming*. Equilibrium logic is based on the nonclassical logic of here-and-there. Since this logic characterises the strong equivalence of programs in ASP (strongly equivalent programs are interchangeable in any context), it follows that any transformation or simplification which preserves here-and-there equivalence can be applied to programs as a pre-processing step. An important area of study is to consider whether some types of transformations suggested by results on strong equivalence can provide enhancements to ASP solvers such as `d1v`.

- (b) *Extensions of the language.* In logic programming, an important class of extended reasoning systems can be obtained by the inclusion of new operators in the basic logic language, for instance, modal, causal or action operators. Existing constructions are defined in the context of stable models, therefore, it is natural to investigate extensions to nested programs and to equilibrium logic as a whole.
- (c) *Abduction and induction.* In previous work by members of the project, equilibrium logic for abductive reasoning was considered. It is planned to continue this work by looking also at the concept of *explanation*. There are close connections between abductive logic programming and inductive logic programming (ILP). Another task in this context is the development of an ILP system for answer set programming with the ability to deal with generalisations such as nested programs.
- (d) *Belief revision and updates.*

In the language of equilibrium logic and extended logic programming, the concept of strong negation is closely related to the idea of incredulity. Previous works showed some relation between the standard methods of belief revision and incredulity. The project, will study how belief revision operators can be defined for languages with strong negation and for formalisms such as equilibrium logic.

5. *New versions of equilibrium logic with different underlying logics .*

- (a) *Extend the equilibrium construction* to other multiple-valued logics (either finite- or infinite-valued) and study the behaviour of the logics obtained.
- (b) *Capturing other LP semantics* by means of variants of equilibrium logic. The equilibrium construction in the logic of here-and-there captures reasoning with stable models or answer sets. By modifying the construction (resp. the underlying logic) one may consider whether other logic programming semantics (eg. well-founded semantics) can also be characterised.

6. *Implementation and related tasks.*

- (a) Search for a polynomial translation for the complete equilibrium logic into disjunctive logic programs, and implement it as a front-end of **dlv**. Compare the resulting system with the current version of equilibrium reasoning as implemented in **QUIP**.
- (b) TAS directly implements proof procedures for multiple-valued logics and their extensions, instead of translating a problem in non-classical logic into a problem in classical logic, and then apply ASP or QBF solvers. A first TAS solver for equilibrium logic is being developed and can be used as a starting point for the implementation of a mechanised reasoning system comparable to the methods mentioned above.

Gantt chart

Tasks	1st year	2nd year	3rd year
Multi-lattices and generalizations	██████████	░░░░░░░░░░	░░░░░░░░░░
Logics valued on lattices and generalizations	░░░░░░██████████	██████████░░░░	░░░░░░░░░░
Extended logic programming frameworks	░░░░░░░░░░	██████████	██████████
Categorical generalised unification algorithms	░░░░░░██████████	██████████	██████████░░░░░░
Mathematical structures for extending TAS	░░░░░░░░░░░░░░	░░░██████████	██████████
Non-monotonicity in many-valued and fuzzy logics	██████████	██████████	░░░░░░░░░░
Abduction and induction in ASP systems	██████████	██████████░░░░░░░░	░░░░░░██████████
Belief revision	░░░░░░░░░░░░░░	░░░░░░░░░░	██████████
Extensions of the language	░░░░░░░░░░░░░░	░░██████████	██████████
Strong equivalence and relation between theories	░░░██████████	██████████	██████████░░░░░░
Implementations in QUIP and dlv	██████████	██████████	░░░░░░░░░░
TAS and non-monotonicity	░░░░░░░░░░░░░░	░░░░░░██████████	██████████

2 Level of success achieved to date

Most of the proposed tasks are being fulfilled as expected. In the rest of the section we will describe the major achievements obtained in each of the tasks.

The publications [4, 5, 18] are specifically oriented to tasks (1a) and (1c) with marginal contribution to task (1b). In particular, in [4, 5] the structures called multi-semilattice, multi-lattice, universal multi-semilattice and universal multi-lattice are introduced, algebraic characterisations are proposed for them, and their mathematical properties are studied. On the other hand, [18] focuses on a prospective study of the use of the structure of ordered multi-lattice as underlying set of truth-values for a generalized framework of logic programming is presented. Specifically, we investigate the possibility of using multilattice-valued interpretations of logic programs and the theoretical problems that this generates with regard to its fixed point semantics.

Task (2a) is covered by publications [12, 13, 14, 15]. In [14] a neural net based implementation of propositional $[0, 1]$ -valued multi-adjoint logic programming is presented. The implementation needs some preprocessing of the initial program to transform it into a *homogeneous* program; then, transformation rules carry programs into neural networks, where truth-values of rules relate to output of neurons, truth-values of facts represent input, and network functions are determined by a set of general operators; the output of the net being the values of propositional variables under its minimal model. Later, in [16, 12], a generalization of the homogenization process needed for the neural implementation is presented in order to deal with a more general family of adjoint pairs, but maintaining the advantage of the existing implementation. A modification of the neural implementation is presented in [15] in order to deal with a more general family of adjoint pairs, including conjunctors constructed as an ordinal sum of a finite family of basic conjunctors. This enhancement greatly expands the scope of the initial approach, since every t-norm (the type of conjunctor generally used in applications) can be expressed as an ordinal sum of product, Gödel and Łukasiewicz conjunctors. Finally, a different underlying lattice is used in [13], where we pursue an extension of the previous

approaches in order to accomodate calculation with truth-intervals.

The modal logic of qualitative order-of-magnitude reasoning $\mathcal{L}(MQ)$ [1] arose as an interesting by-product when studying different approaches to reasoning under imperfect information. Current joint work is being done together with Prof. Ewa Orłowska in order to obtain a Rasiowa-Sikorski proof procedure for $\mathcal{L}(MQ)$.

A more theoretical study of properties related to task (2a) has been published in [7, 9, 10], where termination conditions for various extensions of fuzzy logic programming frameworks were introduced.

In [19] a formal model for similarity-based fuzzy unification in multi-adjoint logic programs was introduced: on the multi-adjoint computational model, a similarity-based unification approach is constructed by simply adding axioms of fuzzy similarities and using classical crisp unification which provides a semantic framework for logic programming with different notions of similarity.

Reference [11] introduces advances towards the generalised categorical unification algorithm: a categorical approach involving set and term functors as monads allows for a study of monad compositions that provide variable substitutions and compositions thereof. In this paper, substitutions and unifiers appear as constructs in Kleisli categories related to particular composed powerset term monads. Specifically, we show that a frequently used similarity-based approach to fuzzy unification is compatible with the categorical approach, and can be adequately extended in this setting.

The references [37, 22, 8] contain different approaches to the mechanisation of reasoning and, therefore, are related to task (3b); the first reference introduces a tableaux-inspired prover for equilibrium logic; the second presents a new implementation of the TAS prover for classical propositional logic; the third one introduces an tabulation proof procedure for residuated logic programming, as a first attempt to obtain a such a proof procedure for full multi-adjoint logic programming.

Regarding the second sub-project, its main effort during this period was directed at the following:

- Consolidation and extension of concepts of equivalence between programs or theories in answer set programming and equilibrium logic.

The importance of this topic is evident from the fact that LPNMR (the most important conference on the topic of logic programming and non-monotonic reasoning) dedicated in its 2004 edition a complete session of four talks to this specific theme. In addition, INFOMIX (currently the most ambitious European project on the application of ASP, see <http://sv.mat.unical.it/infomix/>) initiated the study of variations of this concept for program optimisation for applications in the field of data integration.

Reference [26] proposes a characterisation of uniform equivalence in equilibrium logic and a tableaux system for checking uniform equivalence, a first implementation of which has been developed and described in [37]. A second paper, [27], is concerned to a more significant extension of the concept of equivalence: the synonymy between theories formulated in different languages. Works [21, 2] study strong equivalence relative to the semantics of paraconsistent answer sets (PAS) and propose a new representation of models. Regarding the optimisation of computational systems, an application of strong equivalence to the

transformation and simplification of programs can be seen in [23]. These transformations could be implemented as soon as we have efficient systems of automated deduction for the here-and-there logic.

- Extension of the language of equilibrium logic.

The project foresees a series of possible extensions of the language of equilibrium logic, such as the addition of modal or causal operators, in order to extend the current languages of action based on ASP. In this first period of the project we have focussed on first-order extensions since, in the short term, they seem to have greater potential impact on extensions of existing computational systems. (However a partial, modal extension of equilibrium logic has recently been studied by Wand and Zhang (LPNMR 05).) The first results obtained [28, 25] are promising in that we already have a first-order version for the logic N5 with constant domains for here-and-there models with nice meta-theoretical properties and capturing the inference relation of ASP for programs with variables.

The future research line on the integration of ASP and multi-agent systems has two main goals. The first one is the extension of agent programming languages based on ASP, taking into account the organising context provided by the multi-agent systems in which they are embedded. A preliminary work in this direction (although without mentioning ASP) is [34]. The second goal deals with the specification of protocols of interaction by means of ASP, starting from the conceptual framework established in [33].

- Other variants of equilibrium logic.

Two generalisations have been studied recently. In [20] it is shown how a paraconsistent version of the logic N5 can be used to capture paraconsistent answer sets. In a further paper under submission it is shown how a partial version of equilibrium logic (based on a previously unknown nonclassical logic) captures reasoning with well-founded semantics.

- Contribution to the development of computational systems.

This theme is covered by references [26] (tableaux system), [23] (transformations) and [2], which introduces two types of reduction of propositional theories in equilibrium logic to logic programs. The second method described in [2] seems to be adequate as a basis for the implementation of equilibrium logic as a front-end for `d1v`.

Last but not least, it is worth mentioning the automated system described in [37] for equilibrium logic, which includes both a strong equivalence and a uniform equivalence checker.

- Foundations of fuzzy logic and reasoning

The Madrid partner has also been an active contributor to the area of fuzzy logic, treated mainly in subproject 1. In [35, 36] distinct properties of fuzzy inference are studied, in particular verification of the meta rules of modus ponens/modus tollens and of contrasymmetry for the case of fuzzy implication. The papers [3, 31, 32] deal with the theory of aggregation. [3] proposes a new mechanism for aggregation that takes account of the origin of the information to be aggregated, while [31, 32] deal with the principles of non-contradiction and excluded middle.

3 Progress indicators

After fifteen months since his incorporation to the project, the PhD student associated to the first sub-project has begun to actively collaborate in some specific research tasks. As a result, a workshop contribution has been already accepted and a draft paper has been submitted for publication.

Members of the research team have been invited to give a talk or organise a special session in different international conferences:

- In July 2004, A. Valverde and D. Pearce were invited to the workshop “Negation in Constructive Logic” held in Dresden (Germany). During this workshop, they contacted other researchers whose topics are related to the goals of this project, in particular with Sergei Odintsov from the Sobolev Institute of Mathematics (Russia).
- Simultaneously, J. Medina and M. Ojeda presented an invited talk on a special session at IPMU dedicated to reasoning under imperfect information [17] surveying on recent and prospective results about multi-adjoint logic programming. Later, when attending the ECAI and JELIA conferences, M. Ojeda initiated a collaboration with Umberto Straccia, from CNR (Italy), and Carlos Damásio, from Univ. Nova de Lisboa, in order to develop the mathematical foundations of a further extension of logic programming to manage uncertainty. This collaboration resulted in the application of a proposal of Integrated Action between Spain and Italy for the period 2006–2007.
- In April 2005, A. Valverde and D. Pearce participated, by invitation, in the seminar “Nonmonotonic Reasoning, Answer Set Programming and Constraints” held in Dagstuhl (Germany), where prospective collaboration with other research groups, specially, from the European project WASP was discussed.
- Moreover, a special invited session in EUSFLAT will be organised by J. Medina and M. Ojeda on the topic of *Non-Classical Unification and Generalised Logic Models*.

Regarding publications, here we present a summary of journal publications, peer-reviewed national/international conference proceedings and workshop contributions.

Journal Publications Fundamenta Informaticae [1], Fuzzy Sets and Systems [3, 11, 19], Discrete Mathematics [4], Annals of Mathematics and Artificial Intelligence [5], Mathware & Soft Computing [12], Journal of Applied Logic [6, 14], Studia Logica [25], Revista de la Real Academia de Ciencias [35], International Journal of Intelligent Systems [36].

Intl. Conf. JELIA: Logics in Artificial Intelligence [7, 22, 28, 37], ECAI: European Conference in Artificial Intelligence [8, 27] IPMU: Information Processing and Management of Uncertainty [10, 17], LPNMR: Logic Programming and Non-Monotonic Reasoning [20, 26], ICLP: International Conference on Logic Programming [23], IWINAC: Intl. Conf. on the Interplay between Natural and Artificial Computation [13], IBERAMIA: Ibero-American Conference on Artificial Intelligence [15], EUSFLAT: European Conference on Fuzzy Logic and Technologies [31], SAC: Symposium on Applied Computing [33], PROMAS: Programming Multi-Agent Systems [34].

Intl. Workshops WASP: Workshop on Answer Set Programming [2], ALC: Asian Logic Conference [21], Philosophical Foundations of Information Systems Engineering [24], DGF Workshop on Negation in Constructive Logic [29], Dagstuhl Seminar on Nonmonotonic Reasoning, Answer Set programming and Constraints [30], Aggregation Operators [32]

Natl. Conf. ESTYLF: Spanish Conference on Fuzzy Logic and Technologies [9, 16], LFSC: Lógica Fuzzy & Soft Computing [18].

Last but not least, there is a substantial level of collaboration with other groups, as shown:

1. with Pedro Cabalar of the University of Corunna on the foundations of nonmonotonic reasoning. Prof. Cabalar has joined the project as an associate to the Madrid partner.
2. with Carlos V. Damásio from Univ. Nova de Lisboa, on the topic of termination results and tabulation procedures for different extensions of fuzzy logic programming.
3. with Thomas Eiter, Hans Tompits and Stefan Woltran of the Technical University Vienna on implementing equilibrium logic and on variations of strong equivalence.
4. with Patrik Eklund and M^a Ángeles Galán from Umeå University, on the generalised categorical unification algorithm. Dr. Galán is joining the Málaga partner as a Juan de la Cierva Research Fellow (2005).
5. with Sergei Odintsov of the Sobolev Mathematical Institute Novosibirsk on logical systems with nonclassical negation.
6. with Ewa Orłowska from the National Telecommunications Institute in Warsaw, on Rasiowa-Sikorski proof methods for the logic of qualitative order-of-magnitude reasoning.
7. with Axel Polleres of the University of Innsbruck on answer set programming and the semantic web. Dr. Polleres is joining the Madrid partner as a Juan de la Cierva Research Fellow (2005).
8. with Umberto Straccia from *Consiglio Nazionale delle Ricerche* (Italy) about mathematical foundations of a further extension of logic programming to manage uncertainty.
9. with Peter Vojtáš from Charles University of Prague, on the development of the theory of multi-adjoint logic programming.

References

- [1] A. Burrieza and M. Ojeda-Aciego. A multimodal logic approach to order of magnitude qualitative reasoning with comparability and negligibility relations. *Fundamenta Informaticae*, 2005. To appear.
- [2] P. Cabalar, D. Pearce and A. Valverde. Reducing propositional theories in equilibrium logic to logic programs, *Proceedings Workshop Answer Set Programming, ASP'05*.
- [3] T. Calvo and A. Pradera. Double Aggregation Operators. *Fuzzy Sets and Systems* 142(1):15–33, 2004.
- [4] P. Cordero, J. Martínez, G. Gutiérrez, and I.P. de Guzmán. Generalizations of lattices looking at computation. *Discrete Mathematics* 295:107–141, 2005.
- [5] P. Cordero, G. Gutiérrez, J. Martínez, and I. P. de Guzmán. A new algebraic tool for automatic theorem provers. *Annals of Mathematics and Artificial Intelligence*, 42(4):369–398, 2004.
- [6] C.V. Damásio, J. Medina, and M. Ojeda-Aciego. Termination of logic programs with imperfect information: applications and query procedure. *Journal of Applied Logic*, 2005. To appear
- [7] C.V. Damásio, J. Medina, and M. Ojeda-Aciego. Sorted multi-adjoint logic programs: termination results and applications. In Logics in Artificial Intelligence, JELIA'04, pages 260-273. *Lect. Notes in Artificial Intelligence* 3229, 2004.
- [8] C.V. Damásio, J. Medina, and M. Ojeda-Aciego. A tabulation proof procedure for residuated logic programming. In European Conference on Artificial Intelligence, volume 110 of *Frontiers in Artificial Intelligence and Applications*, pages 808-812, 2004.
- [9] C.V. Damásio, J. Medina, and M. Ojeda-Aciego. A termination theorem for sorted multi-adjoint logic programming. In *Proc. 12th Spanish Congress on Fuzzy Logic and Technology, ESTYLF'04*, 2004.
- [10] C.V. Damásio, J. Medina, and M. Ojeda-Aciego. Termination results for sorted multi-adjoint logic programs. In *Proc. Information Processing and Management of Uncertainty, IPMU'04*, pages 1879-1886, 2004.
- [11] P. Eklund, M. A. Galán, J. Medina, M. Ojeda-Aciego, A. Valverde: Similarities between powersets of terms. *Fuzzy Sets and Systems* 144(1): 213-225, 2004
- [12] J. Medina, E. Mérida-Casermeyro, and M. Ojeda-Aciego. A neural implementation of multi-adjoint logic programs via sf-homogeneous programs. *Mathware & Soft Computing*, 2005. To appear.
- [13] J. Medina, E. Mérida-Casermeyro, and M. Ojeda-Aciego. Interval-valued neural multi-adjoint logic programs. In *Proc. Intl. Conf. on the Interplay between Natural and Artificial Computation*, pages 521-530. *Lect. Notes in Computer Science* 3561, 2005.

- [14] J. Medina, E. Mérida-Casermeiro, and M. Ojeda-Aciego. A neural implementation of multi-adjoint logic programming. *Journal of Applied Logic*, 2/3:301-324, 2004.
- [15] J. Medina, E. Mérida-Casermeiro, and M. Ojeda-Aciego. Decomposing ordinal sums in neural multi-adjoint logic programs. In Ibero-American Conference on Artificial Intelligence, pages 717-726. *Lect. Notes in Artificial Intelligence* 3315, 2004.
- [16] J. Medina and M. Ojeda-Aciego. Extended homogenization for multi-adjoint logic programs. In *Proc. of 12th Spanish Congress on Fuzzy Logic and Technology, ESTYLF'04*, 2004.
- [17] J. Medina and M. Ojeda-Aciego. Multi-adjoint logic programming. In *Proc. of Information Processing and Management of Uncertainty, IPMU'04*, pages 823-828, 2004. Invited talk of special session on Logic Programming with Imperfect Information.
- [18] J. Medina, M. Ojeda-Aciego, and J. Ruiz Calviño. Fuzzy logic programming via multi-lattices: first results and prospects. *Proc. of Lógica Difusa & Soft Computing*, 2005. To appear.
- [19] J. Medina, M. Ojeda-Aciego, and P. Vojtáš. Similarity-based unification: a multi-adjoint approach. *Fuzzy Sets and Systems*, 146(1):43-62, 2004.
- [20] S. Odintsov and D. Pearce. Routley Semantics for Answer Sets. In *Proc. of Logic Programming and Non-Monotonic Reasoning, LPNMR'05*, 2005, forthcoming.
- [21] S. Odintsov and D. Pearce. The Logic of Paraconsistent Answer Sets (extended abstract), *Proceedings of the Asian Logic Conference 2005*, Novosibirsk, 2005.
- [22] M. Ojeda-Aciego and A. Valverde. `tascpl`: TAS solver for Classical Propositional Logic. In Logics in Artificial Intelligence, JELIA'04, pages 731-735. *Lect. Notes in Artificial Intelligence* 3229, 2004.
- [23] D. Pearce. Simplifying logic programs under answer set semantics. Proceedings of ICLP04, Springer, *Lect. Notes in Computer Science* 3132:210-224, 2004.
- [24] D. Pearce. Some Methodological Issues for Intelligent Information Systems, *Proceedings PHISE 05*, to appear
- [25] D. Pearce, A. Valverde: A First Order Nonmonotonic Extension of Constructive Logic. *Studia Logica* 80:323-348 (2005)
- [26] D. Pearce, A. Valverde: Uniform Equivalence for Equilibrium Logic and Logic Programs. In Logic Programming and Nonmonotonic Reasoning, LPNMR'04, *Lecture Notes in Computer Science* 2923:194-206, 2004
- [27] D. Pearce, A. Valverde: Synonymus Theories in Answer Set Programming and Equilibrium Logic. In European Conference on Artificial Intelligence, volume 110 of *Frontiers in Artificial Intelligence and Applications*, pages 388-392, 2004

- [28] D. Pearce, A. Valverde: Towards a First Order Equilibrium Logic for Nonmonotonic Reasoning. In Logics in Artificial Intelligence, JELIA'04, pages 147-160. *Lect. Notes in Artificial Intelligence* 3229, 2004.
- [29] D. Pearce, A. Valverde. The Logic of Here-and-there with Strong Negation Invited talk at the *DGF Workshop on Negation in Constructive Logic*, Dresden University of Technology, Germany, 2004.
- [30] D. Pearce. The Logic of Paraconsistent Answer Sets. Invited talk at the *Dagstuhl Seminar on Nonmonotonic Reasoning, Answer Set programming and Constraints*, 2005.
- [31] A. Pradera, E. Trillas. Aggregation and Non-Contradiction. *Proceedings of the European Symposium on Fuzzy Logic and Technology EUSFLAT'05* (to appear).
- [32] A. Pradera, E. Trillas. Aggregation Operators and the Non-Contradiction Principle. *Proceedings Aggregation Operators AGOP'05* (to appear).
- [33] J. M. Serrano, S. Ossowski. A semantic framework for the recursive specification of interaction protocols. Coordination Models, Languages and Applications, Special Track of the 20th ACM *Symposium on Applied Computing* (SAC 2005), pp. 456-451. ACM Press.
- [34] J. M. Serrano, S. Ossowski, S. Saugar. Reusability issues in the instrumentation of agent interactions. *Proc. of the Third international Workshop on Programming Multi-Agent Systems (PROMAS'05)*, Utrecht, the Netherlands.
- [35] E. Trillas, C. Alsina and A. Pradera. On MPT-implication functions for Fuzzy Logic. *Revista de la Real Academia de Ciencias. Serie A, Matemáticas*, 98(1):259-271, 2004.
- [36] E. Trillas, C. Alsina, E. Renedo and A. Pradera. On contra-symmetry and MPT-conditionality in Fuzzy Logic. *International Journal of Intelligent Systems* 20:313-326, 2005.
- [37] A. Valverde: **tabeq1**: A Tableau Based Suite for Equilibrium Logic. In Logics in Artificial Intelligence, JELIA'04, pages 734-737. *Lect. Notes in Artificial Intelligence* 3229, 2004.
- [38] A. Valverde. Reducing propositional theories to logic programs. Invited talk at the *Dagstuhl Seminar on Nonmonotonic Reasoning, Answer Set programming and Constraints*, 2005.

The incorporation of the Mathematical Morphology, the Formal Concept Analysis and the Fuzzy Logic techniques and tools to the data cleaning, aggregating, reducing and mining stages of the Knowledge Discovery process

TIC2003-08693

Ramón Fuentes-González*
Dpto. de Automática y Computación
Universidad Pública de Navarra
31006-Pamplona

Abstract

The principal aim of this project is to set up new mathematical models together with efficient associated algorithms that allow the incorporation of some elements of the Formal Concept Analysis (FCA), of the Mathematical Morphology (MM), of the Aggregation Functions Theory (AFT) and of the Fuzzy Relation Theory (FRT) in the following phases included in the Knowledge Discovery (KD) processes: data cleaning, data aggregating, data reducing and data mining (DM). Also, this project is intended to compare our algorithms with others associated with the classical methods in DM.

Keywords: Knowledge Discovery, Data Mining, Fuzzy Concept Analysis, Formal Concept Analysis, Fuzzy Mathematical Morphology, Fuzzy Relations.

1 The aim of the project

The Knowledge Discovery in databases (KDD) [1] process consists of the automated extraction of patterns representing knowledge implicitly stored in large databases, data warehouses and other massive information repositories. It is considered that the KDD process consists of an iterative sequence of the following steps: data cleaning, data integration (data grouping and data reduction), data selection, data transformation, data mining, pattern evaluation and knowledge representation [2].

* Email: rfuentes@unavarra.es

At present time, the KDD process is a multidisciplinary field [3], drawing work from areas including database technology, artificial intelligence, machine learning, neural networks, statistics, optimization, pattern recognition, fuzzy logic, soft computing, formal concept analysis [3], rough sets theory, knowledge-based systems, knowledge acquisition, information retrieval, high-performance computing and data visualization.

With this project it is tried to fulfil, in collaboration with the aforementioned ones, two new incorporations to that discipline list: the Mathematical Morphology (Crisp Theory [5] and Fuzzy Theory [6], [7], [8]) and the Fuzzy Concept Analysis [9], [10]. Specifically, our main objective is to set up some mathematical models and its associate algorithms that allow to incorporate elements of the disciplines: Fuzzy Concept Analysis, Fuzzy and Crisp Mathematical Morphology, Aggregation Functions Theory [11], and Fuzzy Relations Theory [13], in the following KDD steps: data cleaning, data grouping, data reduction and data mining.

The attainment of this objective will suppose the incorporation of some new ideas and methods in processes of KDD and Data Mining, as well as their comparison with the present mechanisms of KDD that are developed from the aforementioned different disciplines.

2 Level of reached success in the project

Next, a summary (including the fulfilment of the proposed tasks and the developed activities), is shown.

1. Fulfilment of the initially proposed plan of work.

The scheme of work associated to the used methodology which we proposed in the Scientist-technique Memory of the Project is specified in 14 tasks. To date, the corresponding part of that scheme includes the first eleven tasks. Next, we expose a summary of the made activities, detached by tasks.

Task 01: *Compilation of specialized information that concerns the project* and Task 02: *Analysis and evaluation of generic applications of DATA MINING (DM).* (Whose development was programmed during the first year of the project, 2004).

In the main, the work of the group of investigation related to these two tasks has been completed within the expected term, and it has consisted of:

01a) Compilation of the basic bibliography and of the works of investigation published in KNOWLEDGE DISCOVERY IN DATABASES (KDD) and DATA MINING (DM) specialized journals, fundamentally the one that more directly appears related to the topics including in the memory of the project. Although this task is considered fundamentally complete, it is open to the foreseeable updates that we will do during the period that remains to finalize the project.

The listing of those resources will be included in the final memory, specifying those that have influenced, inspired and complemented the results and investigations related to this project.

01b) Web location of permanent resources over subjects, investigations, events, new bibliography, software and groups of investigation related to the topics: KDD and DM in general, as well as in the specific case of resources in the field of THE FORMAL CONCEPT ANALYSIS applied to the extraction of knowledge in databases.

One of the main achievements obtained as a result of this task has been the incorporation of our group to the Spanish Network of Data Mining. This organization incorporates the most of the DM and KDD working groups in the Spanish Public Universities.

In addition, this activity has allowed to the group to get up to the lists of interest related to publications, events, meetings and location of new resources.

In the final memory of this project, a listing of these resources will be included, along with the details and explanations that we will consider pertinent.

02). With regard to the DM and KDD specific tools analysis, the following software was selected:

Utilities of data analysis with possibilities of graphical representation of the data or summaries of them:

- *3DV8 Enterprise Edition*, *Open VisiCube*, *Miner3D ACCESS*, *S-Plus* (language), *Axum 7*.

The three first utilities implement standards of the graphical data mining, so they will serve us as models for the algorithms that we will develop in the following phases of the project.

So much the *S-Plus* as *Axum 7* utilities have the possibility of implementing algorithms that directly manipulate data tables. (The frequent support of the data presentation in DM techniques).

All these applications have analyzed in Windows, although analogous utilities to *S-Plus* are available in UNIX and LINUX.

(i) Specifically designed utilities for data mining tasks:

- *Clementine*, *Insightful Miner 3.0*, *Weka-3-2*.

The *Clementine* and *Insightful Miner 3.0* utilities are selected since they implement the majority of the classic data mining algorithms.

We have selected those utilities in order to compare our algorithms, with the classic ones, these last implemented in *Clementine* and *Insightful Miner*.

Both are similar utilities, but the utility *Insightful Miner 3.0* has the advantage to use the *S-Plus* language directly.

Finally, *the Weka-3-2* is a specific tool for the investigation in DM. It is a free access utility developed in Java. We have selected it since it can become a standard of the algorithms on KDD and DM, at least at academic level, both in investigation and in implementation.

(ii) Specific languages for the data tables manipulation:

- *S-Plus* (language), *MICROSOFT ACCESS* (version 2003).

Since it has been mentioned in the section (i), we think that in addition to the graphical possibilities for the representation of data, the *S-Plus* language seems suitable for the implementation of algorithms that transform, (via morphologic operators), a data table in other one.

The initial files (on which the developed algorithms that implement the morphologic operators have been proven), have been recovered by means of the utility *Microsoft Access*, so the first versions of our algorithms are implemented in Access. Next, the *S-Plus* language will be used to obtain the later versions, since the structure of this last language will make a more efficient version of those algorithms.

(iii) Complementary utilities:

- *Toscanaj*, *Fuzzy Query*, *GUHA*.

The *Toscanaj* utility is a new version of *Toscana*, the standard utility for the analysis of data associated to contexts using the Theory of *Formal Concept Analysis* methods (*FCA*) This new *Toscanaj* version is implemented in Java, reason by which its integration with *Weka* is a feasible task that we will try to approach when the tables of instances will be interpreted as contexts.

We consider that the utility *Fuzzy Query* is useful to introduce an aspect that non appears in the specialized literature, and that we want to implement in our algorithms: the consideration of the *Fuzzy Logic* (*FL*) as a basic element for the models containing vagueness, circumstance that appears frequently in database analysis. This utility is designed specifically for fuzzy queries in structured databases (*FSQL*).

Finally, *GUHA* is a free access utility with implemented algorithms of analysis of conceptual data, algorithms that are of our interest in this project.

Let us indicate here that this process of search of suitable tools for our investigation, although sufficiently defined, it is not closed. We have verified in our access to the specialized sources on that the increasing interest in KDD subjects has done that many investigators of very diverse matters try to incorporate they tools to problems of data mining.

Task 03: Design of the mathematical model for the morphologic filters in data mining and Task 04: Expression of the morphologic filters for the phases of data cleaning and data grouping in terms of the Mathematical Morphology and the theory of Fuzzy Relations. (Whose development was programmed during the year 2004 and the beginning of the 2005).

This is the task in which we are working now. In particular, we have designed the algorithms that implement the basic filters for the treatment of data tables. Such basic filters are the *erosions* and *dilatations* of data tables, (represented by fuzzy subsets or fuzzy relations), using another fuzzy or crisp relation as a *neighbourhood relation* (or *structuring relation*).

Later, and following the techniques of the *Fuzzy Mathematical Morphology Theory* used in another context (the one of treatment of images [14]), the combination of those basic filters they have provided new filters such as *openings*, *closings*, *hit or miss operators*, etc.

Summary of results obtained in the development of the activities of tasks 03 and 04

With regard to tasks 03 and 04, we have developed and extended the ideas that previously appeared outlined in a published work [12] by members of this group. The purpose of this work is the application of the Mathematical Morphology tools and methods to handle the useful and possibly hidden information of the fuzzy and crisp relational systems.

In order to introduce the basic filters erosion and dilatation of data tables, (fuzzy sets $A \in L^E$, $B \in L^E$, ... or fuzzy relations $S \in L^{E \times F}$), by a neighbourhood relation, (a fuzzy relation $R \in L^{E \times E}$), we consider the usual composition operators $\circ_\tau$ and $\triangleleft_\sim$ in Fuzzy Logic [9] defined by:

$$\begin{aligned} (R \circ_\tau B)(x) &= \sup_{y \in F} (R(x, y) \tau B(y)), \\ (A \circ_\tau R)(y) &= \sup_{x \in E} (A(x) \tau R(x, y)). \\ (R \triangleleft_\sim B)(x) &= \inf_{y \in F} (R(x, y) \rightsquigarrow B(y)), \\ (A \triangleleft_\sim R)(y) &= \inf_{x \in E} (A(x) \rightsquigarrow R(x, y)). \\ (R \circ_\tau S)(x, z) &= \sup_{y \in F} (R(x, y) \tau S(y, z)), \\ (R \triangleleft_\sim S)(x, z) &= \inf_{y \in F} (R(x, y) \rightsquigarrow S(y, z)). \end{aligned}$$

Where L is:

- (i) A complete chain in $[0, 1]$ with the usual order or
- (ii) a complete lattice of closed intervals $[a, b]$ of $[0, 1]$ with the usual induced order.

And where τ and $\rightsquigarrow$ represent a t-norm and a fuzzy implication in L respectively [10].

With those elements, we extend the classical crisp Minkowski operators [11] in $\mathbb{R}^2$: sum ($\oplus$), and difference ($\ominus$) used in Mathematical Morphology [14], to fuzzy subsets by

$$\begin{aligned} A \oplus R &= R \circ_\tau A \quad \text{or} \quad S \oplus R = R \circ_\tau S \quad \text{and} \\ A \ominus R &= R \triangleleft_\sim A \quad \text{or} \quad S \ominus R = R \triangleleft_\sim S. \end{aligned}$$

And finally, we introduce the basic fuzzy morphological filters for data mining:

Erosion of the fuzzy set $\mathcal{A}$ by the structuring relation R : $\mathcal{E}(\mathcal{A}, R) = \mathcal{A} \ominus R^{\phi}$, been R^{ϕ} the opposite relation : $R^{\phi}(x,y) = R(y,x)$

Dilatation of the fuzzy set $\mathcal{A}$ by the structuring relation R : $\mathcal{D}(\mathcal{A}, R) = \mathcal{A} \oplus R$.

Opening and *closing* filters as composition of the last ones:

Opening of the fuzzy set $\mathcal{A}$ by the structuring relation R : $@(\mathcal{A}, R) = \mathcal{D}(\mathcal{E}(\mathcal{A}, R), R)$.

Closing of the fuzzy set $\mathcal{A}$ by the structuring relation R : $C(\mathcal{A}, R) = \mathcal{E}(\mathcal{D}(\mathcal{A}, R), R)$.

We have shown some interesting result about this operators in the data mining context.

Let us indicate finally in this section that we are introducing new operators (*difference operators*) as differences of the basic ones. Later we will undertake the study of new definitions that implement in our context the concepts of *skeleton* and *skeleton by zones of influence* [11], that are used in mathematical morphology in grouping processes.

With this idea , and since the skeleton concept is associate to *distances* between items, our present task is to analyze bibliography of distances defined between fuzzy and crisp subsets of instances.

Also we want to make notice that we analyze data tables in information systems, we have observed certain similarity between the topics of our investigations and the techniques of the Theory of "Rough Sets" (RS) of Pawlak [15]. This is a task that we did not contemplate in our initial project. But our intention is to make a meticulous analysis of those similarities, with the purpose of seeing the possibility of including the results of RS in our investigation.

Task 05: *Design of algorithms associated to the morphologic filters for data cleaning, data grouping and data mining* and Task 06: *First implementation of the algorithms, proof and debugger them* (planned between June of 2004 and July of 2005).

We have developed the algorithms for the formal expressions of the following filters: *erosion*, *dilatation*, *opening*, *closing*, *hit or miss* and *difference operators*. At date, this implementation has been made for crisp and fuzzy subsets and with crisp relations as structuring elements only . Those algorithms are made as generalized products of matrices. Later, we will implement the most complex case of filters obtained with structuring fuzzy relations that are not crisp, problem that it should be undertaken after making an analysis on the more favourable type of implication and t-norm operators. This one seems to be a laborious task, although it seems that, like in other many fields related to the Fuzzy Logic, the operators of *Lukasiewicz* are the most appropriate to define the erosions and dilatations basic filters.

Taking a free access database as an element of test with near 50,000 instances and 14 attributes, (database *Adult* or *Census Income* with information on the census of the United States in 1994), and using Microsoft Access, we have made one first implementation of a great part of the mentioned filters. At the moment, we are dedicated to the interpretation of these results.

At the same time, we are implementing the algorithms to those filters in the mentioned language S-PLUS.

These tasks will continue until end of October of 2005.

Task 07: *Expression of the procedures of data grouping and data mining in terms of the Fuzzy Concepts Analysis and the theory of Fuzzy Relations*, Task 08: *Design of algorithms associated to the previous procedures* and Task 09: *First implementation of the previous algorithms, proof and debugger of them.* (planned between January of 2005 and October of 2005).

In this section we have obtained some results that allow us to extract information from a fuzzy context in which there are unknown values. In this line, and in collaboration with the *Dpto of Matemática Aplicada de la Universidad del País Vasco*, we have proposed a method (using the idea of frequent sets) for the elimination of objects (instances) or attributes (fields) in fuzzy contexts in which these absent values appear. Moreover, we will set up the implications between intervals to replace the absent values and to limit the lack of information in the context.

Finally, we have used the idea of frequent items to define the frequent L-Fuzzy concepts that give us a method to organize the information obtained from a fuzzy context and we have also used the association rules to set up implications between the attributes of the fuzzy context.

As a result of these investigations we have presented a paper in the congress Estylf 04 in September of 2004 and other two have been accepted, one of them in the congress: *EUSFLAT-LFA 2005 Joint Conference* and another in *TAMIDA 2005 (III Taller Nacional de Minería de Datos y Aprendizaje*, organized by the Spanish network of Data Mining and Learning ([CICYT Tic2002-11124-e](#)). Both will be celebrated in September of this year 2005. The references of these works appear in the following section.

Task 10: *Study of the process of the knowledge fusion using Aggregation's Functions* and Task 11: *Design of algorithms associated to the previous procedures.* (Whose development was planned in the period April of 2005 until December of 2005).

The development of these activities will become in October of 2005, since the previous ones have lasted more than planned.

3 Indicators of the results

To date, and like result of the developed work related with our investigations, we have published the following communications:

1. A. Burusco, R. Fuentes-González and C. Alcalde. Use of rules of association in the theory of L-Fuzzy Concepts. *Estylf* 2004. Jaén (2004)
2. Alcalde, A. Burusco and R. Fuentes-González. Treatment of contexts with absent values. (admitted in TAMIDA 2005, III national Factory of Mining of Data and Learning, organized by the Red Española de Minería de Datos y Aprendizaje , Granada. September of 2005).
3. C. Alcalde, A Burusco and R. Fuentes-González. Treatment of the incomplete information in L-fuzzy contexts. (admitted in Eusflat-Lfa 2005 joint Conference, Barcelona. September 2005).

On the other hand, in a term of two months we will be able to send to the specialized journals, two works with the results that we have obtained related to the morphologic operators applied to data tables. These works will turn on the implementation of the filters mentioned in the summary that there are including in this report of the results corresponding to tasks 03 and 04.

4 References

- [1] Frawley, W.J., Piatetsky-Shapiro, G. and Matheus, C.J. Knowledge discovery databases: An overview. In: Piatetsky-Shapiro, G. and Frawley, W.J.(eds) Knowledge Discovery in Databases, AAAI/MIT, (1991) 1-27
- [2] Han J. and Kamber M. Data Mining. Concepts and Thecniques. Academic Press 2001
- [3] Hand, D., Mannila, H., Smyth, P. Principles of Data Mining. The MIT Press. 2001
- [4] Ganter, B., Wille, R. Formal Concept Analysis. Mathematical Foundations. Springer 1999.
- [5] Serra J. Mathematical Morphology Vol 1 and 2 Academic Press 1982
- [6] Bloch I. and Maître, H. Fuzzy Mathematical Morphologies: a comparative study, Télécom Paris 94D001
- [7] Burillo, P., Frago, N. , Fuentes-González, R. Generation of Fuzzy Mathematical Morphologies. Mathware & Soft Computing Vol. VIII, n 1 31-46 (2001).
- [8] De Baets, B. Fuzzy morphology: a logical approach, in Uncertainty Analysis, in B. Ayyub and M. Gupta (Eds.),
- [9] Burusco, A.; Fuentes-González, R. Concept lattices defined from implication operators. Fuzzy Sets and Systems 114 431-436 (2000).
- [10] Umbreit S. Formale Begriffsanalyse mit unscharfen Begriffen. Dissertation, Martin-Luther-Universität Halle-Wittenberg 1995.
- [11] Bouchon-Meunier, B. Aggregation and fusion of imperfect information. Heidelberg: Physica-Verlag 1998
- [12] Fuentes-González, R., Burillo, P. Y Frago, N. Técnicas de Morfología Matemática en el tratamiento de Sistemas Relacionales Difusos. ESTYLF2002. León (2002)
- [13] Kohout L., Bandler W. Use of fuzzy relations in knowledge representation, adquisition and processing. In Zadeh L. and Kacprzyk (Ed) Wiley 1992
- [14] Soille P., Morphological Image Analysis. Principles and Applications. Springer 1999
- [15] Pawlak, Z. Rough Sets. Kluwer 1991

Concept and Behavior Learning by Soft-Computing Techniques applied to the Mobile Robot Navigation problem TIC2003-04900

Raúl Pérez *

Department of Computer Science and Artificial Intelligence
E.T.S. de Ingeniería Informática.
University of Granada. 18071 - Granada (Spain).

Abstract

The development of intelligent and autonomous robots needs the incorporation of knowledge to its software architecture. This knowledge can be incorporated in the system by its designer or it can be acquired or tuned by the own system. This second option has the advantage that makes easier the design of the system and the obtained system presents a better adaptation to the work environment. In the case of the mobile robot navigation problem, we have developed a navigation architecture based on fuzzy behaviors, where the mobile robot can achieve a safe navigation inside of an indoor environment. With this architecture, the robot can detect some important elements of the environment and build a topological map to represent the structure of an office-like floor. However, in the design of the system, we have used a considerable expert information for implementing the rules that describe the control behaviors and for defining the perceptual objects that the robot can find out during its navigation. This fact implies that the robot has not the ability for modifying the knowledge presents in the control behaviors, object definitions and for identifying new objects. The main aim of this project is to research in algorithms based on soft-computing techniques to allow the robot be able to acquire new knowledge or tune the current knowledge to obtain an improved robot navigation architecture.

Keywords: Learning algorithms, Soft-Computing, Mobile Robots.

1 Objectives of the project

1.1 Initial hypothesis and main objective

The knowledge needed for the right performance of many systems, is hard to obtain in some cases as for example in the mobile robot navigation problem. The team that supports this project has worked, on one hand, on issues related to machine learning through the use of fuzzy rules. On the other hand, a software architecture for mobile robot navigation has been also developed. Thanks to this architecture a Nomad 200 mobile robot is able to navigate in a safe way within an office-like indoor environment. Taking into account these two previous

*Email: fgr@decsai.ugr.es

efforts in both areas (machine learning and robotics), the main objective of the project is to enhance the navigation architecture adding the skill of knowledge acquisition and learning. More specifically, the idea is the building of a knowledge acquisition module to be incorporated to the existent architecture in order to the robot be able to:

- learn new perceptual objects or tune its knowledge on certain perceptual object as the doors of the indoor environments.
- learn new control behaviors or tune its knowledge on the behaviors already implemented in the architecture.

The way for carrying out the knowledge acquisition process is intent to be autonomous in some cases and be supervised in other ones. That is, the human-robot interaction will be necessary in certain cases.

The initial hypothesis is supported by the experience of several members of the team on both machine learning and robotic areas. In regard to machine learning, the team has participated in different projects financed by the Spanish National Research and Development Plan: CICYT TIC 92-0665, CICYT TIC95-0453 and CICYT TIC96-0778, being a member of the team the main researcher in the two last projects. In these projects we have developed several machine learning techniques. For this project, we would like to distinguish the interesting properties of an inductive learning algorithm based on the iterative genetic model that allows us to describe the behavior of a system through a fuzzy rules set.

The team has also worked in the European project GENESYS: Fuzzy controllers and smart tuning techniques for energy Efficiency and overall performance of HVAC systems in buildings, European Commission, JOULE Program Code JOE3CT980090. In this project several tuning techniques were developed to tune fuzzy controllers applied to air conditioning equipments.

In regard to the robotics area, we have participated in the project EUROBOT-IBERIA: Network on Physical Agents TAP99-1354-E, Planet: European Network on Planning that has a section dedicated to robotics and in the Andalusian Network on Robotics. In this area we have developed a behavior-based architecture for mobile robot navigation.

1.2 Specific objectives

In order to achieve the main objective shown above, there exist several more specific objectives. They are described following:

1. Development and implementation of sensor data fusion techniques, with the treatment of uncertainty.

The robot has several sensory systems as for example infrared and ultrasound systems. In this project, a camera will be added to the robot to obtain visual information of the environment. The analysis of the visual information supplied by the camera will be the most important aspect in this phase. Furthermore, fuzzy logic will be used to deal with the uncertainty and vagueness as well as to facilitate the data fusion process.

2. Development and implementation of tuning and learning techniques of perceptual objects.

The idea is that the robot be able to detect perceptual objects from its sensorial system. As a perceptual model based on ultrasonic and infrared data have been already developed by the team, therefore, the attention will be focused on visual data. Fuzzy logic and genetic algorithms will be applied to set the perceptual features of certain objects like the doors.

3. Development and implementation of learning techniques of fuzzy rules applied to the control behaviors.

The system will be able to adapt the implicit knowledge existent in the rules that describe the control behaviors of the robot. Thus, it will be possible to achieve a more robust and accurate behavior of the system.

4. Integration of the new developed techniques within the navigation architecture.

The knowledge acquisition module will be inserted within a deliberative-reactive architecture to achieve a safe and robust navigation of the robot. It is hoped that the robot will enhance its perception capacities, map building process and skills to reaction under changes in the environment.

1.3 Tasks and planning

The planning of the project is divided in six phases to carry out during three years. Firstly it would be done a general analysis where all the general and particular (to be achieved in each phase) requisites of the system will be specified. The next four phases correspond to the four specific objectives to achieve and the last one will be of experimenting and testing of the whole system. In the last phase it should be possible to verify the improvement incorporated to the system thanks to the new knowledge acquisition module. Also, new aims could be defined to achieve in a future work. The six phases will be the following:

- Phase 1 - General analysis of the problem.
- Phase 2 - Development of sensor data fusion techniques.
- Phase 3 - Development of tuning and learning techniques of perceptual objects.
- Phase 4 - Development of tuning and learning techniques of fuzzy control behaviors.
- Phase 5 - Integration into the navigation architecture.
- Phase 6 - Global system testing.

Likewise, each phase corresponds to certain tasks that will be accomplished through the following planning.

1.4 Phase 1.

To carry out in the months 1 - 4 in the first year.

- T1.1. Analysis of the general requisites as well as for each one of the phases.
- T1.2. Definition of the interface between each one of the phases.

1.5 Phase 2.

To carry out in the months 5 - 12 in the first year and 1 - 2 in the second year.

- T2.1. Acquisition of the vision equipment and the pan-tilt and its installation in the Nomad200 robot.

- T2.2. Development and implementation of the techniques for extraction of characteristics, segmentation and shape recognition applied to visual data.
- T2.3. Analysis of the problem of data fusion and combination of visual information with distance information.
- T2.4. Development and implementation of methods based in Soft-Computing for data fusion.
- T2.5. Integration of the information supplied by the vision system with the current perception model.
- T2.6. Test of the techniques developed at this phase.

1.6 Phase 3.

To carry out in the months 1 - 10 in the second year.

- T3.1. Analysis of the learning problematic applied to the classification of perceptual information about the objects.
- T3.2. Adaptation of the existing techniques that were already used in previous projects or the development of new learning methods.
- T3.3. Implementation of the chosen techniques for the learning or tuning.
- T3.4. Tests to the impact of the learning or tuning in the capacity of the system.

1.7 Phase 4.

To carry out in the months 9 - 12 in the second year and 1 - 4 in the third year.

- T4.1. Analysis of the learning and tuning methods applied to the design of fuzzy behaviors in mobile robots.
- T4.2. Adaptation of existing techniques that were used in previous projects, or the development of new methods.
- T4.3. Implementation of techniques for the learning or tuning. This will be done in simulation.
- T4.4. Validation of the obtained results by tests in the real world with the new rule databases.

1.8 Phase 5.

To carry out in the months 1 - 10 in the third year.

- T5.1. Incorporation of the knowledge acquisition module in the architecture.
- T5.2. Adaptation of the execution and monitorization module.
- T5.3. Adaptation of the map building process.
- T5.4. Adaptation of the localization method.
- T5.5. Adaptation of the planning level.
- T5.6. Tests the new modifications in the architecture.

1.9 Phase 6.

To carry out in the months 10 - 12 in the third year.

- T6.1. Building of the topological map of the environment.
- T6.2. Development of complex navigation tasks where requisites defined in the analysis phase are needed.
- T6.3. Tests using those navigation tasks.
- T6.4. Evaluation of the test results and accomplishment of all objectives.
- T6.5. Definition of future work.

2 Level of success achieved in the development of the project

2.1 Phases 1 and 2

In the phase 1, the general analysis of the problem was carried out defining the requirements of the other phases and the interface among them. In the phase 2, the vision system was installed in the Nomad 200 robot and the visual information was dealt with as a new sensorial system.

Thanks to the incorporation of the visual information to the system, we have achieved in this second phase, a multi-agent system structured in three layers that uses a concurrent activation of both visual behaviors and fuzzy behaviors for the Nomad 200 navigation. The higher layer receives order from an external operator and generates a high level of abstraction plan consisting in the sequence of rooms and corridors to transverse and doors to cross to go to a desired place. This plan is decomposed in the middle layer in a sequence of skills that the robot is able to accomplish. These skills arises from the interaction of visual and fuzzy behaviors implemented in the lower layer of the architecture. Fuzzy behaviors takes as input the information provided by the range sensor and uses fuzzy rules to manipulate the imprecision and vagueness of the sensor data allowing a safe navigation to accomplish each part of the plan. Visual behaviors gathers the visual information from a single camera placed at the top of our robot and allows the system to know the location of the doors of the environment and indicate the straight direction to cross it. The camera is placed over a pan-tilt unit (PTU) that allows to move the camera independently from the movements of the robot giving to vision an active role in our system. We assume that all doors are opened because our robot cannot open them. To ease the door detection, artificial landmarks with numerical information in it have been placed beside the doors. Once the robot detects the landmark it must keep watching at the landmark indicating the straight direction to the door.

The tasks described in Sections 1.4 and 1.5 have been completed since the robot is able to manage ultrasound and visual information using both distance and visual data. The information has been dealt with through Soft-computing techniques in order to manage the uncertainty and vagueness present in the sensor data. Some experiments have been carried out to validate the system. This test and others have been filmed and can be downloaded at <http://decsai.ugr.es/~eaguirre/research/Videos.htm>.

2.2 Phase 3 and 4

In phase 3 we have tried to characterize some basic objects to facilitate the robot navigation. In previous works some objects like walls, corridors, corners and doors have been defined

using a ultrasound-based model. However this perceptual model is affected negatively by the drawbacks of the ultrasound sensors. In this project we have focused our effort in the use of visual information to detect the doors of the environment since the doors are a very important kind of perceptual object in office-like indoor environments. Using the information provided by the camera placed on our Nomad 200 mobile robot we are going to establish a new belief degree on the existence of doors in indoor environments. This belief degree, obtained using vision, can be easily integrated with other values that the robot can obtain applying the fuzzy perceptual model previously developed.

In this phase we develop a new method for visual door-detection based on the extraction of the segments of the image and the definition of several fuzzy concepts. The segments of the images are analyzed in order to detect which of them belong to the frame of a door. The fuzzy concepts *Simple Frame* (SF) and *Double Frame* (DF) are defined to differentiate those situations in which the frame edges of a door can be seen. Features of the segments like size, direction or the distance between them are measured and analyzed using fuzzy logic in order to establish a membership degree on the defined fuzzy concepts. In a first step, the edges of the image are extracted using Canny, then the Hough Transform is calculated to extract the main segments of the image. Using fuzzy logic, the relationship between the segments is measured to detect the presence of the concepts SF and DF. The method is able to detect typical doors of indoor environments under variable illumination conditions in grey-level images. The method can detect the doors under strong perspective deformations caused by the two degrees of freedom (DOF) allowed for the camera of our robot. Furthermore, according to our experiments it is suitable for real-time applications in a mobile robot Nomad 200 equipped with a vision system and a laptop.

The use of fuzzy logic has allowed, on one hand, to limit the search space of the set of possible segments and to define in a natural way fuzzy concepts. On the other hand, it allows to analyze the segments with different directions and size obtaining a high degree of flexibility in the manipulation of the information contained in the images. This flexibility is necessary to take into account the different perspective deformations caused by the two DOF of our camera.

The experiments carried out show that the proposed method can detect effectively the door of our environment with a high degree of success (92%). We consider that the method is valid for a great variety of doors commonly present in indoor environments.

However this system has several parameters that have been established in an experimental way and using expert knowledge. In order to obtain a more robust and general system some automatic mechanism is needed. In our project we have used evolutionary algorithms to solve this problem. That is, we use genetic algorithms for tuning the membership functions of the visual fuzzy system previously developed by us for door-detection in indoor environments. In fact, the system is formed by different set of fuzzy rules each one with several parameters to take into account. Furthermore, the objectives to obtain are two: On one hand, it is desirable a high detection rate in the images with doors. On the other hand, a low false positive rate is necessary in images where no doors are present. Thus we have to solve a multiobjective problem.

The use of evolutionary algorithms to solve multiobjective problems has experimented a great advance with the introduction of the *Pareto-based fitness* concept. The idea consist of ranking the individuals of the population according to the proximity to the Pareto-optimum. This approach has been referred in the literature by the term *EMO* (Evolutionary Multiobjective). The main advantage of EMO algorithms is the ability of simultaneously searching several non-dominated solutions. EMO algorithms finds in a single run different non-dominated solutions to a problem while several runs of a normal algorithm would be required to obtain similar results.

In this project, we make use of the EMO SPEA2 for tuning the membership functions of

the visual fuzzy system previously developed. The visual fuzzy system uses a hierarchy of fuzzy classifiers that finds, among the segments of an image, those that belong to the frame edges of a doorframe. It can be used on autonomous mobile robots for navigating, map-building and positioning purposes. The variables used, the number of membership functions for each variable and the rule base of the fuzzy systems have been determined based on expert knowledge. Nevertheless, instead of selecting manually the range of each membership function, they have been uniformly distributed covering the domain of each variable. The tuning process will adapt the membership functions of the visual fuzzy system to the conditions of our environment, i.e., to the usual conditions of height and size of the doors of the environment and to the distances and orientations under which they are seen by the robot. The tuning of the 59 parameters of the fuzzy visual system is performed using SPEA2. The use of SPEA2 has allowed the concurrent optimization of the two objectives: increase of the true positive detection and reduction of the false positive detections. As the experimental results show, the technique allows us to adapt the initial visual fuzzy system increasing its success percentage on the images of our database from 56.15% to 95.48%.

Related to the tasks described in Section 1.6, we have focused our attention in the definition of the perceptual object “door” using fuzzy rules and tuning the system properly. A similar process will be done for other perceptual objects. Furthermore, we are working in the adaptation of other learning techniques.

The current state of the project is the beginning of the phase 4 although some tasks of phase 3 are still active.

3 Result rates

The results obtained in the development of the project have been shown in several national and international conferences as it is described below. Also, the project is being useful to allow the training of a post-graduate student that is carrying out his PhD. Thesis into the scope of the project. This student has obtained a grant of the Spanish Ministry of Education to carry out his post-graduate students.

3.1 International conferences

- Aguirre E., García-Silvente M., Gómez M., Muñoz-Salinas R., Ruiz C. A multi-agent system based on active vision and ultrasounds applied to fuzzy behavior based navigation. In Proc. of 10 Th International Symposium on Robotics and Applications (ISORA 2004) (En CD-ROM), Sevilla, 2004.
- Muñoz-Salinas R., Aguirre E., García-Silvente M. and González A. Door-detection using computer vision and fuzzy logic. In Proc. of 6th WSEAS International Conference on Mathematical Methods & Computational Techniques in Electrical Engineering (En CD-Rom ISBN: 9608457076), Atenas (Grecia), 2004.
- Muñoz-Salinas R., Aguirre E., García-Silvente M. and González A. Tuning a visual fuzzy system using SPEA2. In Proc. of First International Workshop on Genetic Fuzzy Systems (GFS'2005), pp. 37-43, Granada, 2005.
- Aguirre E., González A., Muñoz-Salinas R.. Mobile Robot Map-Based Localization using approximate locations and the Extended Kalman Filter. Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2004) pp. 191 - 198. Perugia (Italia), 2004.

- Muñoz-Salinas R., Aguirre E., Garcia-Silvente M. and Gonzalez A., A Fuzzy System for Visual Detection of Interest in Human-Robot Interaction. Accepted in the 2nd International Conference on Machine Intelligence (ACIDCA-ICMI'2005), Tunisia.
- Muñoz-Salinas R., Aguirre E., Garcia-Silvente M. and Gonzalez A., People detection and tracking through stereo vision for robot-human interaction. Accepted in the Mexican International Conference on Artificial Intelligence (MICAI). Monterrey (México), 2005.

3.2 Spanish national conferences

- Aguirre E., Gómez M., Muñoz-Salinas R., Ruiz C. Un Sistema Multi-Agente que emplea Visión Activa y Ultrasonidos aplicado a Navegación con Comportamientos Difusos. Actas del IV Workshop de Agentes Físicos (WAF'2003), pp 63 – 74, Alicante, 2003.
- Aguirre E., García-Silvente M., González A., Muñoz-Salinas R. Detección de puertas mediante información multisensorial y su aplicación a la navegación de robots móviles. Actas del V Workshop en Agentes Físicos (WAF'2004), pp 33 – 42. Gerona, 2004.
- Aguirre E., García-Silvente M., González A., Muñoz-Salinas R. Detección de puertas mediante visión y lógica difusa. Actas del XII Congreso Español sobre Tecnologías y Lógica Fuzzy, pp. 389-394, Jaén, 2004.

3.3 Papers in International Journals

- Aguirre E., Argudo J. F., González A. y Pérez R., A Fuzzy Perceptual Model for Map Building and Navigation of Mobile Robots. Integrated Computer-Aided Engineering, vol. 11:3, pp. 239-258, 2004.
- Muñoz-Salinas R., Aguirre E., García-Silvente M. and González A., Door-detection using computer vision and fuzzy logic. WSEAS Transactions on Systems, Volumen: 10 n. 3 pp. 3047 – 3052, 2004.
- Muñoz-Salinas R., Aguirre E., García-Silvente M. and Gómez M., A Multi-agent system architecture for mobile robot navigation based on fuzzy and visual behavior. Robotica (Cambridge University Press). To appear in 2005.

3.4 Other activities of the member of the team

- A member of the team participated in the Organization Committee of the “First Workshop on Genetic Fuzzy Systems (GFS)” maintained from 17th to 19th March in Granada.
- Several members participate in the Program and Organization Committee of the “VI Workshop en Agentes Físicos” (WAF2005) within of the “I Congreso Español de Informática” (CEDI2005) that will be maintained in Granada.

Two research stays in foreign centers are planned:

- The member Muñoz-Salinas R., will go to the Advanced Mobile Robots & Intelligent Agents Laboratory. De Montfort University, Leicester (Reino Unido) from 30th June to 30th September 2005.
- The Dr. García-Silvente M., will go to the Instituto Tecnológico de Monterrey de México (ITESM) during a month in August 2005.

Architectures, Compilers and Applications in Multiprocessors TIC2003-06623

Emilio López Zapata *

Depto. de Arquitectura de Computadores
Escuela Técnica Superior de Ingeniería Informática
Universidad de Málaga

Abstract

The high performance experienced by current computers is the result of combining advances in fabrication technologies (microelectronics, optics and magnetic), in architecture (new design techniques) and in compilers. All of these allows to adapt different components, with different speeds, between themselves with minimum performance degradation of the fastest components. We are all convinced that this overall performance will increase in the coming years. However new computer applications show up in the horizon that demand for higher computational power, higher storage capacity and faster communications.

In this project we will explore in depth the new multiprocessor systems (clusters) and Grid computing, considering preferentially the important problems around mass storage and data management and processing, typical of multimedia applications. We will approach this subject from three main research lines: 1) Automatic optimization (parallelism, locality) of irregular and dynamic applications, 2) mass storage, multimedia, graphics and video processing architectures, and 3) emerging applications, in fields like audiovisual information, bioinformatics and numerical simulation.

Keywords: Supercomputing, Parallel Computing, Optimizing Compilers, Advanced Architectures, Data Locality Exploitation, Parallelism Exploitation, Storage Architectures, Multimedia, Graphics Processing, Video Processing, Bioinformatics, Parallel Numerical Algorithms.

1 Project Goals

This report is organized according to the three main research lines that were defined in the initial project proposal (and specified at the end of the abstract of this document). The specific goals of each research line are summarized in this section. It is important to note that this project is the natural continuation of the research accomplished in previous projects, that combined basic and applied research.

*Email: ezapata@ac.uma.es

1.1 Automatic optimization (parallelism, locality)

This line of the project is focused in the design and development of new techniques to optimize large and complex scientific and engineering codes (irregular and dynamic, or pointer-based, codes).

a) Automatic optimization/parallelization of irregular and dynamic codes. Current compilers usually fail when optimizing dynamic codes, that include pointer-based, complex data structures and use dynamic memory allocation. Such compilers are unable to extract from the code the information needed to exploit important performance aspects, like parallelism and locality. Many of the emerging applications that we study in the third research line, for instance, need from a deep analysis of the data structures and memory references in order to optimize them. So data analysis, that includes data organization and memory reference pattern, is a key step in the optimization of dynamic codes. There has been much research in this topic, allowing some effective compiler optimizations. However, these results are not enough to tackle more aggressive optimizations. In previous years we have developed new techniques based on shape analysis in order to capture the main characteristics of pointer-based data structures. This module is focused in advance such techniques and propose new ones to allow effective compiler phases, like interprocedural analysis and data dependence analysis, for dynamic codes. In addition, these new techniques will be used to develop new strategies for exploiting parallelism and locality.

1.2 Architectures

This line is focused in the design of new architectures for specific applications.

a) Distributed system for video on demand (VoD). A CDN (Content Distribution Network) is an economically viable solution for a VoD system. A CDN consists of a distributed system of servers interconnected through a specific network. CDNs present diverse difficult problems, like content replication, mass storage consumption, network traffic, and so on. This module proposes to develop a distributed VoD system where the video archives are partially replicated based on their popularity index. In addition different multicast techniques will be evaluated.

b) Mass storage architectures. The fast development of digital multimedia contents during the last years have entailed strong requirements on mass storage. Robust, scalable mass storage systems are unavoidable in this scenario. Typically, a scalable storage system store a small amount of the whole multimedia contents in the secondary level (hard disks), usually the most recently/frequently used data. The rest of the data is stored in a tertiary level (sequential magnetic tapes), of large capacity. This module is focused in the design of a scalable, massive storage architecture model, able to cover the needs of the current and future storage-demanding applications (multimedia, bioinformatics, CFD, astronomy ...).

c) Arithmetic and multimedia processors. Arithmetic hardware for evaluating elementary functions usually works with input arguments within a specific value interval. So when evaluating such functions, a rank reduction of input arguments are required before using the arithmetic hardware. This rank reduction is usually made by means of a series of multiplications and divisions. However, when the input values are very large, the precision of resulting reduced values may be compromised. This module proposes to design a specific hardware to accomplish

the argument rank reduction in a reliable means. On the other hand, hardware for multimedia extensions attached to processors has resulted in new possibilities to be developed. In this module we also study the design of arithmetic elementary functions using the hardware for multimedia extensions.

d) MPEG-4 codec architectures. MPEG-4 is a recent standardized video compressed format. Low profiles of this standard can be covered by current processors. However, medium and high profiles cannot be currently fully developed due to the great computational power required. This module is focused in the design of specific architectures for medium and high profile MPEG-4 codification/decodification.

1.3 Emerging applications

This line is focused in the study of a selection of applications that demand large computational resources.

a) Audiovisual information systems. Storage and management of audiovisual information are experiencing great changes in the last few years, in the context of digitalization of video/audio contents. In fact, recent standards, like MPEG-4, proposes new efficient methods for encoding multimedia data, including content-descriptive elements. In this context, it is clearly needed a common model to describe audiovisual information in a way useful for applications in diverse areas (multimedia, biology, astronomy ...). This module is focused in the design of a series of analysis techniques for processing audiovisual information, like those needed for the detection of key frames in a compressed video (storyboard), based on some characteristics. Or those needed to detect and extract objects in compressed video.

b) Genomics and proteomics. Recent advances in genomics entail an increasingly important role for bioinformatics research, in contrast to applied research in the laboratory. In fact, at present the huge volume of data generated in laboratories requires the development of new and efficient computational solutions for data storage and classification. This module proposes the design and implementation of software architectures for storage, management and analysis of biological data.

c) Simulation of flexible objects, and other scientific/engineering problems. The simulation of flexible objects is an essential tool for realistic computer animations of virtual humans and dynamic scenarios. New technologies, like digital TV and multimedia devices, require the development of applications for cheap, (near) real-time virtual reality simulations. This module tackles this problem from three approaches: the mathematical model, the solution of the equation systems, and the optimization techniques for high-performance computation. In this module we also include research about the design of efficient solutions to other scientific and engineering problems, using mainly parallel computation.

2 Level of Success Reached in the Project

At the time of writing this report, the project has consumed 20 months out of its whole three-year period. At present, a large part of the proposed goals have already been accomplished. In this section we summarize the main research achievements reached until now.

2.1 Automatic optimization (parallelism, locality)

a) Automatic optimization/parallelization of irregular and dynamic codes. We have proposed a methodology to develop efficient parallelization techniques for irregular and dynamic applications [5, 39, 40]. Based on this methodology, a framework using shape analysis was designed to capture properties of pointer-based dynamic data structures [8]. This framework was partially used to propose new strategies to exploit locality and parallelism for such codes [16, 33]. A key data analysis component for such optimizations is an effective data dependence analysis for pointer-based data structures [9, 32, 51]. We have also developed a tool for performance analysis of the cache [11], new techniques for efficient parallelization of irregular reductions [15, 17], and a cache model to design efficient locality-based loop transformations [18, 19]. Finally, we have experimented with a novel approach for exploiting the graphical resources (GPU) of a commodity PC for different applications, like fast retrieval of indexed data [57] and fast execution of irregular computations [58].

2.2 Architectures

a) Distributed system for video on demand (VoD). We have proposed hybrid multicast algorithms inspired by the Maximum Factored Queue Length (MFQL) batching scheme used to decide which video queue will be serviced with a multicast session [13]. We have also designed a new threshold-based patching scheme to control partial stream transmission during an ongoing multicast session [14].

b) Mass storage architectures. New storage technologies, like SAN and iSCSI, are essential for the deployment of audiovisual information systems. Apart from the high storage capacity, the access performance to huge archives is a critical issue. The PhD research of one component of the group tackles this problem. To date, we are developing in collaboration with Tedia an automatic managing system for distributed mass storage in the context of professional TV.

c) Arithmetic and multimedia processors. We have used the CORDIC algorithm to develop different solutions in the arithmetic field: an improvement of the pipelined CORDIC with linear approximation to rotation [4], a CORDIC processor for variable-precision coordinates, that permits a reliable accurate evaluation of elementary functions [27], and a new range reduction technique optimized for CORDIC [26]. We have studied on-line arithmetic for digital signal processing: a design of a comparison operation for signed digit representations [24], a proposal of an on-line full-adder to deal with multioperands [62], and an VLSI architecture for the integer-to-integer wavelet transform [25]. Finally, we have analyzed how to use multimedia extensions to compute elementary functions based on polynomial approximations [6].

d) MPEG-4 codec architectures. We have proposed a FPGA design for fast computing of the minimum sum of absolute differences (SAD), used in block based motion estimation algorithms in MPEG [35, 34, 23].

2.3 Emerging applications

a) Audiovisual information systems. We have developed a video content analysis and metadata organizational system for research videos arising from biological microscopy of living cells [42, 43, 44]. We have also designed solutions to the temporal segmentation of MPEG compressed

video sequences, based on different video characteristics, like edges, luminance and gradual transitions [49, 50]. We have proposed new methods for image processing, like a fully automatic algorithm for extraction of vessels in noisy medical images [59, 60], and a time-frequency method for texture segmentation and texture inhomogeneities detection [22].

b) Genomics and proteomics. We have developed an architecture for XML-based mediator systems and a framework for helping system developers in the construction of mediator-services to integrate heterogeneous data sources in the biological field [1, 2, 3]. We have proposed a mathematical model to extend the dynamic range of gene expression data measured by laser scanners [12]. We have also studied applications of knowledge discovery technology (data mining) to bioinformatics. Basically, we combine exhaustive database searching and automatic pattern recognition strategies with discovering of association rules [41, 7]. Two surveys related to bioinformatics were also accomplished in this period. One is about computational strategies to parallelize most of the typical applications in bioinformatics, specially those based in metaheuristics approaches [55]. A second one studies grid technologies and architectures, and how they are being used in biological sciences [52]. Related to this, a new adaptive task scheduling model was designed for computational grid environments [20, 21], as well as a new component (service) of a computational grid, AnaGram, was developed for protein function assignment [38].

c) Simulation of flexible objects, and other scientific/engineering problems. We have developed efficient parallel solutions to irregular reductions in the context of physically-based fabric simulators. Such operations appear when evaluating forces, a very time consuming task [45]. In the context of simulation, we have proposed a three-level (optical, energetics and physiological) model for simulating the interaction between laser irradiation and human tissues [61, 47, 10]. We have also analyzed different parallel implementations of the (preconditioned) conjugate gradient method for band systems, for several architectures and programming paradigms [46, 36]. From our experience in parallelizing complex applications, we have proposed an uniform interface for the design of efficient hybrid (shared-memory and message-passing) parallel codes [37].

3 Result Indicators

a) Publications

It should be highlighted that this research has already produced an important amount of results that have been published in the most prestigious journals and conference proceedings of the field (many from IEEE and ACM). In summary, during this period, the project produced: 19 publications in journals listed in JCR, 31 papers in international conferences, 10 papers in national conferences, and 2 book chapters.

b) PhD Thesis

Edmundo Sáez Peña, "Automatic Video Segmentation", 2005 (pending final date).

Sonia González Navarro, "Multi-server System for VoD Service", 2005 (pending final date).

In addition, the ongoing research of PhD candidates working in the project should produce three more PhD thesis by the end of the project (december 2006).

c) Cooperation with other groups

An important part of the research described in section 2 is being developed in cooperation with

several prestigious, world-level research groups: Prof. David Padua's group, from the Dept. of Computer Science, Univ. of Illinois at Urbana-Champaign, USA (section 2.1.a), Prof. Joel Saltz's group, from the Biomedical Informatics Dept., The Ohio State Univ., USA (section 2.1.a), Prof. Tomas Lang, from the Dept. of Electrical and Computing Engineering, Univ. of California at Irvine, USA (sections 2.2.c and 2.2.d), and Dr. José M. Carazo's group, from the CNB (Centro Nacional de Biotecnología) (section 2.3.b).

d) Participation in European projects

The group is participating in the European Network of Excellence on High-Performance Embedded Architectures and Compilers (HiPEAC), Project No.: IST-004408.

The group is also participating in the European Project "Express-Fingerprints, Expression Profiles as Fingerprints for the Safety Evaluation of New Strains Including GMOs Used in Bioprocessed Food", Project QLRT-2000-01473.

e) Technology transfer activities

Members of the research group have a permanent collaboration with Tedral company, being in charge of its R&D and transferring technologies appropriate for the analysis and automatic indexing of video, audio and images, as well as the design of high performance processing and mass storage architectures. A similar level of collaboration is being carried out with Integromics company, specialized in the design of bioinformatics applications for genomics and proteomics.

References

- [1] J.F. Aldana, M. Hidalgo-Conde, I. Navas, M.M. Roldán, P.A. de Alarcón and O.Trelles, *An Open Architecture for the Dynamic Integration of Biological Data Sources*, Int'l. Conf. on Computing, Communications and Control Technologies (CCCT 2004), Austin, USA, 14-17 August 2004, pp. 239-244.
- [2] J.F. Aldana, M. Hidalgo-Conde, I. Navas, M.M. Roldán, and O. Trelles, *Bio-Broker: A Biological Data and Services Mediator System*, IADIS Int'l. Conf. Applied Computing 2005, Algarve, Portugal, 22-25 February 2005.
- [3] J.F. Aldana, M. Roldán, I. Navas, A.J. Pérez and O.Trelles, *Integrating Biological Data Sources and Data Analysis Tools Through Mediators*, 19th Annual ACM Symposium on Applied Computing (SAC 2004), Nicosia, Chipre, 14-17 March 2004, p. 127.
- [4] E. Antelo and J. Villalba *Low Latency Pipelined Circular CORDIC*, 17th IEEE Symposium on Computer Arithmetic (ARITH-17), Cape Cod, Massachussets, USA, 27-29 June 2005.
- [5] R. Asenjo, F. Corbera, E. Gutiérrez, M.A. Navarro, O. Plata and E.L. Zapata, *Optimization Techniques for Irregular and Pointer-based Programs* (Invited Talk), 12-th Euromicro Conference on Parallel, Distributed and Network based Processing (PDP'2004), La Coruña, Spain, 11-13 February 2004.
- [6] G. Bandera, M. González, J. Villaba, J. Hormigo and E.L. Zapata, *Evaluation of Elementary Functions Using Multimedia Features*, 18th Int'l. Parallel and Distributed Processing Symposium (IPDPS'04), Santa Fe (NM), USA, 26-30 April 2004.
- [7] P. Carmona-Sáez, M. Chagoyen, A. Rodríguez, O. Trelles, J.M. Carazo and A. Pascual-Montano, *Integrating GO annotations with Expression Data to Analyze Microarray Experiments*, ECAI- 2004 Workshop on Data Mining in Functional Genomics and Proteomics: Current Trends and Future Directions, Valencia, Spain, 23 August 2004.
- [8] F. Corbera, R. Asenjo and E.L. Zapata, *A Framework to Capture Dynamic Data Structures in Pointer-Based Codes*, IEEE Transactions on Parallel and Distributed Systems, Vol. 15, No. 2, February 2004, pp. 151-166.
- [9] F. Corbera, A. Navarro, R. Asenjo, A. Tineo and E.L. Zapata, *A New Loop-Carried Dependence Detection Approach for Pointer-Based Codes*, XV Jornadas de Paralelismo, Almería, Spain, 15-17 September 2004, pp. 432-437.
- [10] V. Espigares, J. Veintimilla, M.A. Trelles, O. Trelles and L.F. Romero, *Simulación en Tiempo Real de la Acción del Láser sobre Tejidos Vivos*, XV Jornadas de Paralelismo, Almería, Spain, 15-17 September 2004, pp. 150-155.
- [11] B.B. Fraguera, R. Doallo, J. Touriño and E.L. Zapata, *A Compiler Tool to Predict Memory Hierarchy Performance of Scientific Codes*, Journal of Parallel Computing, Vol. 30, No. 2, February 2004, pp. 225-248.
- [12] J. García de la Nava, S. van Hijum and O. Trelles, *Saturation and Quantization Reduction in Microarray Experiments Using Two Scans at Different Sensitivities*, Statistical application in genetics and molecular biology, Vol 3. Iss. 1, 2004, Art. 11.

- [13] S. González, A. Navarro, J. López and E. Zapata, *Two Hybrid Multicast Algorithms for Optimizing Resources in a Distributed VoD System*, Int'l Conf. on Multimedia Modelling (MMM'2004), Brisbane, Australia, 5-7 January 2004.
- [14] S. González, A. Navarro, J. López and E. Zapata, *A Threshold based Multicast Technique in a Distributed VoD System with Customer Reneging Behavior*, 10th Int'l Conf. on Distributed Multimedia Systems (DMS'2004), San Francisco Bay, USA, 8-10 September 2004.
- [15] E. Gutiérrez, O. Plata and E.L. Zapata, *Data-Partitioning Based Parallel Irregular Reductions*, Concurrency and Computation: Practice and Experience, Vol. 16, No. 2-3, February-March 2004, pp. 155-172.
- [16] E. Gutiérrez, S. Romero, O. Plata and E.L. Zapata, *Explotando Localidad en Códigos basados en Punteros*, XV Jornadas de Paralelismo, Almería, Spain, 15-17 September 2004, pp. 477-482.
- [17] E. Gutiérrez, S. Romero, L.F. Romero, O. Plata and E.L. Zapata, *Parallel Techniques in Irregular Codes: Cloth Simulation as Case of Study*, Journal of Parallel and Distributed Computing, Vol. 65, Iss. 4, April 2005, pp. 424-436.
- [18] E. Herruzo, G. Bandera and O. Plata, *Applying Loop Tiling and Unrolling to a Sparse Kernel Code*, Int'l. Conf. on Computational Science (ICCS'2004), Krakow, Poland, 6-9 June 2004.
- [19] E. Herruzo, G. Bandera, O. Plata and E.L. Zapata, *Reducing Cache Misses by Loop Reordering*, Int'l. Conference on Parallel Computing (ParCo2005), Málaga, Spain, 13-16 September 2005.
- [20] M. Hidalgo, J. García de la Nava, A. Rodríguez and O. Trelles, *Planificación Adaptativa en Entornos GRID*, XV Jornadas de Paralelismo, Almería, Spain, 15-17 September 2004, pp. 295-300.
- [21] M. Hidalgo-Conde, A. Rodríguez, S. Ramírez and O. Trelles, *Adaptive Task Scheduling in Computational GRID Environments*, European GRID Conference (ECG 2005), Amsterdam, Holland, 14-16 February 2005.
- [22] J. Hormigo and G. Cristóbal, *Image Segmentation Using the Wigner-Ville Distribution*, Advances in Imaging and Electron Physics, Vol. 131, Ed. Elsevier, 2004, pp. 65-80.
- [23] J. Hormigo, J. Olivares, I. Benavides and J. Villalba, *Estimación de Movimiento en MPEG Mediante Técnicas de Aritmética On-Line Sobre FPGA*, 3ª Conferencia Iberoamericana en Sistemas, Cibernética e Informática (CISCI'2004), Orlando, Florida, USA, 21-25 July 2004, Vol II, pp. 213-218.
- [24] J. Hormigo, J. Olivares, J. Villalba and I. Benavides, *New On-Line Comparator with No On-Line Delay*, 8th World Muticonference on Systemics, Cybernetics and Informatics (SCI'2004), Orlando, Florida, USA, 18-21 July 2004, Vol II, pp. 45-49.
- [25] J. Hormigo, J.M. Prades, J. Villalba and E.L. Zapata, *Hardware Implementation of the Wavelet Transform for JPEG2000*, Conference on VLSI Circuits and Systems II (in SPIE Symposium on Microtechnologies for the New Millennium 2005) Seville, Spain, 9-11 May 2005, SPIE vol. 5837-24.
- [26] J. Hormigo, M. Sánchez, M. González, G. Bandera and J. Villalba, *Optimized FPGA Implementation of Trigonometric Functions with Large Input Arguments*, XIX Conference on Desing of Circuits and Integrated Systems (DCIS'2004), Burdeos, France, 24-26 November 2004.
- [27] J. Hormigo, J. Villalba and E. L. Zapata, *CORDIC Processor for Variable-Precision Interval Arithmetic*, Journal of VLSI Signal Processing, Vol. 37, No. 1, May 2004, pp. 21-39.
- [28] M.J. Martín-Vázquez, M.A. Trelles and O. Trelles, *Computer Assisted System for the Objective Assessment of Vein Elimination*, XII Congreso de la Sociedad Española de Láser Médico Quirúrgico, Marbella, Spain, 14-16 May 2004.
- [29] M.J. Martín-Vázquez, M.A. Trelles y O. Trelles, *Lt-skin: A Computer Assisted Platform for Objective Assessment of Tissue Quality*, Journal of Laser in Medical Science (in press)
- [30] A.J. Mesones, E. Herruzo, G. Bandera, J.I. Benavides and O. Plata, *Distributed Architecture System for Computer Performance Testing*, 6th Int'l. Conf. on Parallel Processing and Applied Mathematics (PPAM'2005), Poznan, Poland, 11-14 September 2005.
- [31] N. Moreno-Vergara, J. F. Aldana, P. A. de Alarcón and O. Trelles, *Distance Learning Platform for Bioinformatics*, 2nd Int'l. Conf. on Multimedia and Information & Communication Technologies in Education (m-ICTE 2003), Badajoz, Spain, 3-6 December 2003.
- [32] M.A. Navarro, F. Corbera, R. Asenjo, A. Tineo, O. Plata and E.L. Zapata, *A new Dependence Test based on Shape Analysis for Pointer-based Codes*, 17th Int'l Workshop on Languages and Compilers for Parallel Computing (LPCP'2004), West Lafayette, Indiana, USA, 22-25 September 2004.
- [33] M.A. Navarro, R.G. Valderrama, F. Corbera, E. Gutiérrez, R. Asenjo, O. Plata and E.L. Zapata, *Exploiting Locality and Parallelism in Pointer-based Programs*, 11th Workshop on Compilers for Parallel Computers (CPC'2004), Seeon Monastery, Chiemsee, Germany, 7-9 July 2004.
- [34] J. Olivares, J. Hormigo, I. Benavides and J. Villalba *Alternativas al Cálculo de la Estimación de Movimiento en MPEG*, XV Jornadas de Paralelismo, Almería, Spain, 15-17 September 2004, pp. 84-89.
- [35] J. Olivares, J. Hormigo, J. Villalba and I. Benavides, *Minimum Sum of Absolute Differences Implementation in a Single FPGA Device*, Int'l. Conference on Field Programmable Logic and Applications (FPL'2004), Lisbon, Portugal, 30 August - 1 September 2004, pp. 986-990.
- [36] E.M. Ortigosa, L.F. Romero and J.I. Ramos, *Parallel Scheduling of the PCG Method for Banded Matrices Arising from FDM/FEM*, Journal of Parallel and Distributed Computing, Vol. 63, Iss. 12, December 2003, pp. 1243-1256.

- [37] E.M. Ortigosa, S. Romero, L.F. Romero and E.L. Zapata, *Interfaz de Comunicaciones para Códigos Paralelos Híbridos*, XV Jornadas de Paralelismo, Almería, Spain, 15-17 September 2004, pp. 144-149.
- [38] A.J. Pérez, G. Thode and O. Trelles, *AnaGram: Protein Function Assignment*, Bioinformatics, Vol. 20, No. 2, January 2004, pp. 291-292.
- [39] O. Plata and R. Asenjo, *Exploitation of Locality and Parallelism in Pointer-based Programs* (Tutorial), Int'l. Conf. on Parallel Architectures and Compilation Techniques (PACT'04), Antibes Juan-les-Pins, France, 29 September - 3 October 2004.
- [40] O. Plata, R. Asenjo, E. Gutiérrez, F. Corbera, A. Navarro and E.L. Zapata, *On the Parallelization of Irregular and Dynamic Programs*, Journal of Parallel Computing, Vol. 31, Iss. 6, June 2005, pp. 544-562.
- [41] A. Rodríguez, J.M. Carazo and O. Trelles, *Mining Association Rules from Biological Databases*, Journal of the American Society for Information Science and Technology, Vol. 56, Iss. 5, March 2005, pp. 493-504.
- [42] A. Rodríguez, N. Guil, D.M. Shotton and O. Trelles, *Automatic Analysis of the Content of Cell Biological Videos and Database Organization of Their Metadata Descriptors*, IEEE Transactions on Multimedia, Vol. 6, No. 1, February 2004, pp. 119-128.
- [43] A. Rodríguez, N. Guil, D.M. Shotton and O. Trelles, *Analysis and Description of the Semantic Content of Cell Biological Videos*, Multimedia Tools and Applications, Vol. 25, No. 1, January 2005, pp. 37-58.
- [44] A. Rodríguez, A. Ortega, J.A. Micó and O. Trelles, *Automatic Tracking Analysis in Morris Water Maze Biomedical Videos*, 4th IASTED Int'l. Conf. on Visualization, Imaging and Image Processing, Marbella, Spain, 6-8 September 2004, pp. 647-652.
- [45] S. Romero, E. Gutiérrez, L.F. Romero, O. Plata and E.L. Zapata, *Parallelization Issues of a Code for Physically-based Simulation of Fabrics*, Computer Physics Communications, Vol. 162, Iss. 3, September 2004, pp. 188-202.
- [46] L.F. Romero, E.M. Ortigosa, S. Romero and E.L. Zapata, *Nesting OpenMP and MPI in the Conjugate Gradient Method for Band Systems*, 11th European PVM/MPI Users' Group Meeting (EuroPVM/MPI 2004), Budapest, Hungary, 19-22 September 2004. (Springer-Verlag, Germany, LNCS no. 3241, pp. 181-190).
- [47] L.F. Romero, M.A. Trelles and O. Trelles, *Real-time simulation for laser-tissue interaction model*, Int'l. Conference on Parallel Computing (ParCo2005), Málaga, Spain, 13-16 September 2005.
- [48] H. Rosales, J. Aldana, R. Asenjo and O. Trelles, *Adecuación de un Modelo de Enseñanza Superior en Modelo de Enseñanza Virtual*, 1er. Congreso Virtual Latinoamericano de Educación a Distancia (LatinEduca2004.com), 23 March - 4 April 2004.
- [49] E. Sáez, J.I. Benavides and N. Guil, *Reliable Real Time Scene Change Detection in MPEG Compressed Video*, IEEE Int'l. Conference in Multimedia and Expo (ICME 2004), Taipei, Taiwan, 27-30 June 2004, pp. 567-570.
- [50] E. Sáez, J.I. Benavides and N. Guil, *Combining Luminance and Edge Based Metrics for Robust Temporal Video Segmentation*, Int'l. Conference in Image Processing (ICIP 2004), Singapore, 24-27 October 2004, pp. 2231-2234.
- [51] A. Tineo, F. Corbera, A. Navarro, R. Asenjo and E.L. Zapata, *A Novel Approach for Detecting Heap-based Loop-carried Dependences*, IEEE Int'l. Conf. on Parallel Processing (ICPP'2005), Univ. of Oslo, Norway, 14-17 June 2005.
- [52] G.P. Trabado, O. Plata and E.L. Zapata, *A Survey on Grid Architectures for Bioinformatics*, American Scientific Pub. (in press)
- [53] M.A. Trelles, I. Allones, M.J. Martín Vázquez, O. Trelles, M. Vélez and S. Mordon, *Long Pulse Nd: Yag Laser for Treatment of Leg Veins in 40 Patients with Assessments at 6 and 12 Months*, Lasers in Surgery and Medicine, No. 35, 2004, pp. 68-76.
- [54] M.A. Trelles, I. Allones and O. Trelles, *An 810 nm Diode Laser in the Treatment of Small (≤ 1.0 mm) Leg Veins: A Preliminary Assessment*, Laser in Medical Science, Vol. 19, No. 1, 2004, pp. 21-26.
- [55] O. Trelles and A. Rodríguez, *Bioinformatics and Parallel Metaheuristics*, Parallel Metaheuristics, Wiley Book Series on Parallel and Distributed Computing, Ed. E. Alba (in press)
- [56] M.A. Trenas, F.J. Corbera, E. Gutiérrez, J.R. Cózar and S. Romero, *Experiencia Docente con un Servidor de Prácticas para la Asignatura Tecnología de Computadores*, XV Jornadas de Paralelismo, Almería, Spain, 15-17 September 2004, pp. 489-493.
- [57] M. Ujaldón and J. Saltz, *The GPU as an Indirection Engine for a Fast Information Retrieval*, Int'l. Journal of Electronic Business, Vol. 3, No. 4, July-August 2005.
- [58] M. Ujaldón and J. Saltz, *Exploiting Parallelism on Irregular Applications Using the GPU*, Int'l. Conference on Parallel Computing (ParCo2005), Málaga, Spain, 13-16 September 2005.
- [59] F.L. Valverde, N. Guil, E. Domínguez and J. Muñoz, *A Statistical Validation of Vessel Segmentation in Medical Images*, IX Ibero-American Conf. on Artificial Intelligence (IBERAMIA 2004), Puebla, Mexico, 22-26 November 2004. (Springer-Verlag, Germany, LNAI no. 3315, pp. 617-625).
- [60] F.L. Valverde, N. Guil and J. Muñoz, *Segmentation of Vessels from Mammograms Using a Deformable Model*, Computer Methods and Programs in Biomedicine, Vol. 73, Iss. 3, March 2004, pp. 233-247.
- [61] J. Veintimilla, V. Espigares, M.A. Trelles, O. Trilles and L.F. Romero, *Integrated Computational System for Laser-Tissue Interaction Modelling*, XII Congreso de la Sociedad Española de Láser Médico Quirúrgico, Marbella, Spain, 14-16 May 2004.
- [62] J. Villalba, J. Hormigo, J.M. Prades and E.L. Zapata, *On-line Multioperand Addition Based on On-line Full-Adders*, 16th IEEE Int'l. Conf. on Application Specific Systems, Architectures and Processors (ASAP'2005), Samos, Greece, 23-25 July 2005.

Integrating qualitative and quantitative reasoning spatio-temporal models and their application to robotics TIC2003-07182

M. Teresa Escrig

Universitat Jaume I

Engineering and Computer Science Department

Campus Riu Sec, Avda Sos Baynat s/n E-12071 Castellón (Spain)

Phone: +34 964 728303 fax: +34 964 728486

e-mail: escrigm@icc.uji.es

Abstract

Autonomous navigation of real robots includes four problems: mapping, locating, planning and motion control. The mapping problem is been currently solved by representing the environment of the robot by hybrid information: imprecise quantitative measures provided by sensors and the corresponding qualitative representation, mainly by cognitive maps. The rest of the problems in autonomous navigations is been currently solved using quantitative techniques, usually probabilistic, in order to overcome with the imprecision.

The idea of an autonomous robot navigating by using qualitative or common sense reasoning comes from the last 80th (i.e. the approaches developed by Kuipers et al.), although they have never been applied to real robots because they assume unreal perfect sensor information. In the 90th decade, all around Europe, a lot of theoretical qualitative models of different aspects of space and time were developed. Although the idea of applying them to robot navigation was very attractive, and some results were obtained in representing qualitatively the environment, no qualitative reasoning process has been used for solving any of the navigation problems on real robots.

In this project we want to use qualitative representation enhanced with quantitative imprecise data in a hybrid representation, and qualitative or common sense reasoning to solve the four problems of autonomous real robot navigation. Moreover, we will continue developing theoretical qualitative models of those aspects of space, time and movement that we need in order to solve particular problems. We also are going to deal with sensor integration: sonar, laser and vision based on biological models.

1 Objectives of the project

The integration of qualitative and quantitative spatio-temporal models for real robot navigation is a challenging research subject in Europe and United States. At the end of 80th in Kuipers et al.'s approaches, qualitative models were applied to simulated robot navigation, where the robot perception where perfect and unreal. Afterwards, a lot of theoretic qualitative spatio-temporal models have been developed. However, up to now, real robot navigation has been solved using quantitative techniques, usually probabilistic. The integration of theoretical spatio-temporal qualitative models and quantitative techniques for real robot navigation is a challenging problem.

From 1994, our research group (Cognition for Robotics Research, C4R2, www.c4r2.uji.es) has been working in the definition of a theoretical qualitative model which allows us the integration of several spatio-temporal aspects into the same model [Escrig & Toledo 98a]. Up to now we have integrated 2D orientation [Escrig & Toledo 96a], 3D orientation [Pacheco et al. 01a, 01b, 01c, 02a, 02b, 02c, 02d], cardinal directions [Escrig & Toledo 98c, 00b], topology [Museros & Escrig 00, 01a, b], named distances [Escrig & Toledo 96b, 97a, 97b, 98c, 00b], compared distances [Escrig & Toledo 01a, 01b], velocity [Escrig & Toledo 02a, 02b, 02c], topology and time [Museros & Escrig 02a, 02b, 02c, 02d], movement [Escrig et al. 02], for point and extended objects [Escrig & Toledo 96b]. This integration has been accomplished thanks to consider each aspect to be integrated as an instance of the constraint satisfaction problem [Escrig & Toledo 98d], solved using approaching techniques such as path consistency, by means of constraint logic programming extended with constraint handling rules as tool. This model of integration has been applied to simulated agent navigation by the structured environment of the streets of a city [Escrig & Toledo 98b, 00a, 00c]. Preliminary results in the use of qualitative spatial reasoning for real robot navigation has also been obtained.

The general objectives of the project can be divided into two parts. In the theoretical part of the project we will go on working in the integration of spatio-temporal aspects in the same model, if they were necessary for our practical applications. We are also going to work on sensor integration: infrared, sonar, laser and vision. A qualitative interpretation of sensor information will allow us a treatment of the robot environment and the robot obstacles, similar to human common sense. The advantage of the use of common sense is an enhancement of the level of the robot intelligence, which will allow us an efficient, quick and robust sensor use. We want to go into qualitative vision systems based on biological models.

In the practical part of the project, we will define robot navigation algorithms in which we will integrate the qualitative/quantitative spatio/temporal models which have been theoretically developed. We will work in three applications: the navigation in the structured environments of public buildings with two kind of robots (Khepera and Pioneer); the automatic construction of mosaic design by using qualitative shape recognition of ceramic pieces; and the navigation in an environment similar to brain structure with a legged robot.

The concrete objectives of this project are the following:

1. The demonstration of the equivalence among the qualitative spatio-temporal models based on intervals. Our goal is to obtain a general model, for which all spatio-temporal qualitative named models are instances.

2. An integration of the quantitative imprecise information provided by sensors into a qualitative, more abstract level. As the result of this, the robot will extract the main features of the environment in a quicker and more robust way.
3. We will deal with vision in a more cognitive way, by doing qualitatively shape representation, segmentation and matching processes.
4. We will develop navigation algorithms where the theoretical qualitative models of space and time are going to be incorporated. We are going to use four kind of robots: the Khepera, the Pioneer, the Aibo ERS7 and the Lauron IV.
5. We also want to develop our programs to be independent of the robotic platform. Therefore we are going to use the architecture based on multiagent systems and ontology defined in our group [Manzano & García 02].

In order to afford our objectives, we have defined the following tasks:

TASK 1. Demonstration of the equivalence among qualitative models based on intervals.

SUBTASK 1.1. Study of the state of the art of qualitative models based on intervals.

SUBTASK 1.2. Definition of a general spatio-temporal qualitative model based on intervals.

TASK 2. Qualitative/quantitative sensor integration of distance sensors: sonar, infrared and laser.

SUBTASK 2.1. Sonar, infrared and laser sensors provide distances among the robot and the obstacles. For the case of sonar and infrared, these distances contain a lot of imprecision. In this subtask we will represent distances qualitatively. For ring sensor robots, we will qualitatively integrate distances and orientation. The qualitative treatment of measurements manage the imprecision of data.

SUBTASK 2.2. The previous CICYT project was coordinated with the vision research group of our University. From this collaboration, the vision group provided a 3D reconstruction of any environment by a laser range finder. In this project we want to represent the results of this work in qualitative terms, i.e. describing walls and obstacles. We are interesting in obtaining landmark points of these features, i.e. corners of walls or obstacles.

TASK 3. Qualitative/quantitative sensor integration of sensors: vision.

SUBTASK 3.1. Study of the art on qualitative shape representation and cognitive vision.

SUBTASK 3.2. We will take from the state of the art the most satisfactory approach for our necessities or we will develop our own approach if there is no satisfactory one. We will extract the main features of the image by qualitative representation. We will provide a qualitative recognition of shapes.

SUBTASK 3.3. We will provide an automatic matching process. We will study the possibility of doing a qualitative segmentation of the image.

TASK 4. Development of robot navigation algorithms by using representation and reasoning qualitative spatio-temporal models.

SUBTASK 4.1. Study of the art on mobile real robot control architectures.

SUBTASK 4.2. Adaptation of the control architecture chosen in the previous subtask for the development of a navigation algorithm for the Khepera robot. Gradual incorporation of the representation and reasoning qualitative models of different aspects of space and time into this architecture and the sensor integration algorithm. We will build a structured environment adapted to the size and sensor reach by the Khepera robot. The robot will automatically learn the environment structure in a cognitive map. The robot will navigate towards an objective. The complete algorithm will be tested in a mobile robot simulator software, called Webot developed by Cyberbotics Ltd. (www.cyberbotics.com).

SUBTASK 4.3. We will build an obstacle detection module and another module for recovering from errors when the robot gets lost, for the Khepera robot, which will be added to the algorithms developed in the previous task.

SUBTASK 4.4. The algorithms developed in the previous tasks will be tested in the Khepera real robot.

SUBTASK 4.5. We will adapt the algorithms developed in the previous task for the Pioneer 3-DX (www.activmedia.com). We will test initially the algorithms with the Webots simulator of the Pioneer robot.

SUBTASK 4.6. We will build an obstacle detection module and another module for recovering from errors when the robot gets lost, for the Pioneer robot, which will be added to the algorithms developed in the previous task.

SUBTASK 4.7. The algorithms developed in the previous tasks will be tested in the Pioneer real robot.

SUBTASK 4.8. We will adapt the algorithms developed in the previous tasks for the legged robots: AIBO ERS7 and LAURON IV.

SUBTASK 4.9. Definition of general modules which includes all the functionalities previously solved which can be used by different robotic platforms.

We have planned to execute the previous defined tasks following the next timetable:

Tasks	First year	Second year	Third year
Subtask 1.1			
Subtask 1.2			
Subtask 2.1			
Subtask 2.2			
Subtask 3.1			
Subtask 3.2			
Subtask 3.3			
Subtask 4.1			
Subtask 4.2			
Subtask 4.3			
Subtask 4.4			
Subtask 4.5			
Subtask 4.6			
Subtask 4.7			
Subtask 4.8			
Subtask 4.9			

We are currently starting the second half of the second year of the project.

2 Level of success of the project

We are currently working on task 1 although we have not published any results yet.

Subtask 2.1 has been mostly finished [Falomir & Escrig 04a,b]. It has to be tested in different real robot platforms, which is our current work. The theoretical results are very promising.

Subtask 2.2 has been momentary oriented towards the qualitative interpretation of 2D laser sensor, because for the applications that we have in mind (autonomous real robot navigation) we do not need a precise 3D reconstruction of the environment. We have already obtained a qualitative description of landmarks of the environment (namely corners of walls and obstacles) at different levels of granularity which can automatically be adapted depending on the environment and the application, as well as the definition of qualitative places (where the main features of the environment are invariant) [Peris & Escrig 05].

Subtask 3.1 has been partially accomplished, we need to do a deeper study of the art on cognitive vision. Subtask 3.2 has been partially accomplished. We have completely accomplished the qualitative representation, recognition and matching of shapes of objects in two different applications: the 2D shape recognition of ceramic piece in order to build a mosaic given a vector design [Museros & Escrig 04a, 04b, 04c, 04d]; and the 3D shape recognition of robot obstacles [Museros & Escrig 04e]. The matching process to be developed in subtask 3.3 has been accomplished. The part which needs complementary work is the qualitative image segmentation. Quantitative preliminary results have been obtained in [Museros & Escrig 05]. We still have not provided any results on the integration of distance sensors (sonar, infrared and laser) with vision. However, we will afford this in brief.

Subtask 4.1 has partially been accomplished. In subtask 4.2 we have applied the topology extended with time model developed in [Museros & Escrig 02a, 02b, 02c, 02d] to describe the movement of the Khepera robot [Museros & Escrig 04e] which has been tested both with the Webots simulator and the real robot (at it was pointed out in subtask 4.4). We are interested in translating these results to the rest of the robots, because the sensor reach of the Khepera robot is too short for practical applications. We are going to abandon this platform in the future. Subtask 4.3. and 4.6 have not been solved yet. For the Pioneer and the AIBO robots we have developed other navigation algorithms [Escrig & Peris 05a, 05b], [Peris & Escrig 05]. In fact, we have applied qualitative spatio-temporal models for solving the almost-SLAM problem defined at the ROBOCUP competition. We are currently implementing the algorithms. Although conceptually is solved, we need to test the complete set of programs in the real platforms in order to obtain the statistical results of their behaviour. We think they will be completely finished at the end of the year for both platforms, AIBO and Pioneer. We also are working with the LAURON IV robot. For this task we need to implement a robust module to recover the robot from errors in locating, we need to test final integrated programs in real robots and to transfer programs from different robotic platforms to test the programs in all the platforms (subtask 4.7). For affording the subtask 4.9 we have defined a standard for working

in our group: each program developed by any member of the group will have to be documented in a particular way and stored in an ontology, in order to be reusable for all people and robotic platforms. We also have developed a cross-compiler for reusing directly the code programmed for the Pioneer simulated in Webots and the real robot. We also have built in Webots a simulator of the LAURON IV robot [Pacheco & Benito 05].

3 Results

Funded by this project there are 5 people doing research for their PhD. Lledó Museros is going to defend her thesis at the end of this year. Juan Carlos Peris and Julio Pacheco need at least two years more to finish their PhD. The other two people have done only preliminary work. The group is very active. During this year and a half of the live of this project we have published 12 papers: [Museros & Escrig 05], [Escrig & Peris 05a, 05b], [Peris & Escrig 05], [Falomir & Escrig, 04a, 04b], [Museros & Escrig 04a, 04b, 04c, 04d, 04e], [Pacheco & Benito 05]. We also have edited a journal [Yager et al. 2005]. And we also have published a book [Escrig 05].

We have done a very big effort during this year and a half in the application of the theoretical qualitative models developed by our group to real robot navigation. We have obtained less publications than in the previous CICYT project because it has been tremendously difficult to obtain the first practical results with real robots. In summary we have obtained 6 main results: (1) We have built an application which automatically interprets a vector design of a mosaic and builds it from a set of ceramic piece pieces which are qualitatively recognized in 2D; (2) We already have a robot (Khepera) which qualitatively recognize obstacles in 3D; (3) The Khepera robot already navigates by using a qualitative motion model which integrates topology and time; (4) We already have a robot (Pioneer) which qualitatively recognize landmarks from the quantitative environment information provided by a laser sensor; (5) We have qualitatively integrated the information provided by any kind of distance sensor; and (6) We also have conceptually solved the almost-SLAM problem defined for the AIBO robots at the ROBOCUP competition. We have almost finished its implementation. We are also transferring the results for solving the problem of the automatic sweeping of any surface.

We have started relationships with factories in order to transfer our technology. Among other things, we want to apply the results of our work to build autonomous vacuum cleaner robots.

Last year we participated very actively in the elaboration of a proposal for a Network of Excellence (called Choros) in the VI European Framework Programme (EFP) which unfortunately finally does not succeed. The groups involved in Choros were: U. of Leeds, U. of Zurich, U. of Hamburg, U. of Bamberg, U. of Muenster, Spanish Scientific Research Council, U. of Paris at Orsay, U. of Bremen, Örebro U., T.U. of Vienna, U. of Tuebingen, U. of Seville, U. of Geneva, U. Jaume I of Castellón, New Bulgarian U., and U. of Freifurg. The collaboration with the previous mentioned groups comes from the participation in the "Spacenet" Network of Excellence of the IV EFM for 4 years (1994-1998). We have

particularly good relationships with the Christian Freksa's group in Bremen. Christian is a co-supervisor of Museros's PhD thesis.

Our group is also a member of the MONET European network of excellence (funded by the 5th EFP) on Qualitative Reasoning.

We are also very active members of the ARCA Spanish network on Qualitative Reasoning and Applications (funded by a CICYT project).

We also collaborate with the U. Valencia (Dr. Salvador Moreno) and the U. Valladolid in Segovia (Dr. José Vicente Alvarez) in a very active way.

Another indicator of the relevance of our work and the effort done to communicate the result of our research work has been the concession of two prices: the price on "Science and Technology" of the city of Castellón to M. Teresa Escrig in 2004 [Escrig 05], and the "Castellón Onda Cero Radio" price to the Cognition for Robotics Research group for their research trajectory.

References

- [Escrig 05] M.T.Escrig, "Robots que Razonan. Aplicación de procesos de razonamiento al movimiento de robots simulados por las calles de Castellón y al movimiento de robots reales", Premio Ciudad de Castellón de Ciencia y Tecnología. Ayuntamiento de Castellón, Abril 2005.
- [Escrig & Peris 05a] Escrig, M.T., Peris, J.C., "The use of a reasoning process to solve the almost SLAM problem at the Robocup legged league", accepted to the Catalanian Conference on Artificial Intelligence, CCIA'05, October 2005 (IOS-Press).
- [Escrig & Peris 05b] Escrig, M.T., Peris, J.C., "The use of Qualitative Spatial Reasoning to solve the Simultaneous Localization and Map Building problem in non-structured environments", VII Jornadas de Trabajo ARCA, JARCA'05.
- [Escrig & Toledo 96a] M.T.Escrig, F.Toledo, "Qualitative Spatial Orientation with Constraint Handling Rules", Proceedings of the 12th European Conference on Artificial Intelligence ISBN: 0-471-96809-9, Budapest, Hungary, pp. 486-490, 1996.
- [Escrig & Toledo 96b] M.T.Escrig, F.Toledo, "Enhancing Qualitative Relative Orientation with Qualitative Distance for Robot Path Planning",): Proceedings of the 8th IEEE International Conference on Tools for Artificial Intelligence, ISBN: 0-8186-7686-8, pp. 174-182, 1996.
- [Escrig & Toledo 96b] M.T.Escrig, F.Toledo, "The Treatment of Extended Objects in Qualitative Spatial Reasoning", Proceedings of the Journees Hispano-francaises systemes intelligents et controle avance, ed. UPC-LASS/CNRS, pp. MQ4/1-MQ4/10, 1996.
- [Escrig & Toledo 97a] M.T.Escrig, F.Toledo, "Managing Orientation and Distance on Extended Objects in Qualitative Spatial Reasoning", Proceedings of the International Joint Conference on Artificial Intelligence, Workshop on Spatial and Temporal Reasoning, pp. 5-12. Nagoya, Japan, August 1997.

- [Escrig & Toledo 97b] M.T.Escrig, F.Toledo “The integration of Qualitative Orientation and Named Distances: Application to a Qualitative Navigation Simulator”. *International Journal on Artificial Intelligence Tools*, Vol.6, No.4, pp.451—480, 1997.
- [Escrig & Toledo 98a] Escrig, M.T., Toledo, F., *Qualitative Spatial Reasoning: theory and practice. Application to Robot Navigation*. IOS Press, *Frontiers in Artificial Intelligence and Applications*, ISBN 90 5199 4125. 1998.
- [Escrig & Toledo 98b] M.T.Escrig, F.Toledo, “Applying Qualitative Spatial Reasoning to Robot Navigation”. *Butlletí de l’Associació Catalana d’Inteligència Artificial*, Vols.14-15, pp. 204-212, 1998.
- [Escrig & Toledo 98c] M.T.Escrig, F.Toledo, "Cardinal Directions on Extended Objects for Qualitative Navigation", *Lecture Notes in Artificial Intelligence* Vol. 1415. Springer-Verlag. ISBN 3-540-64582-9, pp.291-301, 1998.
- [Escrig & Toledo 98c] M.T.Escrig, F.Toledo, “A Framework Based on CLP Extended with CHRs for Reasoning with Qualitative Orientation and Positional Information”, *Journal of Visual Languages and Computing* Vol. 9, pp. 81-101, 1998.
- [Escrig & Toledo 00a] Escrig, M.T., Toledo, F., “Autonomous Robot Navigation using human spatial concepts”. *International Journal of Intelligent Systems*, Vol.15, No.3, March 2000, pp.165-196, 2000. ISBN 0884-8173.
- [Escrig & Toledo 00b] Escrig, M.T., Toledo, F., “The use of qualitative orientation, distance and cardinal directions in autonomous robot navigation”, *Current Trends in Qualitative Reasoning and applications*, edition. J.A.Ortega, R.M.Gasca, and M.Toro editors, pp. 87-97, 2000.
- [Escrig & Toledo 00c] Escrig, M.T., Toledo, F., “Autonomous Robot Navigation using human spatial concepts”, *International Journal of Intelligent Systems* ISBN 0884-8173, pp.165-196, 2000.
- [Escrig & Toledo 01a] Escrig, M.T., Toledo, F., “Reasoning with compared distances at different levels of granularity”, in *Proceedings of the 9th Conference of Spanish Association of Artificial Intelligence*, October 2001.
- [Escrig & Toledo 01b] M.T. Escrig, F. Toledo, “Applying Compared Distances to Robot Navigation”, *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI’01)*, Workshop on Spatial and Temporal Reasoning with Agents Focus, Seattle, August 2001.
- [Escrig & Toledo 02a] Escrig M.T., Toledo, F., « Qualitative Velocity», *Lecture Notes in Artificial Intelligence* num. 2504 *Topics in Artificial Intelligence*. 5th Catalanian Conference on AI, CCIA 2002. Eds. M. Teresa Escrig, Francisco Toledo and Elisabet Golobardes. ISBN : 3-540-00011-9, Ed. Springer, pp. 29-39, 2002.
- [Escrig & Toledo 02b] Escrig M.T., Toledo, F., « Qualitative Velocity: How to represent it?», *Sistemas Cualitativos y Diagnosis. Automatización del Razonamiento Cualitativo y Aplicaciones*. Eds. Juan Antonio Ortega, Xavier Parra y Belarmino Pulido. Edición digital @tres S.L.L. ISBN : 84-95499-62-2. pp. 59-62, 2002.

- [Escrig & Toledo 02c] Escrig M.T., Toledo, F., « Qualitative Velocity », Proceedings of Workshop at the European Conference on Artificial Intelligence 2002, pp. 77-80, 2002.
- [Escrig et al. 02] Escrig, M.T., Museros M., Pacheco, J., Toledo, F., "Several Models on Qualitative Motion as instances of the CSP", *Inteligencia Artificial* Número 17 - Otoño /2002 N°17, Revista Ibero-americana de Inteligencia Artificial VI/02, pag. 55-71, 2002. ISSN: 1137-3601 (c).
- [Falomir & Escrig, 04a] Falomir Z., Escrig M. T., "A hybrid model for multi-sensor data fusion", Spatial and Temporal Reasoning Workshop, European Conference on Artificial Intelligence (ECAI), pp. 3-13, August 2004.
- [Falomir & Escrig, 04b] Falomir Z., Escrig M. T., "Qualitative multi-sensor data fusion", Seventh Catalanian Conference on Artificial Intelligence, IOS Press, *Frontiers in Artificial Intelligence and Applications*. Vol. 113, pp. 259-266, ISBN 1-58603-466-9, October 2004.
- [Manzano & García 02] F. Manzano , L.A. Garcia, "OCA: An Open, Modular, Ontology Based Autonomus Robotic Agent Architecture", Ed. Springer-Verlag. *Lecture Notes in Artificial Intelligence*. Vol 2443, pp. 173ff. Varna, Bulgaria, September 4-6, 2002.
- [Museros and Escrig 05] Museros L., Escrig M. T., "Extracting Relevant Features of an Image for Qualitative Shape Matching.", VII Jornadas de Trabajo ARCA, JARCA'05.
- [Museros and Escrig 04e] Museros L., Escrig M. T., "A Qualitative Theory for Shape Matching applied to Autonomous Robot Navigation.", *Frontiers in Artificial Intelligence and Applications*. Recent Advances in Artificial Intelligence Research and Development. ISBN 1 58603 466 9, 2004.
- [Museros and Escrig 04d] Museros L., Escrig M. T., "Qualitative Recognition of objects without holes, with holes and curves", Workshop Notes del en Spatial and Temporal Reasoning Workshop, en el ECAI 2004, 16th European Conference on Artificial Intelligence 2004, 2004.
- [Museros and Escrig 04c] Museros L., Escrig M. T., "A Qualitative Theory for Shape Representation and Matching", *Proceedings QR 2004*, Edited by Johan de Kleer and Keneth D. Forbus., pgs. 3-10, 2004.
- [Museros and Escrig 04b] Museros L., Escrig M. T., "A Qualitative Theory for Shape Representation and Matching for Design", *Proceedings ECAI 2004*, IOS Press ISSN 0922-6389, pgs. 858-862, 2004.
- [Museros and Escrig 04a] Museros L. , Escrig M.T., "A Qualitative Theory for Shape Description", *Inteligencia Artificial*. Revista IberoAmericana de Inteligencia Artificial num. 23, Volumen 8, 2004. ISSN: 1137-3601, 2004
- [Museros & Escrig 00] Museros L., Escrig M.T., « A Topological Calculus for Cartographic Entities», *Lecture Notes in Artificial Intligence* num. 1849. *Spatial Cognition II. Integrating Abstract Theories, empirical studies, formal methods, and practical applications*, C. Freksa, W. Brauer, C. Habel, K. Wender Eds.. ISBN 3-540-67584-I, Ed Springer, pp225-238, 2000.
- [Museros & Escrig 01a] Ll. Museros, M.T. Escrig, "The integration of topological information, qualitative orientation and positional information using constraint logic programming for robot navigation", *Proceedings of the International Joint Conference on Artificial Intelligence*

(IJCAI'01), Workshop on Spatial and Temporal Reasoning with Agents Focus, Seattle, August 2001.

[Museros & Escrig 01b] Ll. Museros, M.T. Escrig, "Enhancing Topological Information with other Qualitative Spatial Aspects", *Diagnosis, Razonamiento Cualitativo y Sistemas Socioeconómicos*. Carlos Alonso y Juan Antonio Ortega Eds. ISBN : 84- 95499- 35- 5, pp. 113-123, 2001.

[Museros & Escrig 02a] Museros L., Escrig M.T., « Modeling Motion Qualitatively : Integrating Space and Time», *Lecture Notes in Artificial Intelligence* num. 2504 *Topics in Artificial Intelligence*. 5th Catalanian Conference on AI, CCIA 2002. Eds. M. Teresa Escrig, Francisco Toledo and Elisabet Golobardes. ISBN : 3-540-00011-9, Ed. Springer, pp. 64-74, 2002.

[Museros & Escrig 02b] Museros L., Escrig M.T., « Applying qualitative spatial reasoning combining space and time to mobile robot navigation», *Sistemas Cualitativos y Diagnosis. Automatización del Razonamiento Cualitativo y Aplicaciones*. Eds. Juan Antonio Ortega, Xavier Parra y Belarmino Pulido. Edición digital @tres S.L.L. ISBN : 84-95499-62-2, pp. 63-72, 2002.

[Museros & Escrig 02c] Museros L., Escrig M.T., "Combining Qualitative Spatial Information: The integration of Topology, Time, Orientation and Distances for Spatial Reasoning", *Proceedings of Workshop at the European Conference on Artificial Intelligence 2002*, pp. 3-12, 2002.

[Museros & Escrig 02d] Museros L., Escrig M.T., « Integrating Qualitatively Time and Topology for Spatial Reasoning », *Proceedings of the QR2002. 16th International WorkShop on Qualitative Reasoning*. Editors : Nuria Agell and Juan A. Ortega. ISBN : 84-95499-60-6, pp. 105-112, Digital@tres, S.L.L., 2002.

[Pacheco & Benito 05] Pacheco, J., Benito, F. "Development of a Webots simulator for the LAURON IV Robot", accepted to the Catalanian Conference on Artificial Intelligence, CCIA'05, October 2005 (IOS-Press).

[Pacheco et al. 01a] Pacheco, J., Escrig, M.T., Toledo, F., "A model for representing and reasoning with 3-D Qualitative Orientation", in *Proceedings of the 9th Conference of Spanish Association of Artificial Intelligence*, October 2001.

[Pacheco et al. 01b] Pacheco, J., Escrig, M.T., Toledo, F., "Representing and Reasoning on Three-Dimensional Qualitative Orientation Point Objects", *Progress in Artificial Intelligence. Lecture Notes in Artificial Intelligence* vol 2258, Springer, pag. 298-305, 2001. ISBN: 3-540-43030-X.

[Pacheco et al. 01c] Pacheco, J., Escrig, M.T., Toledo, F., "Three-Dimensional Qualitative Orientation Point Objects: Model and Reasoning", *Workshop on Logic programming for artificial Intelligence and Information System (LPAI) EPIA 2001*, Porto Portugal. *Proceedings* edited by José Alferes and Salvador Abreu, pag. 59-74, 2001.

[Pacheco et al. 02a] Pacheco, J., Escrig, M.T., Toledo, F., "Integrating 3D Orientation Models", M. Teresa Escrig, Francisco Toledo, Elisabet Golobardes (Eds.). *Topics in Artificial*

Intelligence. Lecture Notes in Artificial Intelligence vol 2504, Springer, pag. 88-100, ISBN: 3-540-00011-9 ISSN 0302-9743, 2002.

[Pacheco et al. 02b] Pacheco, J., Escrig, M.T., Toledo, F., “The First Steps towards Reasoning on 3-D Qualitative Orientation”, in *Inteligencia Artificial, Revista Iberoamericana de inteligencia Artificial*. No 15 pag. 39-48 (2002) ISSN: 1137-3601.

[Pacheco et al. 02c] Pacheco, J., Escrig, M.T., Toledo, F., “Qualitative Spatial Reasoning on Three-Dimensional Orientation Point Objects”, *Proceedings of the QR2002. 16th International Workshop on Qualitative Reasoning*. Editors : Nuria Agell and Juan A. Ortega. ISBN : 84-95499-60-6, pp. 113-124, [Digital@tres](http://Digital@tres.com), S.L.L., 2002.

[Pacheco et al. 02d] Pacheco, J., Escrig, M.T., Toledo, F., “Three dimensional coarse qualitative spatial models”, *IV Jornadas de ARCA Sistema cualitativos y Diagnosis*, Junio 2002, Vilanova i la Geltrú, Barcelona, España. ISBN:84-95499-62-2, pp. 81-90, 2002.

[Peris & Escrig 05] Peris, J.C, Escrig, M.T., “Cognitive Maps for mobile Robot Navigation: A Hybrid Representation Using Referente Systems”, *19th International Workshop on Qualitative Reasoning*, Graz, Austria, pp. 179-185, ISBN 3-9502019-0-4, 18-20 May 2005.

[Yager et al. 05] R. R. Yager, I. Aguiló, M.T. Escrig (eds). *International Journal of Intelligent Systems. Special Issue on Robotics and Computer Vision*. Vol 20(5), No. 5, pp. 475-590, ISBN 0884-8173, May 2005.

Improving the performance and services offered by PC clusters. Development of distributed multimedia applications

TIC2003-08154-C06

José Duato *	Francisco Quiles †
Universidad Politécnica de Valencia	Universidad de Castilla-La Mancha
José Manuel García‡	Juan Manuel Orduña§
Universidad de Murcia	Universitat de Valencia
Francisco Mora¶	José Manuel Claver
Universidad Politécnica de Valencia	Universitat Jaume I

Abstract

PC Clusters have become the most attractive platform to implement supercomputers and high-performance servers. As a consequence, it is necessary to improve these platforms in order to provide higher performance and scalability, and better services, such as availability, reliability, quality of service, low power, etc.

In this project, we propose to improve most of the hardware components of a PC cluster. We will also improve the cluster system software. In particular, we propose to improve processor performance by using superpipelining techniques and by reducing power consumption. I/O system performance will be improved by using a switch fabric with point-to-point links. SMP scalability will be improved with a new memory coherence scheme. With respect to the interconnection network, we propose improving not only its performance (by using adaptive routing and load balancing techniques) but also its scalability (by using congestion management techniques), availability (by using fault-tolerance techniques) and quality of service (by including some hardware/software support in the routers and network interface controllers). We will also develop scheduling techniques that will take into account communication costs. Regarding Internet access, we will enhance IP routers, improving their scalability and bandwidth by means of new congestion control mechanisms and HOL blocking avoidance techniques. We will also research on wireless networks, developing new robust coding schemes based on new UEP techniques and applying them to the transmission of MPEG-4 video. Additionally, we will research on distributed virtual environments, developing new

*Email: jduato@disca.upv.es

†Email: paco@info-ab.uclm.es

‡Email: jmgarcia@ditec.um.es

§Email: Juan.Orduna@uv.es

¶Email: fjmora@eln.upv.es

||Email: claver@inf.uji.es

partitioning methods capable of improving the efficiency and scalability of these systems. We will also develop some distributed applications focusing on those that require some specific hardware to allow real-time execution. In particular, we will develop quantization and coding procedures and their hardware implementation on FPGAs. Finally, we will improve the synthesis of FPGA circuits from high-level specifications.

Keywords: PC clusters, interconnection networks (adaptive routing, congestion management, fault tolerance, network reconfiguration, QoS support, power consumption), wireless networks, shared-memory multiprocessors, circuit synthesis, video transmission, load scheduling, distributed virtual environments.

1 Project objectives

This research project has two primary goals: Improving the performance and services offered by PC clusters, and developing distributed multimedia applications. Each one of these objectives can be split into several subobjectives, organized into subprojects:

- Objective 1.1. Improving the performance and scalability of the interconnection network.
- Objective 1.2. Improving the performance of current interconnection technologies for system area networks (SANs).
- Objective 1.3. Improving the performance of Internet routers.
- Objective 2.1. Improving the quality of service and availability provided by the interconnection network to the applications.
- Objective 2.2. Design of switch architectures optimized for the execution of applications that broadcast multimedia traffic.
- Objective 2.3. Improving the performance of the transmission of MPEG-4 encoded video through wireless local area networks and development of new error-resilient encoding techniques.
- Objective 2.4. Analysis and dynamic parameter tuning for the medium access control layer in the IEEE 802.11 a/b WLAN.
- Objective 3.1. Improving performance and developing power consumption reduction techniques for the cluster nodes.
- Objective 3.2. Improving the services offered by PC clusters.
- Objective 3.3. Efficient development of applications for high-performance platforms.
- Objective 4.1. Improving the scheduling techniques for parallel systems.
- Objective 4.2. Improving the performance and scalability of distributed virtual systems.
- Objective 5.1. Development of a quantizer and real-time video encoder amenable for implementation in a reconfigurable VLSI architecture.
- Objective 6.1. Automatic FPGA programming from high-level specifications.

The time chronogram has been omitted since it occupies 7 pages.

2 Project results

2.1 Research results

This section summarizes the activities developed from the beginning of the project, grouped by objectives and topics within each objective. For each topic we highlight the most relevant scientific and technological results that have been obtained.

2.1.1 Objective 1.1.

A) Reduction of HOL-blocking (congestion control). We have designed a new technique, named RECN (Regional Explicit Congestion Notification), that completely eliminates the HOL blocking produced by congestion trees. When congestion is detected, queues are dynamically allocated to store the packets flowing through the congested point, thus isolating congested traffic and avoiding HOL blocking. This mechanism is able to handle congestion trees that grow from root to leaves and from leaves to root, overlapping congestion trees, and tree merging. RECN only requires up to 8 additional queues per port, independently of network size. RECN is very fast and, to the best of our knowledge, the only truly scalable congestion management technique proposed to date. RECN has been patented in the USA jointly with Xyratex. We also developed another technique to reduce HOL blocking, named DBBM (Destination Based Buffer Management), which is very simple yet effective. It is based on assigning groups of packet destinations to each of the queues, so that each queue can only produce HOL blocking among the packets it contains. The main benefit of this technique is its low cost. However, it does not completely eliminate HOL blocking.

B) Development of adaptive routing techniques. Solutions to out-of-order packet delivery. We have continued the work aimed at improving SAN performance by using adaptive routing strategies. The proposed strategies have been evaluated under different scenarios. One of the main drawbacks of adaptive routing is that it can not guarantee in-order packet delivery, feature that is required by many applications. During the last year, we have addressed the out-of-order packet delivery problem. To this end, we have analyzed, under different scenarios, the number of packets arriving out-of-order at destination, also evaluating the number of resources required for guaranteeing in-order packet delivery. Additionally, we have proposed two effective approaches to tackle the problem. One approach is based on allocating a very limited number of reordering buffers at the destination nodes together with certain end-to-end flow control mechanism. Another approach is based on limiting the packet injection at the source nodes, reordering packets at destination if needed. We have shown that it is possible to guarantee in-order packet delivery when applying adaptive routing by using a small number of additional resources and without significantly penalizing network performance.

C) Balancing techniques based on traffic distribution. We have proposed a cost-effective evaluation methodology for traffic balancing algorithms that is based on the use of genetic algorithms. This methodology obtains a near optimal routing path distribution from the applied routing algorithm, allowing us to assess the effectiveness of the traffic balancing algorithms commonly used in PC clusters and SANs.

D) Fault-tolerant routing for direct networks based on the use of intermediate nodes. We have developed a new methodology that, in case of failure(s), selects an intermediate node and routes every packet

traveling through the faulty region in two phases: from the source to the intermediate node and from there to the destination. This methodology is able to tolerate a relatively high number of failures without degrading performance in the absence of failures, and degrading it minimally when failures occur in the interconnection network. In addition, these features are achieved, unlike for previous proposals, without disabling healthy nodes, without using too many hardware resources, and introducing no penalty for most packets (for example, an increase in latency) when they are routed in a network with failures.

E) New strategy for interval routing. This strategy has been developed for interconnection networks built from commercial switches and has been optimized for the more widely used network topologies (tori and meshes). By optimized we mean that it is scalable from several points of view: (i) the memory space required to store the routing tables does not linearly depend on network size, (ii) the hardware required to apply the routing algorithm is not complex and does not depend on network size, and (iii) the time required to perform a routing operation does not depend on network size.

F) Power consumption reduction in interconnection networks. We have investigated two different schemes: The first proposal adjusts the link width based on the current load in the network. In this case, no link becomes completely disconnected. Therefore, the same routing algorithm can be used independently of whether the power saving mechanism is active or not, which simplifies the router design. This strategy is able to achieve very large reductions in power consumption. The main disadvantage is the increase in the latency when the network is lightly loaded. The second proposal reduces power consumption in interconnection networks based on the use of a regular topology and high-degree switches with channels consisting of multiple serial links running in parallel. This approach is based on the selective connection and disconnection of the high-speed serial links that compose the channels, depending on the current amount of traffic in the network. The behavior of the mechanism depends on the election of two thresholds. The design restrictions regarding the election of the thresholds and the relationship between these parameters and the behavior of the network in terms of latency and power consumption have been carefully analyzed. With a careful election of the thresholds, the system can be tuned to obtain different levels of power saving with a moderate increase in latency.

2.1.2 Objective 1.2.

A) Strategies to compute routing tables for SANs. We have proposed an effective methodology to compute up*/down* routing tables in SANs. This methodology shows that just assigning directions "up" or "down" to the links in an effective way is able to significantly increase the performance achieved by up*/down* routing (throughput increases up to a factor of 2.8).

B) Deadlock-free minimal routing techniques for SANs. We have proposed a new generic routing algorithm for SANs that is able to provide minimal routing by using a number of resources smaller than that required by other approaches. The new routing scheme, developed in collaboration with the Simula Research Laboratory in Oslo, combines the idea of classifying the set of minimal paths and assigning each class to a different virtual network, with the idea of enabling transitions between virtual networks.

C) Fault-tolerant routing techniques for InfiniBand. We have proposed an effective and scalable fault-tolerant routing methodology for InfiniBand under dynamic fault model, which takes advantage of the

automatic path migration (APM) mechanism provided by this network technology. The methodology is able to provide the maximum number of disjoint paths allowed by the topology. Deadlock-free disjoint paths are computed by allowing transitions between virtual channels in an ordered way. Recent results focused on providing fault tolerance by combining routing through intermediate nodes and partial network reconfiguration techniques.

D) Congestion reduction techniques for SANs. We have proposed an effective congestion management strategy for InfiniBand networks based on marking and validation of in-transit packets. This strategy allow us to thoroughly identify the traffic flows giving rise to the congestion and to measure the congestion degree in the network in order to apply effective corrective actions.

2.1.3 Objective 1.3.

See Section 2.2.

2.1.4 Objective 2.1.

A) Switch-fabric design of QoS-aware IP routers. See Section 2.2

B) QoS support in InfiniBand. We have evaluated the performance of InfiniBand via simulation under realistic traffic loads and by changing various system parameters. Specifically, we have performed a study on the buffer space required to meet the QoS needs as specified at connection time by the applications. From our results, we have found out that a buffer space of one can fulfill the application requirements. We have also analyzed the impact of the packet size over the QoS as perceived by the end application. In particular, we have found out that the use of small packet sizes implies a higher level of overhead.

Moreover, we have also developed a complete theory to verify the correctness of our proposal. Our theory has been developed from our previous work on a QoS-aware traffic classification. We propose to configure the InfiniBand mechanisms to guarantee the QoS requirements of a wide variety of applications. Currently, this proposal is being implemented into a prototype making use of a commercial InfiniBand system.

C) Improving the availability of the interconnection network. Under this activity, our main results are the following. First, we designed and evaluated a partially implicit routing algorithm (PIRa), that quickly (25% of the time required by previous alternatives) obtains a valid set of routes. It is based on the utilization of default ports in the upward segment of every up*/down* route.

Next, we designed a hybrid mechanism that combines PIRa with other traditional FERa (fully explicit routing algorithms) to mix their respective advantages, speed of convergence and network performance.

Finally, we defined two deadlock-free dynamic routing table distribution processes based on the deactivation of break ports and break dependencies, respectively. Both mechanisms outstandingly improve the results provided by previous proposals. In particular, the behavior of the second one is quite similar to the one expected for a hypothetical ideal dynamic distribution process.

D) QoS support for source routing SANs. We have proposed several methods to use these mechanisms to provide QoS based on bandwidth and latency requirements. We have evaluated the behavior of our proposals using CBR and VBR traffic. We have compared the VC arbitration table scheduler with our WFQ Credit Aware implementation of the MinBW, using CBR and VBR traffic. The results obtained have shown that both methods are able to provide a differentiated treatment to the different service classes. Results have also shown that the VC arbitration table is the only specified mechanism that can be used to provide both bandwidth and latency guarantees. However, the WFQ Credit Aware method obtains very good results for both bandwidth and latency, but it is unable to provide guarantees. The results obtained have shown that the injection flows must be controlled in order not to exceed the reserved bandwidth limits. We have found out that even if the flows are not controlled, the bandwidth requirements still can be guaranteed, but not the latency requirements.

2.1.5 Objective 2.2.

The main objective of this activity has been the design of a Multimedia Router (MMR) architecture: a single-chip router to be used in local area networks and cluster environments. The architecture has been designed to efficiently handle CBR and VBR multimedia traffic as well as best-effort traffic. An in-depth performance evaluation of various proposed link and switch scheduling algorithms has been carried out via simulation. As part of this effort, we have developed a simulation tool.

During this year, we have been enhancing the proposed architecture. In particular, we have proposed a novel link scheduling algorithm: Temperature-IABP (TIABP). Our results show that the use of this algorithm can significantly reduce the hardware complexity of the switch. An FPGA implementation of the proposed architecture clearly shows that it is feasible. Our results also show the effectiveness of the proposed architecture in meeting the QoS requirements of the various applications.

2.1.6 Objective 2.3.

Two main results have come out of this research activity: 1) the optimization of robust coding techniques UEP to be used by DCT-based coders, such as, MPEG-2 and MPEG-4, and 2) an analysis of the impact of the handover procedures over the video quality services. Throughout the first phase of the project, we have focused our efforts on the MPEG-4 video coding standard. Due to new developments in the area of video coding, we have recently undertaken the design of fast transcoding techniques. In particular, we are designing fast prediction techniques to be incorporated into the MPEG-2/H.264 transcoding process. Our choice has been based bearing in mind that the H.264 standard will become the most widely used video coding standard in the near future. It is also well known that the MPEG-2 standard has been widely used, i.e., there exists a large quantity of MPEG-2 encoded video material. Furthermore, the results of this research task should enable the design of transcoding techniques that in turn will enable the deployment of adaptive video services over WLANs.

2.1.7 Objective 2.4.

Under this activity, we have first carried out an in-depth analysis of the most up-to date QoS mechanisms for IEEE 802.11 WLANs. We have then developed various simulators and carried out an extensive set of simulations. Throughout this study, we have considered a WLAN cell working in the infrastructure mode supporting MPEG-4-based video applications. The various QoS mechanisms under study are enabled by modifying the system parameters: IFS, CWmin, CWmax. We have also

undertaken the study of the EDCA (802.11e) mechanism. EDCA is called to become the IEEE standard for QoS provisioning in IEEE 802.11 WLANs. Recently, we have designed a QoS framework fully compatible with EDCA. This mechanism is fully compatible with the DCF mechanism of the IEEE 802.11 standard. This feature ensures the interoperability of DCF-based stations with the EDCA mechanism: a key feature given the large number of DCF-based stations already deployed.

2.1.8 Objective 3.1.

A) Performance improvement and power consumption reduction for processors. We have performed an analysis of how superpipelining affects the energy consumption in high-performance processors. We evaluated pipeline depths from 5 to 42 stages, seeing how power consumption grows, mainly due to the increasing speculation. We made a proposal that reduces the power consumption by 16% for 42 pipeline stages.

Also, we have looked at embedded processors power consumption. We have proposed an instruction fetch unit that makes a more efficient use of the instruction cache. This mechanism, called “Fetch Mask Predictor”, identifies which instructions of each line are actually going to be used and performs a partial access. This way, we obtain a reduction of 44% in the power consumption of the instruction cache, contributing to improve the reliability, stability and lifetime of processors and batteries.

Finally, we have looked at power consumption for high-performance processors. In this case, peak currents and dissipated power can cause many faults and even damage the processor. We have generalized the previous proposal to apply it to high-performance processors. Again, doing partial accesses to instruction cache lines allows a significant power reduction without compromising performance.

B) Design of a new generation of Journaling File systems. We have added extensions to DualFS and studied its performance with large files. Using extensions has allowed us to reduce the space used by metadata, specially for large files.

To improve the performance of directory operations we have implemented them using B+ trees. To maintain the transactionality when using B+ trees, we have designed a mechanism called “metafiles” that will allow us to simplify DualFS implementation and add new features in the following years.

Also, we have tuned our current implementation determining the best default values for a number of parameters.

C) Scalable shared-memory multiprocessors. We have adapted our simulation tools to simulate high-performance SMP architectures, cc-NUMA architectures and CMP architectures. We are also studying how to connect the SIMICS simulator to our current simulation tools to be able to simulate commercial workloads.

Also, we are currently developing a scalable directory-based memory architecture and comparing it to high-performance SMP systems.

2.1.9 Objective 3.2.

A) User-transparent resource management for networks of workstations. We plan to develop an efficient resource management system, develop low-overhead communication protocols and develop fault-tolerance support.

We have already developed an environment for running applications based on MPI using novel interconnection network features and a new methodology for cluster software development that enables

better stability and predictability. Our methodology will ease the development of debugging schemes and fault-tolerance features.

B) Development of user management techniques for wireless environments. We are currently analyzing available tools for access control, designing techniques for access control management in structured environments and designing techniques for building ad-hoc networks.

We have designed a first prototype which brings a solution for the integration of authentication and authorization systems in a concrete context of access control, using SAML and XACML standards for information representation.

C) Integration of a component model (CORBA-LC) to the Grid (Globus). We are studying the non-functional requirements for Grid environments, developing a requirements document for component containers specifying the services offered, and implementing and testing the CORBA-LC component container model.

The work on this activity has focused on implementing a code generator for components, adapting the generator to generate Globus Web Services skeletons that offer the services provided by CORBA-LC components and implementing a proxy generator that allows wrapping Globus Web Services as CORBA-LC components.

We have been working on integrating the CORBA-LC component model and the Globus Grid, specially studying the meta-information facilities provided by Globus.

We have developed mechanisms for integrating version 2 and version 3 of the Globus platform, which are incompatible. These mechanisms ease the migration between versions.

Finally, we have proposed a policy-based management architecture (PBM) which aims at providing a flexible auto-configuration method for Grid nodes.

2.1.10 Objective 3.3.

A) Optimization of the compression and decompression processes for medical video using the 3D wavelet transform. We have studied and applied multithreading technology to video compression using the 3D wavelet transform. We have studied the behavior of general-purpose processors with SMT capabilities (Intel Pentium IV Xeon with Hyper-Threading) running the video encoder. We have proposed several parallelization techniques using different work decompositions. We have obtained speedups as high as 1.34 using Hyper-Threading and a functional decomposition. Additionally, we have implemented this parallelization scheme using both OpenMP and Pthreads to compare both implementations with respect to performance, ease of implementation, and maintainability of the resulting programs.

We have performed several studies comparing manual vectorization using SIMD instructions and automatic vectorization using modern compilers. We have found that automatic vectorization is not good enough yet. After studying the different SIMD instruction sets (Intel's MMX, SSE and SSE2; AMD's 3D-NOW; PowerPC's AltiVec; Sun's VIS; HP's MAX; and MIPS's MDMX) we have concluded that the available SIMD instruction sets allow the execution of the 3D wavelet transform in real time and no new instructions need to be proposed to achieve that goal.

B) Design, implementation and acceleration of image processing algorithms for artificial vision. Our work has focused on developing the basic infrastructure for the implementation and optimization of algorithms for image processing and artificial vision using conventional processors with multimedia and vector processing extensions. In particular, we have reached a fairly advanced development state for a

general purpose library for these purposes called SVNVISION which is based on Intel IPP (Integrated Performance Primitives) and achieves high speed in basic tasks for image processing.

Finally, we are preparing an extensive work about auto-calibration for optical sensors for autonomous agents using odometry.

2.1.11 Objective 4.1.

A) Development of a joint and synchronized simulation tool, capable of simulating both the execution of processes and the interconnection network. We have designed a simulation tool capable of jointly simulating the execution of processes as well as the interconnection network. Also, we started the implementation of such a tool. Currently, we are at the stage of testing the alpha version of this tool.

2.1.12 Objective 4.2.

A) Development of heuristic techniques for solving the partitioning problem in Distributed Virtual Environments. We have implemented several different heuristic search methods (including constructive, evolutionary and bio-inspired heuristics) that provides better partitions (assignments of clients to the servers in the system) while requiring shorter execution times than the methods currently proposed in the literature.

B) Characterization of Distributed Virtual Environment (DVE) systems. Surprisingly, none of the partitioning methods proposed in the literature had been correlated with system performance. We have correlated the partitioning method that is currently considered to provide the best results in terms of system performance. Since the correlation results have shown a total absence of correlation, we have studied the performance of DVE systems as the two main parameters in the partitioning method (average message latency and average CPU utilization of the servers in the system) vary. We have shown that, in order to design efficient and scalable DVE systems, the partitioning method should be targeted at avoiding that any of the servers in the system reaches 100% of CPU utilization, regardless of the message latency.

C) Development of dynamic, efficient partitioning methods for DVE systems. Taking into account the results obtained in topic B), we have developed a new partitioning method, based on a dynamic load balancing technique, that actually improves the DVE system throughput without significantly affecting system latency.

D) New goals not defined in the initial proposal. Due to some difficulties described below, we have led our efforts to propose and fulfill new goals. In particular, we have designed and implemented a new partitioning method, based on heuristic search, that not only maximizes system throughput but it also provides Quality of Service to the maximum number of clients as possible. This method starts from the method described in topic C) and uses the remaining bandwidth (if any) to reduce the latency provided to avatars to a given threshold value.

2.1.13 Objective 5.1.

A) Study of the needs for the image compression system. We have done a deep study of the different techniques—from quantization and coding of images and video based on the wavelet transform—to

determine the specifications of the electronic system that we have to design.

B) Development of a test bench. We have developed a test bench for verifying the image compression system. The tool VERA has been used for the verification of the digital electronic system. With this tool, the system is built as a set of blocks, where each block can be described using different languages (C, VHDL, Verilog, System Verilog, OpenVera, etc.).

C) Implementation of the coder in C. Once the design of the structure of the test bench has been completed, the different blocks are being implemented in C. This modularity has allowed us to do an analysis about the most suitable coder for the image compression system. In the analysis a key factor is the compression rate.

D) Use and verification of the implemented wavelet transform. In the test bench, the block of the wavelet transform made in C has been replaced by its physical implementation (obtained from a synthesizable description VHDL).

E) Implementation of the system with a microprocessor. We have worked on the implementation of the coder block, using a microprocessor designed by us.

2.1.14 Objective 6.1.

A) Extension of automatic FPGA programming tools from high-level specifications. We have extended and improved our automatic FPGA programming tool based on high-level specifications (based on an object oriented language like SmallTalk-80) for SIMD applications. This tool has been tested over current FPGA based platforms and compared with commercial high-level synthesis tools. We have obtained better area results and comparable response time implementations, under the same experimental conditions, in the design of SIMD algorithms. Recently, we have improved the type inference technique used in our tool with backward data propagation, improving previous results. Also, we have started the formalization of our methodology in order to promote the use of these techniques in future commercial hardware compilers.

B) Extension and improvement of a Multimedia Router (MMR) prototype. We have extended the design of a Multimedia Router prototype, named as Simple MMR (SMMR), on an FPGA. We have evaluated its implementation in terms of required silicon area and maximum achievable frequency. Also, we have introduced some techniques to simplify some elements of SMMR like the Link Scheduler (LS) and the Switch Scheduler (SS). As a consequence of this work we have designed a new LS based on the temperature (thermometer) encoding of priority connections. This new LS maintains QoS (Quality of Service) guarantees of MMR router while reducing its implementation area and power consumption.

C) Proposal of a multimedia server with QoS based on FPGA. We have proposed a new link scheduler for a server with QoS for LAN/SAN and high performance cluster networks. The architecture of this link scheduler uses an FPGA to maintain and update the state of each connection. We have used credit-based flow control to avoid losing data at the flit level as used in the MMR router. The first experimental results on a single Xilinx Virtex 2000 show that this architecture grows linearly with the number of connections but logarithmically with respect to the response time. We have tested configurations supporting from 4 to 128 connections in a Celoxica RC10000 platform.

D) Study of new FPGA organizations and high-speed internal lines for the design of NoC. We have carried out a first study of new FPGA organizations to evaluate their routing possibilities in the design of efficient NoC (Network on Chip). There are several possibilities to interconnect heterogeneous elements of current FPGA architectures. Their study will provide more efficient NoC designs, and help us to select the best topologies for a multiprocessor on chip. We have obtained first experimental results regarding the parameterization of routing resources for an FPGA in terms of latency and bandwidth.

2.2 Difficulties during project development

Objective 1.3 “Improving the performance of Internet routers” and Objective 2.1.A “Switch-fabric design for QoS-aware IP routers” have been canceled. These objectives were included due to the collaboration with the IBM Zurich Research Laboratory. Unfortunately, the crisis in the telecommunications market led IBM to close its Communications Department in Zurich, suddenly leading to the end of our collaboration. Thus, we decided to cancel Objective 1.3, but keeping its technical goals (enhanced switch architecture, congestion management), and developing them in a more generic form (thus making those goals to become part of Objective 1.1) or applying them to SANs (thus making those goals to become part of Objective 1.2). Regarding Objective 2.1A, we have decided to pursue a new research line aiming at reducing the complexity of cluster switches with QoS support.

Objective 2.4 “Analysis and dynamic parameter tuning for the medium access control layer in the IEEE 802.11 a/b WLAN”. Due to the development of various WLAN technologies, we have undertaken the study of other WLAN technologies: HIPERLAN/2, IEEE 802.16, IEEE 802.11 g. As part of these efforts, a novel QoS scheme for HIPERLAN/2 has been proposed. Our results show the efficiency of the proposed scheme when supporting MPEG-4 video communications. We have introduced a framework comprising various resource request mechanisms to meet the quality of service requirements of multiple applications; and resource assignment mechanisms. Our numerical results show the effectiveness of the proposed framework when operating in a multiservice network.

Objective 4.1 “Integral task scheduling” has been postponed. Due to an extremely critical personal situation, the researcher who was implementing the simulator required for this objective had to temporarily leave his research tasks since July 2004. Due to the great effort required for another researcher to finish the implementation of such a complex tool, we decided to postpone this objective until the personal situation of the researcher allows him to return to his research tasks. We expect to accomplish some of the goals planned within this objective in a few months.

Objective 5.1. “Development of a quantizer and real-time video encoder amenable for implementation in a reconfigurable VLSI architecture”. Our hardware implementation uses integer numbers, so the quality of the image is not very high as a consequence of truncating the data.

Objective 6.1. Sub-objective entitled “Design of a parallel 3D wavelet on a cluster of PCs with FPGAs from high-level specifications” has been canceled. Other subgroups of this project have successfully advanced in the development of parallel and distributed algorithms, and hardware circuitry to compute the 3D wavelet. Moreover, there is an increasing interest in the evaluation of different elements to provide QoS on SAN and high performance clusters. Thus, we decided to cancel this sub-objective and to focus our efforts in the new sub-objective entitled “Proposal of a multimedia server with QoS based on FPGA”.

Overall, despite having canceled a few research goals, the current set of research objectives is more ambitious than the original one. Effectively, new research lines, not initially planned, have been

started, such as fault tolerance in InfiniBand, fault-tolerant routing for direct networks based on the use of intermediate nodes, a routing strategy based on interval routing, power consumption reduction in interconnection networks, and reduction of the number of necessary VCs for QoS in PC clusters.

3 Result indicators and main achievements

The project started one year and a half ago. Since then, we have filed an international joint patent with a large company in UK (Xyratex) and are close to file a second one. We have also filed an international joint patent titled “Perceptual cryptography in file-sharing environments”. Also, 6 PhD students presented their dissertation, obtaining the maximum score. Moreover, we published a total of 138 papers in national and international conferences and journals. See Appendices for complete lists.

Additionally, we have become members of the European network of excellence named HiPEAC, have submitted two research proposals as workpackage leaders in the February 2005 call for Future and Emerging Technologies (FET) program on Advanced Computer Architecture (ACA), and have consolidated international collaborations and performed student and researcher exchanges with Simula Research Laboratory (Oslo, Norway), University of Southern California (Los Angeles, USA), Los Alamos National Laboratory (Los Alamos, USA), INRIA (Sophia Antipolis, France), CARV laboratory (Crete, Greece), Multimedia Communications Research Laboratory (University of Ottawa, Canada), Florida Atlantic University (Boca Raton, USA), and Virtual Reality Applications Center (Iowa State University, USA).

A Patents

Patent title: A Method of Congestion Management of a Network, a Signalling Protocol, A Switch, an End Station and a Network

Inventors: J. Duato, I. D. Johnson, F. Naven, J. Flich

Date: March 5, 2004

Country: USA

Patent title: Perceptual Cryptography in File-Sharing Environments

B PhD thesis

1. Aurelio Bermúdez, “Design of efficient subnet management mechanisms for InfiniBand networks”. PhD thesis. University of Castilla-La Mancha, September 2004.
2. Gregorio Bernabé García, “Diseño, Evaluación y Optimización de la Transformada Wavelet para Codificación de Vídeo Médico en Arquitecturas Monoprocesador”. PhD thesis. University of Murcia, November 2004.
3. Pedro Morillo Tena, “Mejora de las Prestaciones de los Entornos Virtuales Distribuidos”. PhD thesis. University of Valencia, December 2004.
4. Juan Piernas Cánovas, “Diseño, Implementación y Evaluación de un Sistema de Ficheros Transaccional con Separación de Datos y Metadatos”. PhD thesis. University of Murcia, October 2004.
5. Salvador Coll, “A strategy for efficient and scalable collective communication in the Quadrics network”. PhD thesis. Technical University of Valencia, July 2005.
6. Joaquin Cerdá, “Arquitectura VLSI de autómatas celulares para modelado físico”. PhD thesis. Technical University of Valencia.

C Publication list

1. Manuel E. Acacio and José M. García. “Techniques for Improving the Performance and Scalability of Directory-based Shared-Memory Multiprocessors: A Survey”. *Journal of Computer Science & Technology*, Vol. 3, No. 2, pp. 1-8, Octubre 2003 (Invited paper).
2. M. E. Acacio, J. González, J. M. García and J. Duato. “A Two-level Directory Architecture for Highly Scalable cc-NUMA Multiprocessors”, *IEEE Transactions on Parallel and Distributed Systems*, Vol. 16, No. 1, pp. 67-79, January 2005.
3. M. E. Acacio, J. González, J. M. García and J. Duato. “An Architecture for High-Performance Scalable Shared-Memory Multiprocessors Exploiting On-chip Integration”. *IEEE Transactions on Parallel and Distributed Systems*, Vol. 15, No. 8, pp. 755-768, Agosto 2004.
4. F.J. Alfaro, J.L. Sánchez, Studying the Influence of the InfiniBand Packet Size to Guarantee QoS, *The 10th IEEE Symposium on Computers and Communications (ISCC 2005)*. La Manga (SPAIN), 27-30 June 2005.

5. F.J. Alfaro, J.L. Sánchez, Tuning buffer size in InfiniBand to guarantee QoS, In proceedings of the 10th International Euro-Par Conference (EuroPar'04), Lecture Notes in Computer Science 3149, Springer, pp. 873-881, ISBN 3-540-22924-8, 2004
6. F. J. Alfaro, J. L. Sánchez, J. Duato, QoS in InfiniBand subnetworks, IEEE Transactions on Parallel and Distributed Systems, Vol. 15, n. 9, pp. 810-823, 2004.
7. M. Alonso, J. M. Martínez, V. Santonja, P. López, Reducing Power Consumption in Interconnection Networks by Dynamically Adjusting Link Width, Lecture Notes in Computer Science, Número: 3149 Número: 3149, Año: 2004, Páginas: 882-890 I.S.S.N.: 0302-9743
8. M. Alonso, J. M. Martínez, V. Santonja, P. López, J. Duato, Power Saving in Regular Interconnection Networks Built with High-Degree Switches, 19th International Parallel and Distributed Processing Symposium, IEEE Computer Society, Año: 2005 I.S.B.N.: 0-7695-2312-9
9. M. Alonso, J. M. Martínez, V. Santonja, P. López, J. Duato, Reducción del consumo de potencia en redes de interconexión construidas con conmutadores de alto grado, Enviado a: XVI Jornadas de Paralelismo, Año: 2005.
10. J.L. Aragón, D. Nicolaescu, A. Veidenbaum and A. M. Badulescu. "Energy-Efficient Design for Highly Associative Instruction Caches in Next-Generation Embedded Processors". In Proceedings of the International Conference on Design, Automation and Test in Europe (DATE), Paris, France, Feb. 2004.
11. J.L. Aragon and A. Veidenbaum. "Low-power Fetch Unit Design for Superscalar Processors". In Proceedings of the XV Jornadas de Paralelismo, Almeria (Spain), September 2004.
12. V. Arnau, R. Avendaño, Simulador de Técnicas de Predicción Dinámicas", XV Jornadas de Paralelismo, pp. 460-464. I.S.B.N.: 84-8240-714-7. Septiembre de 2004.
13. V. Arnau, I. Marín. Fast analysis of highly repetitive sequences in human chromosomes using a novel search algorithm, WSEAS Transactions on Biology and Biomedicine, pp. 107-111. ISSN: 1109-9518. Grecia, Enero de 2004.
14. A. Bermúdez, R. Casado, F. J. Quiles, J. Duato, Use of Provisional Routes to Speed-up Change Assimilation in InfiniBand Networks, IEEE International Workshop on Communication Architecture for Clusters (CAC'04), Santa Fe (NM), USA, IEEE Computer Society, ISSN 1530-2075, ISBN 0-7695-2132-0, pp. 186-196. Abril 2004.
15. A. Bermúdez, R. Casado, F. J. Quiles, J. Duato, Distributing InfiniBand Forwarding Tables, 10th International EuroPar Conference, Pisa, Italia, Springer-Verlag, ISSN 0302-9743, ISBN 3-540-22924-8, pp. 864-872, Septiembre 2004.
16. A. Bermúdez, R. Casado, F. J. Quiles, J. Duato, Fast Routing Computation on InfiniBand Networks. Aceptada su publicación en la revista IEEE Transactions on Parallel and Distributed Systems, IEEE Computer Society Press, 2005.
17. G. Bernabé, J. M. García y J. González. "An Efficient 3D Wavelet Transform on Hyper-Threading Technology". Technical Report. UM-DITEC-2004-02.

18. G. Bernabé, J. M. García y J. González. "Reducing 3D Fast Wavelet Transform Execution Time Using Blocking and the Streaming SIMD Extensions". *Journal of VLSI Signal Processing Systems*, Volume 41, Number 2, September 2005.
19. A. Bierbaum, P. Hartling, P. Morillo and C. Cruz-Neira, Immersive Clustering with VR Juggler, in *Int. Conf. in Computational Science and Its Applications (ICCSA' 2005)*, Singapore, May 2005. Springer Verlag.
20. B. Caminero, C. Carrión, F. J. Quiles, J. Duato y S. Yalamanchili, Traffic Scheduling Solutions with QoS Support for an Input-Buffered MultiMedia Router. Aceptada su publicación en la revista *IEEE Transactions on Parallel and Distributed Systems*, 2005.
21. Óscar Cánovas, Gabriel López, Antonio F. Gómez-Skarmeta. "A Credential Conversion Service for SAML-based scenarios". *Proceedings of First European PKI Workshop. Lecture Notes in Computer Science 3093*, Springer Verlag, pp. 297-305, Samos (Greece), Junio 2004.
22. M. Canseco, J. M. Claver, G. León, I. Vilata. "Primeras experiencias en la implementación de un encaminador QoS sobre una FPGA". *FPGAS: Computación & Aplicaciones. JCRA'04*, pp. 289-296. ISBN 84-688-7667-4. Almería, September, 2004.
23. M. Canseco, J. M. Claver, G. León, I. Vilata. "Prototipado de un MMR Simple en una FPGA". *XV Jornadas de Paralelismo '04*, pp. 345-350. ISBN 84-8240-714-7. Almería, September, 2004.
24. R. Casado. Disponibilidad de la red, *Curso de Verano UCLM: Una ventana abierta hacia el futuro de la arquitectura de computadores*, Universidad de Castilla-La Mancha, ISBN 84-921873-6-0, pp. 95-120. July 2004.
25. J. Cerdá, R. Gadea, J.D. Martinez and A. Sebastiá. "A tool for implementing and exploring SBM models: universal 1D invertible cellular automata". *Lecture Notes in Computer Science: Mechanisms, Symbols, and Models Underlying Cognition*, No. 3561, pp. 279-289, 2005.
26. J. Cerdá, O. Amoraga, R. Torres, R. Gadea and A. Sebastiá. "FPGA implementations of the RNR cellular automata to model electrostatic field". *Lecture Notes in Computer Science: High Performance Computing for Computational Science*, No. 3402, pp. 382-395, 2004.
27. J. Cerdá, R. Gadea and G. Paya. "Implementing a Margolus neighborhood cellular automata on a FPGA". *Lecture Notes in Computer Science: Artificial Neural Nets Problem Solving Methods*, No. 2687, pp. 121-128, 2003.
28. J. Cerdá, R. Gadea, V. Herrero and A. Sebastiá. "On the Implementation of a Margolus neighborhood cellular automata on FPGA". *Lecture Notes in Computer Science: Field-Programmable Logic and Applications*, No. 2778, pp. 776-785, 2003.
29. J. M. Claver, M.C. Carrión, M.Canseco, M.B. Caminero, and F.J. Quiles. "A New Efficient Link Sheduling Algorithm to Guarantee QoS on Clusters". *Euro-Par 2005*. Lisboa, Portugal, 8/30-9/2, 2005. Accepted and pending to be published in *Lecture Notes in Computer Science (LNCS)*.
30. S. Coll, J. Duato, F. J. Mora, F. Petrini, and A. Hoisie, Collective Communication Patterns on the Quadrics Interconnection Network, in V. Getov, M. Gerndt, A. Hoisie, A. Malony, and B. Miller (Ed.), *Performance Analysis and Grid Computing*, Kluwer, pp. 93-107, ISBN 1-4020-7693-2, 2004.

31. R.J. Colom, R. Gadea, F.J. Ballester and M. Martinez. "Flexible architecture for the implementation of the two-dimensional discrete wavelet transform (2D-DWT) oriented to FPGA devices". *Microprocessors and Microsystems*, No. 18, pp. 509-518, 2004.
32. P. Cuenca, L. Orozco-Barbosa, F. Delicado, A. Garrido, Breakpoint Tuning in DCT-Based Non-linear Layered Video Codecs, Vol. 16, pp. 2555-2570. *Journal on Applied Signal Processing*. November 2004.
33. F. Delicado, P. Cuenca, L. Orozco-Barbosa, Developing a QoS Framework for Media Streaming over TDMA/TDD Wireless Networks, *Journal of Wireless and Mobile Computing*. 2005.
34. F. Delicado, P. Cuenca, L. Orozco-Barbosa, A. Garrido, Design and Evaluation of a QoS-aware Framework for Wireless TDMA-TDD Networks, *Wireless Personal Communications Journal*. 2005
35. F. Delicado, P. Cuenca, L. Orozco-Barbosa, Multiservice Communications over TDMA/TDD Wireless LANs, *Lecture Notes in Computer Science*. Vol. 3510, pp. 107-116, 2005.
36. F. Delicado, P. Cuenca, L. Orozco-Barbosa, QoS-aware Resource Request Mechanisms for HIPER-LAN/2, *IEEE Global Telecommunications Conference (GLOBECOM 2004)*. December, 2004
37. F. Delicado, P. Cuenca, L. Orozco-Barbosa, A QoS-aware WLAN Resource Request Mechanism for Delay Sensitive Traffic, *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC-04)*. September, 2004
38. F. Delicado, P. Cuenca, L. Orozco-Barbosa, A QoS-aware Resource Request Mechanism for Delay Sensitive Services over TDMA/TDD Wireless Networks, *1st International Conference on E-business and Telecommunication Networks*. August, 2004
39. F. Delicado, P. Cuenca, L. Orozco-Barbosa and A. Garrido, A Class-based Allocation Mechanism for Delay Sensitive Traffic in WLAN, *18 th International Conference on Advanced Information Networking and Applications*. Fukuoka, Japan. March 2004.
40. J. Duato, J. Flich, T. Nachiondo, A Cost - Effective Technique to Reduce HOL Blocking in Single- Stage and Multistage Switch Fabrics , presentado en *12th Euromicro Conference on Parallel, Distributed and Network - based Processing*, Febrero, 2004, A Coruña.
41. J. Duato, I. Johnson, J. Flich, F. Naven, P. J. García, T. Nachiondo, A New Scalable and Cost - Effective Congestion Management Strategy for Lossless Multistage Interconnection Networks, presentado en *11th International Symposium on High- Performance Computer Architecture*, San Francisco, EEUU, Febrero 2005.
42. J. Duato, O. Lysne, R. Pang and T. M. Pinkston, A Theory for Deadlock-Free Dynamic Network Reconfiguration, *IEEE Transactions on Parallel and Distributed Systems* 16(5), 2005.
43. G. Fernández-Escribano, P. Cuenca, L. Orozco-Barbosa, A. Garrido, A Fast Intra-Frame prediction algorithm for MPEG- 2/H.264 video transcoders, *Proceedings of the IEEE International Conference on Image Processing, ICIP-05*, September 11-14, 2005, Genova, Italy.

44. G. Fernández-Escribano, P. Cuenca, L. Orozco-Barbosa, A. Garrido, Computational complexity reduction of Intra-Frame prediction in MPEG-2/H.264 video transcoders Proceedings of the IEEE International Conference on Multimedia & Expo, ICME-05, July 6-8 de 2005, Amsterdam, Holanda.
45. G. Fernández-Escribano, J. Villalón, P. Cuenca, L. Orozco-Barbosa, A. Garrido, On the Capabilities of Intra-Frame prediction in H.264 video encoders Proceedings of the IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, PacRIM-05, Victoria, B.C., Canada, August 24-26 de 2005.
46. G. Fernández-Escribano, P. Cuenca, L. Orozco-Barbosa, A. Garrido, FIMDA: A Fast Intra-Frame Mode Decision Algorithm for MPEG-2/H.264 Transcoding Lecture Notes in Computer Science (LCNS), Advanced Concepts for Intelligent Vision Systems, ACIVS'05 (in press), September 20-23, 2005, Antwerp, Belgium.
47. G. Fernández-Escribano, P. Cuenca, L. Orozco-Barbosa and A. Garrido, Transcodificación Heterogénea de Video MPEG-2/H.264: Cuestiones Abiertas, Jornadas de Paralelismo. Actas de las Jornadas de Paralelismo, pp. 132-137. September 2004
48. R. Fernández, G. Bernabé, J. M. García y M. E. Acacio. "Codificador de vídeo basado en Wavelet 3D usando OpenMP y Pthreads". Actas de las XV Jornadas de Paralelismo, Universidad de Almería, ISBN: 84-8240-714-7, Almería-España, Septiembre 2004.
49. R. Fernández, G. Bernabé, J. M. García y M. E. Acacio. "Optimizing a 3D-FWT Video Encoder for SMPs and HyperThreading Architectures". 13th EUROMICRO Conference on Parallel, Distributed and Network-Based Processing, 2005.
50. Ricardo Fernández y José M. García. RSIMx86: A Cost Effective Performance Simulator. The High Performance Computing & Simulation (HPC&S) Conference. Riga (Latvia), Junio 2005.
51. Juan Fernández, Eitan Frachtenberg, Fabrizio Petrini, Kei Davis and José Carlos Sancho. "Architectural Support for System Software on Large-Scale Clusters". Actas de International Conference on Parallel Processing (ICPP'04), Montreal, Québec (Canadá). Agosto de 2004.
52. Juan Fernández, Eitan Frachtenberg, Fabrizio Petrini, Kei Davis and José Carlos Sancho. "On the Design of System Software for Large-Scale Clusters". Actas de XV Jornadas de Paralelismo. Almería. Septiembre de 2004.
53. Juan Fernández, Fabrizio Petrini and Eitan Frachtenberg. "Monitoring and Debugging Parallel Software with BCS-MPI on Large-Scale Clusters". Workshop on System Management Tools for Large-Scale Parallel Systems. International Parallel and Distributed Processing Symposium, Denver, CO (USA). Abril de 2005.
54. Juan Fernández, Fabrizio Petrini and Eitan Frachtenberg. "Achieving Predictable Performance with BCS-MPI". Capítulo del libro "Engineering the Grid". Jack Dongarra, Hans Zima, Adolfo Hoisie, Laurence Yang and Beniamino di Martino editors. Pendiente de publicación en 2005.
55. J. L. Ferrer, E. Baydal, A. Robles, P. López, and J. Duato, MVCC: An Effective Marking/Validation-based Control Management Strategy for InfiniBand Networks, XVI Jornadas de Paralelismo, Septiembre 2005.

56. Eitan Frachtenberg, Kei Davis, Fabrizio Petrini, Juan Fernández y José Carlos Sancho. "Designing Parallel Operating Systems via Parallel Programming". Actas de Euro-Par 2004, LNCS 3149, pp. 689-696. Ed. Springer-Verlag. Pisa, Italia. Septiembre de 2004.
57. Eitan Frachtenberg, Fabrizio Petrini, Dror G. Feitelson and Juan Fernández. "Adaptive Parallel Job Scheduling with Flexible CoScheduling". IEEE Transactions on Parallel and Distributed Systems. Pendiente de publicación en 2005.
58. R. Gadea, R.J. Colom, J. Cerda and A. Sebastia. "FPGA implementation of a pipelined on-line backpropagation". Journal of VLSI signal processing, No. 40, pp. 189-213, 2005.
59. R. Gadea, A. Ramírez, J. Cerdá and R. Colom. "FPGA Implementation of Adaptive Non-Linear Predictors for Video Compression". Lecture Notes in Computer Science: Field-Programmable Logic and Applications, No. 2778, pp. 1016-1019, 2003.
60. Pablo E. García, Juan Fernández, Fabrizio Petrini and José M. García. "Cluster Computing with QsNetII". Actas de XVI Jornadas de Paralelismo. Congreso Español De Informática (CEDI' 2005). Granada. Septiembre de 2005.
61. Félix J. García, Gregorio Martínez, Óscar Cánovas, Antonio F. Gómez-Skarmeta. "A proposal of a CIM-based Policy Management Model for the OGSA Security Architecture". In Proceedings of GADA 2004, Lecture Notes in Computer Science 3292, Springer, pp. 165-174, Agia Napa (Cyprus), October 2004.
62. F. J. García, O. Cánovas, G. Martínez, y A.F. Gómez-Skarmeta: "Self-configuration of grid nodes using a policy-based management architecture", APGAC Workshop, ICCS, 6/2004.
63. I. García, R. Mollá and P. Morillo, From Continuous to Discrete Games, Proc. of 2004 Computer Graphics International (CGI'04), pp. 626-630, IEEE Computer Society Press, Crete, Greece. June, 2004.
64. P. J. García, J. Flich, J. Duato, F.J. Quiles, I. Johnson, F. Naven, On the Correct Sizing on Meshes through an Effective Congestion Management Strategy , aceptado para presentar en Euro- Par 2005, Lisboa, Agosto 2005.
65. G. Garcia-Mateos, A. Garcia-Meroño, C. Vicente, A. Ruiz, Pedro E. Lopez-de-Teruel, "Time and Date OCR in CCTV Video", Proceedings of the 13th International Conference on Image Analysis and Processing (ICIAP 2005), Cagliari (Italy), 2005.
66. M. Giménez, J. M. Benlloch, J. Cerdá, B. Escat, M. Fernández, E. N. Giménez, C. W. Lerche, J.D. Martínez, F.J. Mora, N. Pavón, F. Sánchez and A. Sebastia "Medium field of view multiflat panel-based portable gamma camera", Nuclear Instruments and Methods. ISSN: 0168-9002. Vol: 525 Issues I-2 pags. 298-302, 2004.
67. M. E. Gómez, J. Duato, J. Flich, P. Lopez, A. Robles, N.A. Nordbotten, T. Skeie, O. Lysne. "A New Adaptive Fault-Tolerant Routing Methodology for Direct Networks," International Conference on High Performance Computing (HiPC), 2004.
68. M. E. Gómez, J. Flich, P. López, A. Robles, J. Duato, N.A. Nordbotten, O. Lysne, T. Skeie. "An Effective Fault-Tolerant Routing Methodology for Direct Networks," in Proc. International Conference on Parallel Processing (ICPP), 2004, pp. 222-231.

69. M. E. Gómez, J. Duato, J. Flich, P. López, and A. Robles, N.A. Nordbotten, O.Lysne, T. Skeie. "An Efficient Fault-Tolerant Routing Methodology for Meshes and Tori," Computer Architecture Letters, Vol. 3, May 2004.
70. M. E. Gómez, P. Lopez, J. Duato. " A Memory-Effective Routing Strategy for regular Interconnection Networks". Proc. International Parallel and Distributed Processing. Best paper award. Denver (USA), 2005.
71. M. E. Gómez, P. Lopez, J. Duato. " A Memory-Effective Fault-Tolerant Routing Strategy for Direct nterconnection Networks". International Symposium on Parallel and Distributed Computing Proc. International Parallel and Distributed Processing. Lille (Francia), 2005.
72. F. Grimaldo, M. Lozano, F. Barber, J. M. Orduña, Integrating Social Skills in Task-Oriented 3-D IVA, to appear in Proceedings of 5TH International Working Conference on Intelligent Virtual Agents, 2005.
73. M. Koibuchi, J. C. Martinez, J. Flich, A. Robles, P. López, and J. Duato, Enforcing In-Order Packet Delivery in PC Clusters using Adaptive Routing Information, 2004 Workshop on High Performance Computing and Networking (HPCN-04), Tokyo, Japón, 2004
74. M. Koibuchi, J. C. Martinez, J. Flich, A. Robles, P. López, and J. Duato, "Enforcing In-Order Packet Delivery in System Area Networks with Adaptive Routing", To appear in Special Issue on Design and Performance of Networks for Super, Cluster and Grid Computing to appear in the Journal of Parallel and Distributed Computing (JPDC), 2005.
75. G. León, J. M. Claver, G. Fabregat. "Optimizing Area on the Generation of Specific Circuits in FPGAs for SIMD Applications". International Workshop on Applied Reconfigurable Computing 2005, pp. 160-167. ISBN 972-99353-8-6. Algarve, Portugal, February 22-23, 2005.
76. G. León, J. M. Claver, G. Fabregat. "Optimizing Area on the Automatic Generation of Circuits in FPGAs for SIMD Applications". XV Jornadas de Paralelismo '04, pp. 438-443. ISBN 84-8240-714-7. Almería, September, 2004.
77. Ch. W. Lerche, J. M. Benlloch, F. Sánchez, N. Pavón, N. Giménez, M. Fernández, M. Giménez, A. Sebastián, J. Martínez and F.J. Mora. "Depth of interaction detection with enhanced position-sensitive proportional resistor network", Nuclear instruments and methods in physics research ISSN: 0168-9002. Vol.: 537 Issue 1-2. Pags. 326-330, 2005
78. Gabriel López, Óscar Cánovas, Antonio F. Gómez-Skarmeta. "Servicio de acceso a la red basado en autorización SAML". En Avances en Criptología y Seguridad de la Información, Actas de la VIII Reunión Española sobre Criptología y Seguridad de la Información, pp. 625-635, Leganés (España), Septiembre 2004.
79. Gabriel López, Oscar Cánovas, Antonio F. Gómez Skarmeta, Rafael Marín. "A Network Access Control Approach based on the AAA Architecture and Authorization Attributes", International Workshop on System and Network Security. In conjunction with IEEE 19th IPDPS.
80. Gabriel López; Cánovas, O., Gómez, A. F., y Marín, R. "Different Network Access Control Methods based on the AAA Architecture." Journal of Network and Computer Applications. Aceptado. Pendiente de publicación. 2005.

81. Gabriel López; Cánovas, O., Gómez, A. F., Otenko, O., y Chadwick, D.W. "A heterogeneous network access service based on PERMIS and SAML." Proceedings of 2nd European PKI Workshop, Canterbury (UK). Aceptado. 2005.
82. O. Lysne, T. M. Pinkston and J. Duato, A Methodology for Developing Deadlock-Free Dynamic Network Reconfiguration Processes, IEEE Transactions on Parallel and Distributed Systems 16(5), 2005.
83. A. Martínez, F. J. Alfaro, J. L. Sánchez, J. Duato, A strategy to reduce the number of necessary VCs for QoS support in clusters. Accepted for HiPC 2005.
84. J. C. Martínez, J. Flich, A. Robles, P. López, and J. Duato, Supporting Adaptive Routing in InfiniBand Networks, Journal of System Architecture. No. 49 (2003), Elsevier, pp. 441-456. ISSN: 1383-762/0165-6074.
85. J. C. Martínez, J. Flich, A. Robles, P. López, J. Duato, Solving out-of-order packet delivery in interconnection networks using adaptive routing, XV Jornadas de Paralelismo, Almería 2004.
86. J. C. Martínez, M. Koibuchi, J. Flich, A. Robles, P. López, and J. Duato, In-Order Packet Delivery in Interconnection Networks using Adaptive Routing, Aceptado en 19th International Parallel & Distributed Processing Symposium (IPDPS'05), Denver (EEUU), 2005.
87. J.D. Martínez, A. Sebastián, J. Cerdá, R. Esteve, F.J. Mora, J.F. Toledo, J.M. Benlloch, N. Giménez, M. Giménez, Ch. W. Lerche, "Data acquisition electronics for positron emission mammography (PEM) detectors", Nuclear Instruments and Methods In Physics Research ISSN: 0168-9002 . Vol: 537 Issue 1-2. Pags. 335-338, 2005.
88. M. Martínez, F.J. Ballester, G. Paya, R. Colom, R. Gadea and J. Belenguer. "FPGA Custom DSP for ECG Signal Analysis and Compression". Lecture Notes in Computer Science: Field-Programmable Logic and Applications, No. 3203, pp. 954-958, 2004.
89. R. Martínez, J.L. Sánchez, F.J. Alfaro, V. Chirivella, J. Flich, Studying the effect of the design parameters on the interconnection network performance in NOWs, In proceedings of the 13th Euromicro Conference on Parallel, Distributed and Network-based Processing (PDP'98), IEEE Computer Society Press, pp. 102-109, ISBN 0-7695-2280-7, 2005.
90. F. Micó, P. Cuenca, L. Orozco-Barbosa, QoS in IEEE 802.11 wireless LAN: current research activities, in Proc. of IEEE Canadian Conference on Electrical and Computer Engineering, pp. 447- 452. May 2004.
91. F. Micó, P. Cuenca, and L. Orozco-Barbosa, QoS Mechanisms for IEEE 802.11 Wireless LANs, Lecture Notes in Computer Science, Vol. 3079, pp. 609-623. ISSN: 0302-9743. 2004.
92. J. M. Montañana, J. Flich, A. Robles, P. López, and J. Duato, A Transition-Based Fault-Tolerant Routing Strategy for InfiniBand Networks, Proceedings of the Workshop on Communication Architecture for Clusters (CAC'04), April 2004.
93. J. M. Montañana, J. Flich, A. Robles, P. López, and J. Duato, Providing Fault Tolerance To InfiniBand Networks, Actas de las XV Jornadas de Paralelismo, Septiembre 2004.

94. J. M. Montañana, J. Flich, A. Robles, and J. Duato, An Scalable Methodology for Computing Fault-free Paths in InfiniBand Torus Networks, Submitted to 6th International Symposium on High Performance Computing (ISHPC-VI), Japan, September 2005.
95. J. M. Montañana, J. Flich, A. Robles, and J. Duato, Reachability-Based Fault-Tolerant Routing Methodology, XVI Jornadas de Paralelismo, Septiembre 2005.
96. P. Moreno, P. Morillo, J. M. Orduña, and J. Duato, Calidad de Servicio en Entornos Virtuales Distribuidos, accepted for presentation in XVI Jornadas de Paralelismo, I Congreso Español de Informática. Granada (SPAIN), 2005.
97. P. Morillo, J. M. Orduña, M. Fernández, and J. Duato, A Comparison Study of Metaheuristic Techniques for Providing QoS to Avatars in DVE Systems, in Proceedings of the International Conference on Computational Science and its Applications 2004 (ICCSA'04), pp. 661-670, ISBN 3-540-22056-9, May 2004.
98. P. Morillo, J.M. Orduña, M. Fernández, J. Duato, A New Partitioning Approach for Distributed Virtual Environment Systems, Proceedings of XV Jornadas de Paralelismo, pp. 12-17. ISBN 84-8240-714-7. 2004.
99. P. Morillo, J.M. Orduña, M. Fernández, J. Duato, Metaheuristic Solutions for Providing QoS in DVE Systems, Proceedings of XV Jornadas de Paralelismo, pp. 18-23. ISBN 84-8240-714-7. 2004.
100. P. Morillo, J. M. Orduña, M. Fernández, and J. Duato, A Fine-Grain Method for Solving the Partitioning Problem in Distributed Virtual Environment Systems, in Proceedings of the IASTED International Conference on Parallel and Distributed Computing and Systems (PDCS 2004), Iasted Acta Press, pp. 292-297, ISBN 0-88986-421-7, 2004.
101. P. Morillo, J.M. Orduña and M. Fernández, A Comparison Study of evolutive algorithms for Solving the Partitioning Problem in Distributed Virtual Environment Systems, Parallel Computing, Vol. 30, No. 5, pp. 585-610, 2004.
102. P. Morillo, P. López, J.M. Orduña and M. Fernández, A New Genetic Approach for the Partitioning Problem in Distributed Virtual Environment Systems, Lecture Notes in Artificial Intelligence, Vol. 3040, pp. 76-85, 2004.
103. P. Morillo, J. M. Orduña, M. Fernández, and J. Duato, A Method for Providing QoS in Distributed Virtual Environment Systems, in Proceedings of the EuroMicro Conference on Parallel, Distributed and Network-Based Processing (PDP 2005), IEEE Computer Society Press, pp. 152-159, ISBN 0-7695-2280-7, 2005.
104. P. Morillo, J. M. Orduña, M. Fernández, and J. Duato, Improving the Performance of Distributed Virtual Environment Systems, IEEE Transactions on Parallel and Distributed Systems, Vol., No 7, pp. 637-649, 2005.
105. P. Morillo, M. Fernández, and J. M. Orduña, Solving the Partitioning Problem in Distributed Virtual Environment Systems Using Evolutive Algorithms, Handbook of Bioinspired Algorithms and Applications, chapter 30, pp. 529-552, Chapman&Hall, 2005.

106. T. Nachiondo, J. Flich, J. Duato, Efficient Reduction of HOL Blocking in Multistage Networks , presentado en 2005 Workshop on Communication Architecture for Clusters, dentro del 19th International Parallel and Distributed Processing Symposium, Abril 2005, Denver, USA.
107. T. Nachiondo, J. Flich, J. Duato, M. Gussat, Cost /Performance Trade - offs and Fairness Evaluation of Queue Mapping Policies, aceptado para presentar en Euro- Par 2005, Lisboa, Agosto 2005.
108. N. A. Nordbotten, M. E. Gómez, J. Flich, P. López, A. Robles, T. Skeie, O. Lysne, J. Duato. "A Fully Adaptive Fault-Tolerant Routing Methodology Based on Intermediate Nodes," IFIP International Conference on Network and Parallel Computing (NPC), 2004.
109. J. M. Orduña, F. Silla, and J. Duato, On the Development of a Communication-Aware Task Mapping Technique, Journal of Systems Architecture, Vol. 50, No. 4, pp. 207-220, 2004.
110. Alejandro Orozco, Manuel E. Acacio and José M. García. "Estudio y Evaluación del Encaminamiento Multidestino en Arquitecturas cc-NUMA con Directorios Comprimidos". Actas de las XVI Jornadas de Paralelismo, Septiembre 2005.
111. L. Orozco-Barbosa, M. Toukourou, P. Cuenca, A. Miri, A QoS Framework for Scalable MPEG-2 Video-on-Demand Services over IP-based Networks, Journal of Internet Technology. Vol. 5(4). pp.389-403. October 2004.
112. G. Paya, M. Martinez, F.J. Ballester and F.J. Mora. "Fully parameterized discrete wavelet packet transform architecture oriented to FPGA". Lecture Notes in Computer Science: Field-Programmable Logic and Applications, No. 2778, pp. 533-542, 2003.
113. Fabrizio Petrini, Adam Moody, Juan Fernández, Eitan Frachtenberg and Dhabaleswar K. Panda. "NIC-based Reduction Algorithms for Large-Scale Clusters". International Journal of High Performance Computing and Networking (IJHPCN). Pendiente de publicación en 2005.
114. J. Piernas, T. Cortes, and J. M. García. "Traditional File Systems versus DualFS: a Performance Comparison Approach". IEICE Transactions on Information and Systems, Vol.E87-D No.7 p. 1703-1711, July 2004.
115. R. Risueño, P. Cuenca, F. Delicado and L.Orozco-Barbosa, On the Effect of Handover Mechanisms on the Performance of Video Communications in WATM Networks, Journal of Wireless and Mobile Computing. 2005 (in press).
116. R. Risueño, P. Cuenca, F. Delicado, L. Orozco-Barbosa, and A. Garrido, On the Traffic Disruption Time and Packet Lost Rate during the Handover Mechanisms in Wireless Networks, 18 th International Conference on Advanced Information Networking and Applications. Fukuoaka, Japan. March 2004.
117. R. Risueño, P. Cuenca, F. Delicado, L. Orozco-Barbosa, and A. Garrido, On the Effect of the Handover Mechanisms in QoS Performance in Wireless Multimedia Networks, 6th International Workshop on Multimedia Network Systems and Applications. Tokyo, Japan. March 2004.
118. Alberto Ros, Manuel E. Acacio and José M. García. "A Novel Lightweight Directory Architecture for Scalable Shared-Memory Multiprocessors". Proc. of the Euro-Par 2005, LNCS, Agosto 2005.

119. Alberto Ros, Manuel E. Acacio and José M. García. "Diseño y Evaluación de una Arquitectura de Directorio Ligero para Multiprocesadores de Memoria Compartida Escalables". Actas de las XVI Jornadas de Paralelismo, Septiembre 2005.
120. S. Rueda, P. Morillo, J. M. Orduña, and J. Duato, A Sexual Elitist Genetic Algorithm for Providing QoS in Distributed Virtual Environment Systems, Proc. of 2005 Int. Parallel and Distributed Processing Symposium (IPDPS' 2005). ISBN: 0-7695-2312-9. 2005.
121. S. Rueda, P. Morillo, J. M. Orduña, and J. Duato, Algoritmos Genéticos para Proporcionar Calidad de Servicio en Entornos Virtuales Distribuidos, accepted for presentation in XVI Jornadas de Paralelismo, I Congreso Español de Informática. Granada (SPAIN), 2005.
122. F. Sanchez, J. M. Benlloch, B. Escat, N. Pavón, F.J. Mora, A. Sebastià, E. Porras, D. Kadi-Hanifi, J. A. Ruiz, "Design and tests of a portable mini gamma camera", Medical Physics. ISSN: 0094-2405. Vol. 31 N° 6. Pag. 1384-1397, 2004.
123. Manuel Sánchez, Óscar Cánovas, Diego Sevilla. "Firma de trabajos en la integración Globus 2 y Globus 3". In Avances en Criptología y Seguridad de la Información, Actas de la VIII Reunión Española sobre Criptología y Seguridad de la Información, pp. 341-352, Leganés (Spain), September 2004.
124. Manuel Sánchez, Óscar Cánovas, Diego Sevilla, Antonio F. Gómez-Skarmeta. A Service-Based Architecture for Integrating Globus 2 and Globus 3. In Advances in Grid Computing - EGC 2005. LNCS 3470. The Netherlands, February 14-16, 2005.
125. José Carlos Sancho, Fabrizio Petrini, Greg Jonhson, Juan Fernández y Eitan Frachtenberg. "On the Feasibility of Incremental Checkpointing for Scientific Computing". Actas de International Conference on Parallel and Distributed Processing Symposium (IPDPS'04), Santa Fe, NM (USA). Abril de 2004.
126. J. C. Sancho, A. Robles, J. Duato, An effective methodology to compute up*/down* routing tables in NOWs, IEEE Transaction on Parallel and Distributed Systems Vol. 15, No. 6, 2004, IEEE Computer Society, ISSN: 1045-9219.
127. T. Skeie, O. Lysne, J. Flich, P. López, A. Robles, J. Duato, LASH-TOR: A Generic Transition-Oriented Routing Algorithm, 10th International Conference on Parallel and Distributed Systems (ICPADS 2004), IEEE Computer Society Press, Newport Beach, CA, USA, July 2004, ISBN: 0-7695-2152-5.
128. Francisco J. Villa, Manuel E. Acacio and José M. García. "Evaluating IA-32 Web Servers through Simics: A Practical Experience". Journal of Systems Architecture, Vol. 51 (2005) 251-264.
129. Francisco J. Villa, Manuel E. Acacio and José M. García. "Thread-Level Speculation and Chip-Multiprocessors: An Overview". Proc. of the PARA'04 Workshop on State-of-the-Art Scientific Computing, Lyngby (Dinamarca), Junio 2004.
130. Francisco J. Villa, Manuel E. Acacio and José M. García. "On the Evaluation of x86 Web Servers Using Simics: Limitations and Trade-Offs". Proc. of the 4th International Conference on Computational Science, LNCS 3036, pp. 541-544, Cracovia (Polonia), Junio 2004.

131. Francisco J. Villa, Manuel E. Acacio and José M. García. "Evaluación de un Servidor Web Multiprocesador mediante Simics". Actas de las XIV Jornadas de Paralelismo, Septiembre 2003.
132. Francisco J. Villa, Manuel E. Acacio and José M. García. "Memory Subsystem Characterization in a 16-core Snoop-Based Chip-Multiprocessor Architecture". 2005 International Conference on High Performance Computing and Communications. Septiembre, 2005.
133. Francisco J. Villa, Manuel E. Acacio and José M. García. "Performance Evaluation of a Bus-Based 16-core Chip-Multiprocessor". Actas de las XVI Jornadas de Paralelismo, Septiembre 2005.
134. J. E. Villalobos, J. L. Sánchez, J. A. Gámez, J. C. Sancho, A. Robles, A Methodology to Evaluate the Effectiveness of Traffic Balancing Algorithms, Lecture Notes in Computer Science. No. 3149. Springer-Verlag, 2004, ISBN: 3-540-22924-8.
135. J. Villalón, P. Cuenca and L. Orozco-Barbosa, QoS Provisioning Mechanisms for IEEE 802.11 WLAN: A Performance Evaluation, Proc. of 10th IEEE/IFIP International Conference on Personal Wireless Communications PWC'05. Colmar, August 2005.
136. J. Villalón, F. Mico, P. Cuenca and L. Orozco-Barbosa, QoS Support for Time-Constrained Multimedia Communications in IEEE 802.11 WLANs: a Performance Evaluation, Proc. of IEEE International Conference on Multimedia Communications Systems (ICMCS-05). Montreal, August 2005.
137. J. Villalón, P. Cuenca and L. Orozco-Barbosa, Limitations and Capabilities of QoS Support in IEEE 802.11 WLANs, Proc. of IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, PACRIM-05. Victoria, August 2005.
138. J. Villalón, P. Cuenca and L. Orozco-Barbosa, On the Effectiveness of IEEE 802.11e QoS Support in Wireless LAN: A Performance Analysis, Submitted to the International Conference on High Performance Computing and Communications, HPCC-05, Naples, Italy, September 2005.

Computación natural con redes de procesadores evolutivos, y sus relaciones con las redes de neuronas y con los autómatas biomoleculares de Watson-Crick

TIC2003-09319-C03

Carlos Martín Vide * José María Sempere Luna
Universidad Rovira i Virgili Universidad Politécnica de Valencia

Luis Fernando Mingo López
Universidad Politécnica de Madrid

Abstract

The project aims at exploiting networks of evolutionary processors as a new bio-inspired computational architecture and at exploring their connection with neural networks and Watson-Crick biomolecular automata. Issues of computational power, universality, complexity and applicability are addressed. The project is basically theoretical and mostly of a mathematical character, and intends to develop the field in close contact with other biomolecular unconventional models of computing.

Keywords: network of evolutionary processors, finite automaton, neural network, biomolecular computing, unconventional computing, theoretical computer science, formal language theory

1 Aims

1.1 Intended aims (URV = Rovira i Virgili University)

1. Consider systematically the string operations suggested by biological processes, devise computing (language generating) mechanisms based on them, and investigate their power and properties.
2. Simplify the networks of evolutionary processors (NEPs) defined so far and improve the existing results.
3. Examine the optimality of universality results in terms of the number of processors. Which are the size and the properties of sub-universal families of sets of vectors of natural numbers and of languages?

*Email: cmv@astor.urv.es

4. What is the descriptonal complexity of these networks with respect to different criteria similar to those investigated for many generative devices in formal language theory? Such criteria might be statical ones like the number of evolution rules, the size of the initial population, or dynamical ones like the number of evolutionary events.
5. Examine systematically the possibilities of solving NP-complete problems in polynomial/linear time and with polynomially bounded resources. Consider sets of NP-complete problems which can be uniformly treated in this framework.
6. It seems natural to investigate the effect of processing different objects, not only strings: trees, graphs, circular strings, pictural words. Consider networks with an enriched structure, maybe without good computability properties, but (hopefully) modeling the functioning of a biological system in a way relevant for biochemistry.
7. Examine the linguistic interpretation and relevance of the many operations with strings suggested by genome processes, and examine the relevance of NEPs for modelling the brain linguistic behaviour.

1.2 Intended aims (UPV = Technical University of Valencia)

1. Propose models of biomolecular finite automata as an alternative to present-day models. Refine and consider variants for present-day models.
2. Characterize computationally and from the viewpoint of complexity the models in 1, with a special emphasis on descriptonal complexity aspects.
3. Study inductive inference aspects of biomolecular finite automata and of some languages based on operations with DNA (duplication, cross-over, splicing, etc.).
4. Develop synthetic and real test banks for the inductive inference of the models.
5. Elaborate biomolecular sequence processing tools based on finite automata: simulation of models and processing of DNA and proteomic sequences (search of genes, search of contexts, 'in silico' simulation of DNA micro-arrays, etc.).
6. Integrate biomolecular finite automata models into NEPs: dynamic extension of NEPs, by means of varying filters and evolutionary rules according to an inductive criterion.

1.3 Intended aims (UPM = Technical University of Madrid)

1. Simulate distributively a neural network.
2. Describe the similarities between neural networks and NEPs.
3. Formally define NEPs which involve learning.
4. Distributively implement NEPs.

2 Level of achieved success

2.1 Achieved aims (URV)

1. This goal was achieved by the results described in the third item of the next section. The operations considered in [1, 2, 12, 13, 24, 35, 52, 53] were splicing, translocation, duplication and Watson-Crick superposition. Furthermore, we considered two biological phenomena like cell-differentiation and gene shuffling. [12] is continued in [88], where duplication codes are investigated.
2. The papers [22, 47] are contributions to this objective. Some results are quite important, but there are still some problems which require a deeper investigation. A new direction was opened in [47], where so-called homogeneous and free NEPs were introduced. This direction seems very attractive offering a lot of possibilities and posing hard problems.
3. The paper [33] establishes the first result in this area. A more detailed discussion was started in [77] and [81]. The same program is to be applied to networks of splicing processors in [84].
4. Contributions [11, 10, 21, 41, 45] contribute to accomplish this objective. Some other papers like [77, 84] are in progress.
5. New solutions to NP-complete problems are presented in [22, 27]. Furthermore, in [22] one discusses the mathematical basis for a systematic study of NEPs as solving problem mechanisms. By the new characterizations in [77, 81, 84], we hope to shed more light on this topic.
6. We identified serious difficulties in our attempt to extend some of the results presented above to structures other than strings, especially tree-like structures. However, some work in progress seems to represent a promising starting point towards this direction.
7. The paper [23] represents a first step in this direction. Actually, a possible implementation is being undergone. Other computer scientists are working on different implementations. Several researchers and students from Spain, France, Germany, Hungary, India, Romania, etc. are working on different aspects offered by this new biologically inspired computational model.

2.2 Scientific-technical results (URV)

The results obtained so far by the members of the team can be divided into four main directions of research, all of them appearing as objectives and goals in the planned work programme:

1. *Computational aspects of NEPs: power, computational complexity, descriptive complexity.* Several computational aspects of NEPs were studied in [10, 21, 27, 33, 41], and it was shown that these mechanisms are able to solve hard problems which are intractable on usual computers. These mechanisms can either generate languages, accept languages or solve mathematical problems.
In [10], one shows that these devices viewed as language generators are computationally complete, that is they have the same computational power as a Turing machine, which

is the classical computing model widely investigated in theoretical computer science. The same result holds for accepting NEPs ([21]). The proof presented there allowed us to obtain deeper results. First, a new characterization of the famous class NP: namely, this class is exactly the class of languages accepted by NEPs in polynomial time. It is worth mentioning that such results are quite rare in the literature.

Second, we propose in [33] a construction of an accepting NEP which behaves as a universal device in the class of all these devices. The construction is mainly based on the simulation of a Turing machine, as reported in [21], which can, in its turn, simulate any accepting NEP, and then an accepting NEP which simulates this Turing machine. It is clear that this approach can be applied to other bio-inspired computing models which are computationally complete.

On the other hand, we continue the research line started in some previous papers proposing solutions, having a linear number of computational steps, to other NP-complete problems based on NEPs as solving problem devices: 3-CNF-SAT and the Hamiltonian Path Problem in [27]. [41] is a survey discussing the most recent results and proposing some directions for further research.

The descriptonal complexity of NEPs was studied in [11, 45, 47]. The measures investigated in these papers were the size, number of rules, and number of symbols. The results are rather unexpected, since NEPs with limited (polynomially bounded) resources are able to solve hard problems. Thus, [11] gives a rather complete picture of the size (number of nodes) of the NEPs generating or accepting some well-known classes of languages. One proves ([11]) that in all cases the underlying structure does not depend on a given language (this is the common case for many other bio-inspired computational models). [47] is a first step in an investigation regarding homogeneous and free NEPs. This is an extremely simple variant of generating NEP shown to be a computationally complete model in [10]. Two simplifications are considered: all filters are completely removed such that words can freely navigate in the networks, and all evolutionary operations are applied uniformly. In our view, this variant seems most suitable for a biological implementation. A survey presenting the most important results concerning the descriptonal complexity of NEPs is given in [45].

The NEP model has drawn the attention of other researchers so that it was considered one of the main topics of a special workshop of the European Conference on Artificial Intelligence in 2004. As well, it was the subject of a special session of the International Work-Conference on the Interplay between Natural and Artificial Computation in 2005.

2. *Possible applications and implementations of NEPs. Networks with other type of bio-inspired processors.* In [22], we replace the evolutionary processors placed in the nodes of NEPs with splicing processors. Actually, we replace the point mutations associated with each node in the original NEPs with another operation widely mentioned in the Darwinian evolution: that of splicing. We present linear time solutions to two NP-complete problems, namely SAT and the Directed Hamiltonian Path Problem, based on accepting networks of splicing processors having all resources (size, number of rules and symbols) linearly bounded by the size of the given instance.

The paper [23] proposes a very simple implementation of NEPs for accepting and analysing linguistic structures with the shape $NP \vee NP$. The formalization takes advantage of

NEPs' features -modularity, specialization and parallelism- to develop a syntactic recognizer able to distinguish correct sentences working with lineal strings as input and lineal labeled structures as output.

3. *Combinatorial and algorithmic properties of some formal operations inspired by the genome evolution.* In [1], we propose a computational model for cell differentiation and answer several decidability questions. The work was published in a journal with a wide visibility in the biological community, being previously presented in a very selective conference dedicated to bioinformatics.

Another approach to the main topic is [52], where we define a new operation inspired by Watson-Crick complementarity and annealing of DNA sequences. This operation turned out to be very useful in a formal description of some DNA experiments previously reported in the literature.

One of the chromosome rearrangements which appears very often in the genome evolution is duplication. Viewed as a formal operation on strings, it is investigated in [12]. Despite that several problems remained unsolved, the case of uniformly bounded duplication was solved completely.

A new type of shuffle inspired from genetics is introduced in [53]. We extend the basic shuffle on words and languages, a well-known operation in theoretical computer science, by introducing three synchronized shuffles. These synchronized shuffles have some relevance to molecular biology, since they may be viewed as the formal representations of various kinds of gene linkage during genome shuffling. More precisely, each synchronized shuffle preserves the genetic backbone of the organisms, as well as the linked genes, by requiring the synchronization of some predefined genes while all other genes are arbitrarily shuffled.

Several formal language based approaches in biology are discussed in [2], which was published in a high standard biological journal.

The problem of computing distances between genomes, in spite of huge advances, is still far from being solved. We contributed to this area with the paper [13], where we proposed a new translocation distance and gave two algorithms for computing it: a polynomial one and an approximate one.

The classical definition of splicing (a formal model of recombination of DNA molecules induced by restriction enzymes) is reconsidered in [35], where we started from the premise that the results reported so far spurred research aimed at finding finite H systems, with a definition that is both conceptually simple and as biochemically realistic as possible, which are able to perform universal computations. In a certain way, the systems we proposed in this paper represent just what the theory of splicing has been looking for, namely finite splicing systems that are computationally complete. Moreover, we consider that the new definitions are quite natural and well-motivated extensions of the classical splicing operation.

4. *Interrelationships with other parallel or distributed computing systems like grammar and automata systems.* We have considered also the applicability of a rather mature theory, that of automata and grammar systems, to attack some proposed problems in our work

plan. This aspect was studied in several papers: [24, 30, 42, 50]. The paper [24] is a good example of how a contextual system could help in modelling a real biological phenomenon. It turned out that automata systems were more appropriate for such purposes than grammar systems.

2.3 Scientific-technical results (UPV)

1. We have obtained a representation theorem for Watson-Crick finite automata.
2. We have formalized classes of languages accepted by Watson-Crick finite automata.
3. We have studied the descriptive complexity of minimal Watson-Crick finite automata.
4. We have obtained inference algorithms for subclasses of regular and context-free languages.
5. We have introduced learning and covering techniques for cellular models with membranes.

2.4 Scientific-technical results (UPM)

1. We have formally defined neural networks with axo-axonic connections. Among the biological neural connections, one has to distinguish: axo-synaptic, axo-somatic and axo-axonic connections. Artificial neural networks are usually based on axo-synaptic connections. A neural network with axo-axonic connections has been modelled. Such networks are a superset of the set of synaptic connections and show universality properties.
2. We have proved the universality of neural networks with axo-axonic connections, on the basis of Funahashi and Sprecher theorems, which show the universality of multi-level perceptrons.
3. We have applied axo-axonic networks to market prediction and temporal series analysis.
4. We have started to formalize NEPs with filters at the connections in order to incorporate learning, in a way similar to self-organizing maps.
5. We have compared neural models and NEPs to conclude that the best architecture for the latter ones is similar to self-organizing maps in neural networks.

3 Indicators of results

3.1 Staff being trained

3.1.1 (URV)

The dissertations of the following students of our International PhD School in Formal Languages and Applications (having been awarded the 2003, 2004 and 2005 MEC Mención de Calidad) have some connection with NEPs:

Artiom Alhazov: URV fellow, dissertation expected in 2006

Madalina Barbaiani: GENCAT fellow, dissertation expected in 2008

Leonor Becerra Bonache: MEC fellow, dissertation expected in 2006
 Cristina Bibire: MEC fellow, dissertation expected in 2008
 Liliana Cojocaru: MAEC-AECI fellow, dissertation expected in 2006
 Carlos Cruz Reyes: GENCAT fellow, dissertation expected in 2008
 Adrian Horia Dediu: URV fellow, dissertation expected in 2006
 Szilárd Zsolt Fazekas: MEC fellow, dissertation expected in 2008
 Adela Grando: MEC fellow, dissertation expected in 2007
 Mihai Ionescu: MEC fellow, dissertation expected in 2008
 Peter Leupold: MEC fellow, dissertation expected in 2006
 Guangwu Liu: GENCAT fellow, dissertation expected in 2008
 Remco Loos: MEC fellow, dissertation expected in 2007
 Gabriela Martín: URV fellow, dissertation expected in 2007
 Benedek Nagy: URV fellow, dissertation expected in 2007
 Tsetsegkhand Namsrai: GENCAT fellow, dissertation expected in 2008
 Anthonath Roslin Sagaya Mary: MAEC-AECI fellow, dissertation expected in 2007

3.1.2 (UPV)

Gloria Álvarez: UPV fellow, near DEA defense
 Marcelino Campos: UPV fellow, 1st year PhD student
 Piedad Peris: UPV fellow, 1st year PhD student

3.1.3 (UPM)

A new research group on natural computing has been created where three lecturers at the Technical School of Computer Science are starting to work for their dissertation on NEPs and their hardware implementation.

3.2 Publications (URV + UPV + UPM)

Note: The authors who are members of the research teams are written in bold.

3.2.1 Papers in refereed journals

1. I. Ardelean, M. Gheorghe, **C. Martín-Vide** & **V. Mitrana** (2004), A Computational Model for Cell Differentiation, *BioSystems*, 76(1-3): 169-176.
2. M. Gheorghe, **V. Mitrana** (2004), A Formal Language Based Approach in Biology, *Comparative and Functional Genomics*, 5: 91-94.
3. **J.M. Sempere** (2004), A Representation Theorem for Languages Accepted by Watson-Crick Finite Automata, *Bulletin of the EATCS*, 83: 187-191.
4. **M. Cavaliere** & P. Leupold (2004), Evolution and Observation: A Non-Standard Way to Generate Formal Languages, *Theoretical Computer Science*, 321(2-3): 233-248.

5. **L.F. Mingo**, L. Aslanyan, **J. Castellanos**, M. Díaz & V. Riazanov (2004), Fourier Neural Networks: An Approach with Sinusoidal Activation Functions, *International Journal on Information Theories and Applications*, 11: 52-55.
6. R. Freund, **C. Martín-Vide** & G. Paun (2004), From Regulated Rewriting to Computing with Membranes: Collapsing Hierarchies, *Theoretical Computer Science*, 312(2-3): 143-188.
7. **M. Cavaliere** & D. Genova (2004), P Systems with Symport/Antiport of Rules, *Journal of Universal Computer Science* 10(5): 540-558.
8. **L.F. Mingo**, **J. Castellanos**, G. López & **F. Arroyo** (2004), Time Series Analysis with Neural Networks, *WSEAS Transactions on Business and Economics*, 4(1): 303-310.
9. **M.D. Jiménez-López** & G. Bel-Enguix (2005), Biosyntax. An Overview, *Fundamenta Informaticae*, 64: 1-12.
10. E. Csuhaj-Varjú, **C. Martín-Vide** & **V. Mitrana** (2005), Hybrid Networks of Evolutionary Processors Are Computationally Complete, *Acta Informatica*, 41(4-5): 257-272.
11. **J. Castellanos**, P. Leupold, **V. Mitrana** (2005), On the Size Complexity of Hybrid Networks of Evolutionary Processors, *Theoretical Computer Science*, 330(2): 205-220.
12. P. Leupold, **C. Martín-Vide** & **V. Mitrana** (2005), Uniformly Bounded Duplication Languages, *Discrete Applied Mathematics*, 146: 301-310.

3.2.2 Papers in LNCS volumes

13. **C. Martín-Vide** & **V. Mitrana** (2004), A New Uniform Translocation Distance, in A. Fred, T. Caelli, R.P.W. Duin, A. Campilho & D. de Ridder, eds., *Structural, Syntactic, and Statistical Pattern Recognition*, *Lecture Notes in Computer Science* 3138: 279-286. Springer, Berlin.
14. **M. Cavaliere**, N. Jonoska, S. Yogeve, R. Piran, E. Keinan & N.C. Seeman (2004), Biomolecular Implementation of Computing Devices with Unbounded Memory, in C. Ferretti, G. Mauri & C. Zandron, eds., *DNA Computing*, *Lecture Notes in Computer Science* 3384: 35-49. Springer, Berlin.
15. A. Alhazov & **M. Cavaliere** (2004), Computing by Observing Bio-systems: The Case of Sticker Systems, in C. Ferretti, G. Mauri & C. Zandron, eds., *DNA Computing*, *Lecture Notes in Computer Science* 3384: 1-13. Springer, Berlin.
16. **M. Cavaliere** & P. Leupold (2004), Evolution and Observation: A Non-standard Way to Accept Formal Languages, in M. Margenstern, ed., *Machines, Computations, and Universality*, *Lecture Notes in Computer Science* 3354: 153-163. Springer, Berlin.
17. P. Leupold, **V. Mitrana** & **J.M. Sempere** (2004), Formal Languages Arising from Gene Repeated Duplication, in N. Jonoska, G. Paun & G. Rozenberg, eds., *Aspects of Molecular Computing: Essays Dedicated to Tom Head on the Occasion of His 70th Birthday*, *Lecture Notes in Computer Science* 2950: 297-308. Springer, Berlin.

18. R. Ceterchi, **C. Martín-Vide** & K.G. Subramanian (2004), On Some Classes of Splicing Languages, in N. Jonoska, G. Paun & G. Rozenberg, eds., Aspects of Molecular Computing: Essays Dedicated to Tom Head on the Occasion of His 70th Birthday, Lecture Notes in Computer Science 2950: 84-105. Springer, Berlin.
19. L. Kari, **C. Martín-Vide** & A. Paun (2004), On the Universality of P Systems with Minimal Symport/Antiport Rules, in N. Jonoska, G. Paun & G. Rozenberg, eds., Aspects of Molecular Computing: Essays Dedicated to Tom Head on the Occasion of His 70th Birthday, Lecture Notes in Computer Science 2950: 254-265. Springer, Berlin.
20. A. Alhazov, **C. Martín-Vide** & L. Pan (2004), Solving Graph Problems by P Systems with Restricted Elementary Active Membranes, in N. Jonoska, G. Paun & G. Rozenberg, eds., Aspects of Molecular Computing: Essays Dedicated to Tom Head on the Occasion of His 70th Birthday, Lecture Notes in Computer Science 2950: 1-22. Springer, Berlin.
21. M. Margenstern, **V. Mitrana** & M.J. Pérez-Jiménez (2005), Accepting Hybrid Networks of Evolutionary Processors, in C. Ferretti, G. Mauri & C. Zandron, eds., DNA Computing, Lecture Notes in Computer Science 3384: 235-246. Springer, Berlin.
22. F. Manea, **C. Martín-Vide** & **V. Mitrana** (2005), Accepting Networks of Splicing Processors, in S.B. Cooper, B. Löwe & L. Torenvliet, eds., New Computational Paradigms, Lecture Notes in Computer Science 3526: 300-309. Springer, Berlin.
23. G. Bel-Enguix & **M.D. Jiménez-López** (2005), Analysing Sentences with Networks of Evolutionary Processors, in J. Mira & J.R. Álvarez, eds., in Artificial Intelligence and Knowledge Engineering Applications: A Bioinspired Approach, part II, Lecture Notes in Computer Science 3562: 102-111. Springer, Berlin.
24. **C. Martín-Vide** & **V. Mitrana** (2005), Contextual Information Systems, in A. Dey, B. Kokinov, D. Leake & R. Turner, eds., Modeling and Using Context, Lecture Notes in Computer Science 3554: 304-315. Springer, Berlin.
25. **M. Cavaliere**, M. Ionescu & T.-O. Ishdorj (2005), Inhibiting/De-Inhibiting Rules in P Systems, in G. Mauri, G. Paun, M.J. Pérez-Jiménez, G. Rozenberg & A. Salomaa, eds., Membrane Computing, Lecture Notes in Computer Science 3365: 224-238. Springer, Berlin.
26. **C. Martín-Vide** & T.-O. Ishdorj (2005), Modeling Neural Processes in Lindenmayer Systems, in J. Cabestany, A. Prieto & F. Sandoval Hernández, eds., Computational Intelligence and Bioinspired Systems, Lecture Notes in Computer Science 3512: 145-152. Springer, Berlin.
27. F. Manea, **C. Martín-Vide** & **V. Mitrana** (2005), Solving 3CNF-SAT in Linear Time Using WWW, in M. Margenstern, ed., Machines, Computations, and Universality, Lecture Notes in Computer Science 3354: 269-280. Springer, Berlin.
28. **M. Cavaliere** & D. Sburlan (2005), Time-Independent P Systems, in G. Mauri, G. Paun, M.J. Pérez-Jiménez, G. Rozenberg & A. Salomaa, eds., Membrane Computing, Lecture Notes in Computer Science 3365: 239-258. Springer, Berlin.

3.2.3 Papers in international conference proceedings

29. G. Bel-Enguix & **M.D. Jiménez-López** (2004), A Membrane Systems Explanation of Semantic Change, in E. Csuhaj-Varjú & G. Vaszil, eds., Proceedings of Grammar Systems Week 2004: 30-49. Hungarian Academy of Sciences, Budapest.
30. M.A. Grando & **V. Mitrana** (2004), Can PC Grammar Systems Benefit from Concurrent Programming?, in E. Csuhaj-Varjú & G. Vaszil, eds., Proceedings of Grammar Systems Week 2004: 200-211. Hungarian Academy of Sciences, Budapest.
31. **D. López**, **A. Cano**, M. Vazquez de Parga, B. Calles, **J.M. Sempere**, T. Pérez, **J. Ruiz** & P. García (2004), Detection of Functional Motifs in Biosequences: A Grammatical Inference Approach, in X. Messeguer & G. Valiente, eds., Proceedings of the 5th Annual Spanish Bioinformatics Conference: 72-75. Technical University of Catalonia, Barcelona.
32. L. Pan & **C. Martín-Vide** (2004), Solving Multidimensional 0-1 Knapsack Problem by P Systems with Input and Active Membranes, in G. Paun, A. Riscos-Núñez, Á. Romero-Jiménez & F. Sancho-Caparrini, eds., Second Brainstorming Week on Membrane Computing, Research Group in Natural Computing, Technical Report 01/2004: 342-353. Universidad de Sevilla, Sevilla.
33. F. Manea, **C. Martín-Vide** & **V. Mitrana** (2005), A Universal Accepting Hybrid Network of Evolutionary Processors, in M. Fernández & I. Mackie, eds., Preliminary Proceedings of the First International Workshop on Developments in Computational Models: 15-24. Instituto Superior Tecnico, Lisboa.
34. G. Bel-Enguix & **M.D. Jiménez-López** (2005), LP-Colonies for Language Evolution: A Preview, in R. Freund, G. Lojka, M. Oswald & G. Paun, eds., Pre-Proceedings of the 6th Workshop on Membrane Computing: 179-192. Vienna University of Technology, Vienna.
35. R. Loos & **V. Mitrana** (2005), Non-Preserving Splicing with Delay, in Proceedings of DNA Based Computers, DNA11: 237-249.
36. **L.F. Mingo**, **J. Castellanos** & **F. Arroyo** (2005), Time Series Analysis with Axo-Axonic Connections in Neural Networks, in Proceedings of the 2nd International Conference on Information Technology: 246-254.

3.2.4 Book chapters

37. **C. Martín-Vide** & G. Paun (2004), Language Generating by means of Membrane Systems, in G. Paun, G. Rozenberg & A. Salomaa, eds., Current Trends in Theoretical Computer Science: The Challenge of the New Century, vol. II: Formal Models and Semantics: 599-611. World Scientific, Singapore.
38. **C. Martín-Vide**, A. Paun & G. Paun (2004), Membrane Computing: New Results, New Problems, in G. Paun, G. Rozenberg & A. Salomaa, eds., Current Trends in Theoretical Computer Science: The Challenge of the New Century, vol. II: Formal Models and Semantics: 613-623. World Scientific, Singapore.

39. **C. Martín-Vide** & G. Paun (2004), P Systems with Symport/Antiport Rules: A Survey, in G. Ciobanu & G. Rozenberg, eds., *Modelling in Molecular Biology*: 175-192. Springer, Berlin.
40. G. Bel-Enguix & **M.D. Jiménez-López** (2005), Linguistic Membrane Systems and Applications, in G. Ciobanu, G. Paun & M.J. Pérez- Jiménez, eds., *Applications of Membrane Computing*: 347-388. Springer, Berlin.
41. **C. Martín-Vide** & **V. Mitrana** (2005), Networks of Evolutionary Processors: Results and Perspectives, in M. Gheorghe, ed., *Molecular Computational Models: Unconventional Approaches*: 78-114. Idea Group, Hershey, PA.

3.2.5 Accepted papers in press

42. M.A. Grando & **V. Mitrana**, A Possible Connection between Two Theories: Grammar Systems and Concurrent Programming, *Fundamenta Informaticae*.
43. F. Manea, **C. Martín-Vide** & **V. Mitrana**, A Universal Accepting Hybrid Network of Evolutionary Processors, *Electronic Notes in Theoretical Computer Science*.
44. G. Bel-Enguix & **M.D. Jiménez-López**, Computing Dialogues with Membranes, *Electronic Notes in Theoretical Computer Science*.
45. **C. Martín-Vide** & **V. Mitrana**, Descriptive and Computational Aspects of Generating and Accepting Hybrid Networks of Evolutionary Processors, *Grammars*.
46. G. Bel-Enguix & **M.D. Jiménez-López**, Dynamic Meaning Membrane Systems: An Application to the Description of Semantic Change, *Fundamenta Informaticae*.
47. J. Dassow, **C. Martín-Vide** & **V. Mitrana**, Free Generating Hybrid Networks of Evolutionary Processors, in *Essays in Honour of Prof. Rani Siromoney*. World Scientific, Singapore.
48. K.S. Dersanambika, K. Krithivasan, **C. Martín-Vide** & K.G. Subramanian, Local and Recognizable Hexagonal Picture Languages, *International Journal of Pattern Recognition and Artificial Intelligence*.
49. P. Dömösi, **C. Martín-Vide** & A. Mateescu, On Polyslender Context-Free Languages, *Publicationes Mathematicae Debrecen*.
50. A. Choudhary, K. Krithivasan & **V. Mitrana**, Returning and Non-Returning PCFA Are Equivalent, *RAIRO - Theoretical Informatics and Applications*.
51. R. Gramatovici & **C. Martín-Vide**, Sorted Dependency Insertion Grammars, *Theoretical Computer Science*.
52. P. Bottoni, A. Labella, V. Manca & **V. Mitrana**, Superposition Based on Watson-Crick-like Complementarity, *Theory of Computing Systems*.
53. M.H. ter Beek, **C. Martín-Vide** & **V. Mitrana**, Synchronized Shuffles, *Theoretical Computer Science*.

3.2.6 Other publications less directly related to the subject of the research project: refereed journals

54. **P. García & J. Ruiz** (2004), A Note on Minimal Cover-Automata for Finite Languages, *Bulletin of the EATCS*, 83: 193-199.
55. A.H. Dediu & **C. Martín-Vide** (2004), Derived Trees Evolution for Tree Adjoining Grammars Parsing, *Grammars*, 7: 161-178.
56. **D. López, J.M. Sempere & P. García** (2004), Inference of Reversible Tree Languages, *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics*, 34(4): 1658-1665.
57. **P. García & J. Ruiz** (2004), Learning k-Testable and k-Piecewise Testable Languages from Positive Data, *Grammars*, 7(1): 125-140.
58. J. Dassow, **V. Mitrana**, G. Paun & R. Stiebe (2004), On Functions and Related Languages of Context-Free Grammars, *RAIRO - Theoretical Informatics and Applications*, 38: 257-267.
59. **A. Cano Gómez & J.-É. Pin** (2004), Shuffle on Positive Varieties of Languages, *Theoretical Computer Science*, 312(2-3): 433-461.
60. M.H. ter Beek, E. Csuhaaj-Varjú & **V. Mitrana** (2004), Teams of Pushdown Automata, *International Journal of Computer Mathematics*, 81.2: 141-156.
61. **V. Mitrana** (2005), Marcus External Contextual Grammars: From One to Many Dimensions, *Fundamenta Informaticae*, 54.1-4: 307-316.

3.2.7 Other publications less directly related to the subject of the research project: LNCS volumes

62. G. Bel-Enguix & **M.D. Jiménez-López** (2004), Computing Speech Acts, in C. Bussler & D. Fensel, eds., *Artificial Intelligence: Methodology, Systems, and Applications*, Lecture Notes in Computer Science 3192: 236-245. Springer, Berlin.
63. K.S. Dersanambika, K. Krithivasan, **C. Martín-Vide** & K.G. Subramanian (2004), Hexagonal Pattern Languages, in R. Klette & J.D. Zunic, eds., *Combinatorial Image Analysis*, Lecture Notes in Computer Science 3322: 52-64. Springer, Berlin.
64. J.A. Laxminarayana, **J.M. Sempere** & G. Nagaraja (2004), Learning Distinguishable Linear Grammars from Positive Data, in G. Paliouras & Y. Sakakibara, eds., *Grammatical Inference: Algorithms and Applications*, Lecture Notes in Computer Science 3264: 279-280. Springer, Berlin.
65. P. Dömösi, **C. Martín-Vide** & **V. Mitrana** (2004), Remarks on Sublanguages Consisting of Primitive Words of Slender Regular and Context-Free Languages, in J. Karhumäki, H. Maurer, G. Paun & G. Rozenberg, eds., *Theory is Forever: Essays Dedicated to Arto Salomaa on the Occasion of His 70th Birthday*, Lecture Notes in Computer Science 3113: 60-67. Springer, Berlin.

- 66. **A. Cano, J. Ruiz & P. García** (2004), Running FCRPNI in Efficient Time for Piecewise and Right Piecewise Testable Languages, in G. Paliouras & Y. Sakakibara, eds., *Grammatical Inference: Algorithms and Applications*, Lecture Notes in Computer Science 3264: 275-276. Springer, Berlin.
- 67. I. Perea & **D. López** (2004), Syntactic Modeling and Recognition of Document Images, in A. Fred, T. Caelli, R.P.W. Duin, A. Campilho & D. de Ridder, eds., *Structural, Syntactic, and Statistical Pattern Recognition*, Lecture Notes in Computer Science 3138: 416-424. Springer, Berlin.

3.2.8 Other publications less directly related to the subject of the research project: international conference proceedings

- 68. L. Cojocaru & **C. Martín-Vide** (2004), Parallel Communicating Pushdown Transducer Systems, in **J.M. Sempere & D. López**, eds., *Proceedings of the ECAI 2004 Workshop on Symbolic Networks*: 41-56. Universidad Politécnica de Valencia, Valencia.
- 69. **C. Martín-Vide & V. Mitrana** (2005), Further Properties of Path-Controlled Grammars, in *Proceedings of Formal Grammar/Mathematics of Language, FG-MOL10*: 139-150. CSLI, Stanford, CA.
- 70. **C. Martín-Vide & V. Mitrana** (2005), Remarks on Arbitrary Multiple Pattern Interpretations, in Z. Ésik & Z. Fülöp, eds., *Automata and Formal Languages*: 200-212. University of Szeged, Szeged.

3.2.9 Other publications less directly related to the subject of the research project: book chapters

- 71. **C. Martín-Vide & A. Mateescu** (2004), GSMs and Contexts, in G. Paun, G. Rozenberg & A. Salomaa, eds., *Current Trends in Theoretical Computer Science: The Challenge of the New Century*, vol. II: Formal Models and Semantics: 581-587. World Scientific, Singapore.
- 72. P.H. Chandra, K.G. Subramanian, **C. Martín-Vide**, D.L. Van & P.S.P. Wang (2005), Parallel Contextual Array Grammars with Trajectories, in C.H. Chen & P.S.P. Wang, eds., *Handbook of Pattern Recognition and Computer Vision*: 55-70. World Scientific, Singapore.

3.2.10 Other publications less directly related to the subject of the research project: books

- 73. **C. Martín-Vide, V. Mitrana & G. Paun**, eds. (2004), *Formal Languages and Applications*. Springer, Berlin.
- 74. **C. Martín-Vide, G. Mauri, G. Paun, G. Rozenberg & A. Salomaa**, eds. (2004), *Membrane Computing*, Lecture Notes in Computer Science 2933. Springer, Berlin.

3.2.11 Papers in progress

75. L.B. Goncharova, Y. Jacques, **C. Martín-Vide**, A.O. Tarakanov & J.I. Timmis, Biomolecular Immune-Computer: Theoretical Basis and Experimental Simulator.
76. P.H. Chandra, **C. Martín-Vide**, K.G. Subramanian, D.L. Van & P.S.P. Wang, Contextual Grammars for Pictorial Pattern Generation.
77. F. Manea, **C. Martín-Vide** & **V. Mitrana**, Descriptive Complexity of Some Universal Accepting Networks of Evolutionary Processors.
78. L. Pan & **C. Martín-Vide**, Further Remark on P Systems with Active Membranes and Two Polarizations.
79. E. Csuhaj-Varjú & **V. Mitrana**, Further Remarks on CD Pushdown Automata Systems.
80. J.D. Emerald, **C. Martín-Vide**, K.G. Subramanian & D.G. Thomas, Grammatical Inference of Some Subclasses of Contextual Grammars.
81. F. Manea, M. Margenstern & **V. Mitrana**, M.J. Pérez-Jiménez: New Characterizations of NP and PSPACE.
82. H. Bordihn, H. Fernau, M. Holzer, V. Manca & **C. Martín-Vide**, On Iterated Sequential Transducers.
83. A. Atanasiu, **C. Martín-Vide** & K.G. Subramanian, On Regular Array Grammars.
84. F. Manea, **C. Martín-Vide** & **V. Mitrana**, On the Complexity of Accepting Networks of Splicing Processors.
85. A. Alhazov, **C. Martín-Vide** & Y. Rogozhin, On the Number of Nodes in Universal Networks of Evolutionary Processors.
86. L. Pan & **C. Martín-Vide**, P Systems with Active Membranes, Two Polarizations, and Restrictions of Rules.
87. R. Gramatovici, **C. Martín-Vide** & **V. Mitrana**, Some Remarks on Partially Blind One-Way Multicounter Machines.
88. P. Leupold & **V. Mitrana**, Uniformly Bounded Duplication Codes.

3.3 Technology transfer (UPM)

In the field of neural networks, joint implementation work is being done with the Russian Academy of Sciences under the INTAS scheme. The software will be available in Spanish, English and Russian. A distributed implementation of that software is being considered as well. Its commercialization or free circulation among public institutions and private companies can come after its applications have been exhibited in international conferences.

In the field of NEPs, a software implementation will be made available with the same purpose. The hardware implementation can be considered too, but likely will have to wait until a future research project will be carried out.

3.4 Collaboration with national and international groups (URV)

Intensive teaching (through lecturing participation in the International PhD School in Formal Languages and Applications) and research collaborations (in all the projects of the research group, including NEPs) exist with the following institutes:

- Austria: Department of Discrete Mathematics and Geometry, **Technical University of Vienna** (W. Kuich)
- Canada: Department of Computer Science, **Carleton University** (J. Oommen); Department of Mathematics and Statistics, **University of Ottawa** (D. Sankoff); School of Computing, **Queen's University** (K. Salomaa); Department of Computer Science, University of Waterloo (J. Brzozowski) Department of Computer Science, **University of Western Ontario** (L. Ilie, S. Yu)
- Czech Republic: Department of Computer Science, **Masaryk University Brno** (J. Gruska); Department of Computer Science, **Silesian University** (J. Kelemen)
- Finland: Research Center, **Nokia** (V. Niemi); Department of Mathematics, **University of Turku** (T. Harju, J. Karhumäki, J. Kari, M. Steinby)
- France: Laboratory of Algorithmic Informatics: Foundations and Applications, **University Paris 7** (C. Choffrut, J.-É. Pin); Department of Computer Science, **University of Marne-la-Vallée** (M. Crochemore); Department of Computer Science, **University of Saint-Étienne** (C. de la Higuera)
- Germany: Department of Mathematics and Computer Science, **University of Bremen** (H.-J. Kreowski); Department of Computer Science, **University of Dortmund** (C. Moraga); Department of Computer Science, **Technical University of Dresden** (F. Baader); Department of Computer Science, **University of Frankfurt** (D. Wotschke); Department of Mathematics, **University of Freiburg** (J. Flum); Department of Computer Science, **University of Hamburg** (M. Kudlek); Department of Mathematics and Informatics, **University of Kassel** (F. Otto); Department of Computer Science, **University of Leipzig** (M. Droste); Department of Computer Science, **University of Magdeburg** (J. Dassow); Department of Computer Science, **Technical University of Munich** (M. Holzer); Department of Computer Science, **University of Potsdam** (H. Bordihn); Department of Computer Science, **University of Stuttgart** (H. Petersen)
- Hungary: Computer and Automation Research Institute, **Hungarian Academy of Sciences** (E. Csuhaj-Varjú); Department of Computer Science, **University of Szeged** (Z. Fülöp)
- India: Department of Mathematics, **Chennai Mathematical Institute** (R. Siromoney)
- Israel: Department of Computer Science, **University of Haifa** (S. Wintner)
- Italy: Department of Information Engineering, **University of Padua** (G. Satta)
- Japan: Department of Computer Science, **University of Electro-Communications** (S. Kobayashi); Department of Information Engineering, **Hiroshima University** (K. Morita); Department of Mathematics, **Kyoto Sangyo University** (M. Ito)

- The Netherlands: Department of Humanities Computing, **University of Groningen** (M.J. Nederhof); Leiden Institute of Advanced Computer Science, **Leiden University** (H.J. Hoogeboom, J. Kleijn); Institute of Linguistics, **Utrecht University** (M. Moortgat)
- Spain: Department of Computer Science, **Pompeu Fabra University** (R. Baeza-Yates)
- Taiwan: Department of Computer Science, **National Taiwan University** (H.-C. Yen)
- United Kingdom: Department of Computer Science, **University of Hertfordshire** (H. Fernau)
- United States: Department of Mathematical Sciences, **Binghamton University** (F. Guzmán, T. Head); Department of Computer Science, **University of California Santa Barbara** (O. Ibarra); Department of Computer Science, **Earlham College** (J. Rogers); Courant Institute of Mathematical Sciences, **New York University** (M. Mohri); Department of Computer Science, **University of Rochester** (M. Ogihara); Department of Computer Sciences, **University of Texas Austin** (R. Miikkulainen)

3.5 Joint activities organized

The following two events have been co-organized:

- Workshop on Symbolic Networks, in the framework of the European Conference on Artificial Intelligence (ECAI 2004, Valencia, 22 August 2004)
- Pre-Organized Session on Computing with Networks of Evolutionary Processors, in the framework of the 1st International Work-Conference on the Interplay between Natural and Artificial Computation (IWINAC-2005, Las Palmas de Gran Canaria, 15-18 June 2005)

Funding has been recently applied for from the European Science Foundation in order to organize an international symposium on NEPs in Tarragona in March 2006.

3.6 International projects

3.6.1 (URV)

In the field of the project, the following European Union projects are being developed at present:

- 2 Marie Curie MERG (reintegration) grants.
- 1 Marie Curie EIF (intra-European) grant.

Also, an application to the scheme NEST (New and Emerging Sciences and Technologies) on immune computing is currently under refereeing.

3.6.2 (UPV)

The group has become a member of the European Molecular Computing Consortium.

3.6.3 (UPM)

- INTAS. The following projects are being developed at present: INTAS INNO 182 (collaborating researchers), INTAS YSF 03-55-1696 (principal investigators), INTAS 2000-0626 (collaborating researchers). The topic of these projects is neural networks, data mining and their implementations.
- IST (Information Sciences and Technologies): IST-2001-32008 is being carried out on molecular computing (DNA and membranes). A new project has been recently applied for.

3.7 Summary of journals (+ LNCS) which the researchers of the project published in

Acta Informatica
 BioSystems
 Bulletin of the EATCS (2)
 Comparative and Functional Genomics
 Discrete Applied Mathematics
 Electronic Notes in Theoretical Computer Science (2)
 Fundamenta Informaticae (4)
 Grammars (3)
 IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics
 International Journal of Computer Mathematics
 International Journal of Pattern Recognition and Artificial Intelligence
 International Journal on Information Theories and Applications
 Journal of Universal Computer Science
 Lecture Notes in Computer Science (23)
 Publicationes Mathematicae Debrecen
 RAIRO - Theoretical Informatics and Applications (2)
 Theoretical Computer Science (6)
 Theory of Computing Systems
 WSEAS Transactions on Business and Economics

3.8 Summary of conferences which the researchers of the project contributed to

- Second Brainstorming Week on Membrane Computing (Sevilla, Spain, 2-7 February 2004) [URV, UPV]
- 10th International Meeting on DNA Computing (DNA 10, Milan, Italy, 7-10 June 2004) [URV]
- 5th Workshop on Membrane Computing (WMC 2004, Milan, Italy, 14-16 June 2004) [URV, UPV]
- Second International Conference on Information Research, Applications and Education (i.tech 2004, Varna, Bulgaria, 14-24 June 2004) [UPM]

- Grammar Systems Week 2004 (Budapest, 5-9 July 2004) [**URV**]
- Joint IAPR International Workshops on Structural and Syntactical Pattern Recognition (SSPR 2004) and Statistical Pattern Recognition (SPR 2004) (Lisboa, Portugal, 18-20 August 2004) [**URV**, **UPV**]
- ECAI 2004 Workshop on Symbolic Networks (Valencia, Spain, 22 August 2004) [**URV**]
- 11th International Conference on Artificial Intelligence. Methodology, Systems, and Applications (AIMSA 2004, Varna, Bulgaria, 2-4 September 2004) [**URV**]
- International Conference on Machines, Computations and Universality (MCU'2004, St. Petersburg, Russia, 21-24 September 2004) [**URV**]
- 7th International Colloquium on Grammatical Inference (ICGI 2004, Athens, Greece, 11-13 October 2004) [**UPV**]
- 5th Annual Spanish Bioinformatics Conference (Barcelona, Spain, 29-30 November 2004) [**UPV**]
- Tenth International Workshop on Combinatorial Image Analysis (IWCIA 2004, Auckland, New Zealand, 1-3 December 2004) [**URV**]
- Third Brainstorming Week on Membrane Computing (Sevilla, Spain, 31 January - 4 February 2005) [**UPV**]
- 2nd International Conference on Information Technology (ICIT 2005, Amman, Jordan, 3-5 May 2005) [**UPM**]
- Turing Days '05: DNA Computing (Istanbul, Turkey, 13-14 May 2005) [**URV**]
- 11th International Conference on Automata and Formal Languages (AFL'05, Dobogókő, Hungary, 17-20 May 2005) [**URV**]
- 11th International Meeting on DNA Computing (DNA 11, London, Canada, 6-9 June 2005) [**URV**]
- 8th International Work-Conference on Artificial Neural Networks: Computational Intelligence and Bioinspired Systems (IWANN2005, Vilanova i la Geltrú, Spain 8-10 June 2005) [**URV**]
- First Conference on Computability in Europe: New Computational Paradigms (CiE 2005, Amsterdam, The Netherlands, 8-12 June 2005) [**URV**]
- Pre-Organized Session on Computing with Networks of Evolutionary Processors, 1st International Work-Conference on the Interplay between Natural and Artificial Computation (IWINAC-2005, Las Palmas de Gran Canaria, Spain, 15-18 June 2005) [**URV**]
- 10th International Conference on Implementation and Application of Automata (CIAA 2005, Nice, France, June 27-29 2005) [**UPV**]

- Third International Conference on Information Research, Applications and Education (i.tech 2005, Varna, Bulgaria, 27-30 June 2005) [**UPM**]
- Fifth International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT 05, Paris, France, 5-8 July 2005) [**URV**]
- First International Workshop on Developments in Computational Models (DCM, Lisboa, Portugal, 10 July 2005) [**URV**]
- 6th Workshop on Membrane Computing (WMC 2005, Vienna, Austria, 18-21 July 2005) [**URV**, **UPV**]
- 3rd International Workshop on Logic and Communication in Multi-Agent Systems (LCMAS 2005), Nineteenth International Joint Conference on Artificial Intelligence (IJCAI-2005, Edinburgh, United Kingdom, 1 August 2005) [**URV**]
- 10th Conference on Formal Grammar and 9th Meeting on Mathematics of Language (FG-MOL 2005, Edinburgh, United Kingdom, 5-7 August 2005) [**URV**]
- 4th International Conference on Artificial Immune Systems (ICARIS-2005, Banff, Canada, 14-17 August 2005) [**URV**]
- VIIIth European Conference on Artificial Life (ECAL 2005, Canterbury, United Kingdom, 5-9 September 2005) [**URV**]
- International Conference on Artificial Neural Networks (ICANN 2005, Warsaw, Poland, 11-15 September 2005) [**URV**]
- First Conference on Algebraic Informatics (Thessaloniki, Greece, 20-23 October 2005) [**URV**]

GRID-IT: Investigación y Desarrollo de Servicios Grid - Aplicación a Modelos Cliente Servidor, Colaborativos y de Alta Productividad. TIC2003-01318

Vicente Hernández García *
Departamento de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia

Abstract

GRID-IT focus on easing the migration of applications to a Grid environment by means of the development of a set of components and services. These high-level tools will be extracted from a selection of applications that will be deployed during the project. The applications have been selected taking into account the experience of the research group involved in the project and the model of interaction of the software components. The project aims at covering the whole process of Grid application development, starting from the set up of a testbed, the development of application-oriented components and services on top of standard Grid middlewares and the implementation, validation and study of possible business models for several relevant applications.

Keywords: Grid Services, Components and Applications, DataGrids, High Throughput Computing, Client-Server.

1 Project objectives

The objective of the GRID-IT project is to develop a set of generic Grid services and components on the basis of the migration and implementation of a group of applications. The development considers different paradigms of interaction in a Grid environment. This set of components and services will form a framework to ease the further development of other Grid-based applications in different scientific and technological fields. Other objectives covered in the project include the deployment of a Grid infrastructure, the testing of state-of-the-art middlewares and the validation and study of business models of the developed applications. The schema shown in Figure 1 reflects the structure of the developments performed in the framework of the GRID-IT project. The different applications form the top level layer, under which the GRID-IT components (inferred from the applications) are integrated. These components

*Email: vhernand@dsic.upv.es

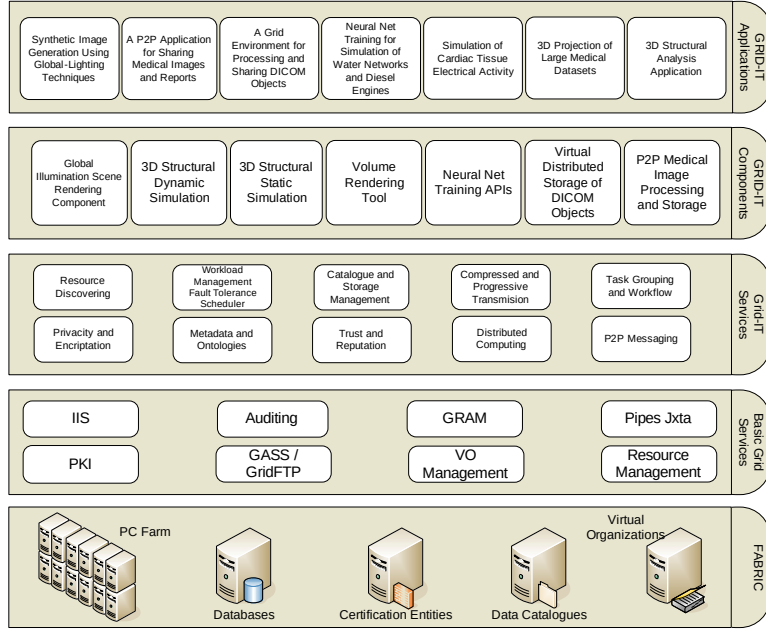


Figure 1: Structure of the GRID-IT project.

are supported by high-level services (the GRID-IT services), which use, in turn, a set of several services provided by the middleware. In some cases these GRID-IT services are re-engineered basic components of existing middlewares.

Some concepts have changed from the start of the project due to the natural evolution of Grid technologies during the last years, leading to an evolution of the project itself. As an example, the definition of OGSi (Open Grid Service Infrastructure), proposed in GT3 [1], has been deprecated and substituted by the specifications defined by the WSRF (Web Service Resource Framework) in GT4 [2, 3]. In the same way, the paradigms of interaction proposed initially (Client-Server, Collaborative and High Throughput) were reconsidered under this situation and the concept of the collaborative model, as it was defined originally, is nearest to the current concept of Datagrid.

The applications that are being developed under each one of the proposed models are the following:

1. **Client-Server Model:** This model includes those applications where the client (commonly a GUI) is a separate application from the server and the clients request from the server most of the information that they need. Two applications have been identified:
 - (a) 3D structural analysis.
 - (b) Synthetic image generation using of global-lighting techniques (radiosity).
2. **Datagrid Model:** This model comprises a wide set of applications that share data and provide a workbench for the execution of small distributed applications. Two activities have been identified:

- (a) A Peer-to-peer application for sharing medical images and reports.
 - (b) A Grid environment for processing and sharing DICOM objects.
3. **High Throughput Computing Model:** This model implies the execution of large sets of experiments on distributed computing nodes on the Grid. Efficiency, fault tolerance and robustness are the main requirements. The applications identified are:
- (a) Neural nets training for the simulation of water networks and diesel engines.
 - (b) 3D projection of large medical datasets.
 - (c) Simulation of cardiac tissue electrical activity.

2 Success Level Achieved in the Project

In the last two years, Europe has become the driving force for the evolution of Grid technologies through different initiatives, such as the UK e-Science Programme [5], some priorities in the Spanish R&D National Programme, or the same VI Framework Programme where the development of Grid technologies has been promoted by means of several important projects.

This trend and the synergy between GRID-IT and other national and international Grid projects where our research group is participating, have intensively promoted the development of the project and fostered the achievement of its objectives.

Figure 2 provides a quick view of the present status of the project. During the first 18 months the following milestones has been covered:

1. A Grid infrastructure has been deployed. At the moment, the operative middlewares are the following: Globus 2.4, Globus 3.9.5 (beta version of Globus 4), LCG 2.4.0 (Large Hadron Collider Computing Grid) and InnerGrid Nitya. The Grid fabric counts with approximately 350 processors, that include computers dedicated exclusively to the project, non-dedicated desktop computers and shared computing resources. During the second stage of the project, other middlewares oriented to services will be deployed, such as Globus 4 or gLite, an evolution of the LCG middleware developed in the Enabling Grids for E-sciencE project (EGEE [4]) where the research group is participating in the deployment of biomedical applications.
2. The design of the different pilot applications has been completed. Some applications have been deployed using Globus 2.4, and the migration to service-based middlewares, like Globus 4 or gLite, is in progress.
3. The GRID-IT components and services that will form the GRID-IT middleware have been identified. The GRID-IT services are those tools with a more general and horizontal conception which are applicable to a wide range of applications. On the other side, the GRID-IT components are modules that implement some functionalities common to a set of applications in different scientific and technological fields. The services and components identified, extrapolated from the design of the applications, are the following:
 - (a) GRID-IT Services:
 - i. Work Management:

- A. Workload management, fault tolerance and scheduler.
 - B. Task grouping and workflow.
 - C. Resource discovering (including multiprocessors).
- ii. Data Management:
 - A. Catalogue and storage management.
 - B. Metadata and ontologies.
 - C. Privacy and encryption.
 - D. Compressed and fault tolerant progressive transmission.
- iii. P2P Management:
 - A. Trust and reputation.
 - B. P2P messaging.
 - C. Distributed computing.
- (b) GRID-IT Components
 - i. 3D structural static simulation component.
 - ii. 3D structural dynamic simulation component.
 - iii. Scene rendering using global illumination methods.
 - iv. P2P medical image processing and storage.
 - v. Virtual distributed storage of DICOM objects.
 - vi. Neural net training APIs.
 - vii. Volume rendering tool.
- 4. The design of the interfaces of the different services and components has been completed. The integration of the components and services in the middleware and applications have been clearly defined and the interfaces and the functionality described.
- 5. The implementation of the selected generic Grid services has started. Several services are already available in beta or preliminary versions, such as the Workload Management and the cataloguing and transmission functionalities in Data Management. Those components will be improved in the second part of the project, especially in terms of robustness and efficiency.

The next subsection describes the different applications and components developed in the framework of the GRID-IT project. More information on the services can be obtained in the bibliography or contacting directly the project manager.

2.1 GRID-IT Applications and Components

This section describes the applications which are guiding the development of the project and the components stemmed from these applications. The applications are the base for the extrapolation of the components and services that conforms the main contribution of the GRID-IT project. The components are high-level objects that could be used for a wide set of applications. Not all the applications in the project have contributed in the same way to the GRID-IT middleware. Some of them have produced components or basic services and others both of them.

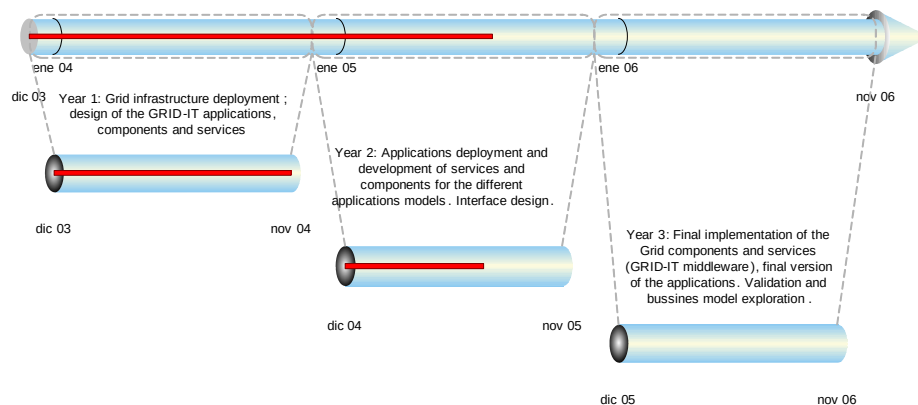


Figure 2: Planification and status of the GRID-IT project

- 3D structural analysis application:** The objective of this application is to perform the complete 3D structural simulation of a building. The current application is composed by a Grid structural calculation component that receives the structural data to be analysed (from an authenticated user) and using several of the basic services (GRID-IT services and middleware services) performs the simulations under the best conditions. The access to the application is made through a graphic interface that enables the user to access the Grid infrastructure transparently.

 - Generic component associated: Parallel computing-based static and dynamic 3D structural simulation components. These components perform the structural analysis of a building model under different load and sismic hypotheses.
- Synthetic image generation using global-lighting techniques (radiosity):** This application is included within a more general application of electronic commerce. The objective is to generate highly realistic synthetic images in a minimum time. The initial application used CORBA protocols to implement the communication with a cluster of PCs (where the parallel algorithm was running). In the framework of the GRID-IT project this application has been migrated into a Grid environment to benefit from features such as uniformity, transparency, reliability, ubiquity, security, etc. The core of the application is a rendering component that will choose the most suitable resources for the generation of the scene taking into account parameters such as the availability of resources, the priority of the job and the like.

 - Generic component associated: Global illumination scene rendering component. This Grid component allows the generation of a VRML scene (using different algorithms) from a given geometric data.
- A Peer-to-peer application for sharing medical images and reports:** This system follows a P2P architecture using the JXTA middleware as basic platform. The system publishes and searches for anonimized information, through a simple interface in a secure

virtual community of medical professionals. The search could be done by key words in the DICOM headers, using tools for content-like searching or by a framework of meta-data associated to the shared information. Moreover, the client incorporates some basic processing tools (zoom, panning, thresholding, etc.). The system is totally extendable to other communities where not only could the information be shared but also there would be the possibility of performing lightweight local processes. The application uses some basic services like those related with the compression and transference of data and other high level services.

- Generic component associated: P2P medical image storage and process. This component provides some tools to develop P2P collaborative applications such as distributed searching, eigen trust evaluation of a virtual community, security and privacy, etc.

4. **A Grid environment for processing and sharing DICOM objects:** This application consists in a Grid middleware that implements some components and services for the storage, processing and management of DICOM objects (images, structured documents, signals, etc.). The objective is to create a huge virtual repository of heterogeneous DICOM relevant to a community of users, with the possibility of associating structured meta-information data for performing searches through ontologies frameworks. The application makes use of some basic services related with data management. The architecture of the application is easily extendible to other problems even in other fields different from the medical one.

- Generic component associated: DICOM objects virtual storage component. This component defines a set of clients and servers that provide the needed tools for the creation of ontologies, virtual repositories, DICOM object sharing and tools for performing intelligent searches through the ontologies.

5. **Neural nets training for the simulation of water distribution and diesel engines:** Some problems, like the simulation of water networks or the contaminants emission of diesel engines, can be simulated by using Neural Nets. This approach, opposite to the use of physic models, is very useful when the computational time is an important factor, e.g. in real-time control. However a model based in Neural Nets must be correctly trained to obtain valuable results. The application developed under the GRID-IT project enables training a high number of Neural Nets increasing the productivity.

- Generic component associated: Neural Net multi-training component. This component aims to perform a set of Neural Net training tasks, taking into account different configuration parameters.

6. **3D projection of large medical datasets:** The objective of this application is to perform the volumetric reconstruction of CT (Computed Tomography) and MR (Magnetic Resonance) medical images. The system allows taking advantage of the distributed memory and the distributed computing capacity by reconstructing small pieces of the whole image in the distributed resources of the Grid. The generic rendering component of the application splits the whole data volume into small parts and distributes this information among the available resources.

- Generic component associated: Volume rendering component. This component provides the rendering of large image datasets by splitting the data in small pieces and processing this sub-volumes in the Grid nodes.
7. **Simulation of cardiac tissue electrical activity:** The simulation of the electrical activity of the heart is a process with high computational requirements. Typically, a case of study is composed by a set of parametric simulations completely independent that could be executed concurrently. The application, currently developed under GT2.4, allows to perform multiple parallel executions of a given problem by using some of the basic services developed in the framework of the project.

3 Results Indicators

The GRID-IT project has catalysed a very important scientific production. We have to emphasize the considerable number of publications related with the results obtained during the first 18 months of the project (see references [6] to [33]). Moreover, three research reports based on the topics of the project (inside the Parallel and Distributed Computing PhD programme of the Universidad Polit cnica de Valencia) have been presented by three researchers of the project to obtain the Advanced Studies Degree (DEA) in Parallel and Distributed Computing.

Apart from these scientific results, derived directly from the project results, other related effects have been motivated by the course of the project. In this sense the research group is involved in several national and international initiatives like the coordination of the creation of the **Metacentre of Supercomputing and Grid Technologies Applications of the Valencian Community** (Universidad Polit cnica de Valencia - CSIC - Universitat de Valencia), the participation in the PROFIT TIC project **gCitizen - Development of a Grid Middleware for the Open Management of the Administrative Information and Procedures in the Field of the Public Administration** (FIT-350101-2004-54), in the **Collaborative Science Portal** of Sun Microsystems and Fundaci n Telef nica, in the **EGEE**[4] project, etc. In the last call of the IST priority of the VI FP, the group has participated in a total of 3 proposals related with Grid technologies, coordinating one of them (an IP proposal called HealinG). At this moment, the group is involved in the preparation of several proposals also related with Grid technologies.

References

- [1] *Open Grid Services Infrastructure (OGSI) Version 1.0. Global Grid Forum.* GTuecke, S., Czajkowski, K., Foster, I., Frey, J., Graham, S., Kesselman, C., Maguire T., Sandholm, T., Snelling, D., and Vanderbilt, P. June 2003.
- [2] *From Open Grid Services Infrastructure to WSResource Framework: Refactoring & Evolution.* Karl Czajkowski, Don Ferguson, Ian Foster, Jeff Frey, Steve Graham, Tom Maguire, David Snelling, Steve Tuecke. 2004.
- [3] <http://www.globus.org/toolkit/docs/4.0/>
- [4] *Enabling Grids for E-Science.* <http://www.eu-egee.org>.

- [5] <http://www.rcuk.ac.uk/escience/>.
- [6] *A Grid-Enabled Scene Rendering Application*. M. Caballer, V. Hernández y J. E. Román. International Conference on Computational Science 2004 (ICCS 2004). LNCS 3038. (ISBN 3-540-22116-6). Ed. Springer Verlag. 54 - 57. 2004.
- [7] *Parallel and Grid Computing in 3D Analysis of Large Dimension Structural Systems*. José M. Alonso, Carlos de Alfonso, Gabriel Garcia, Vicente Hernandez. Euro-Par 2004 Parallel Processing, 10th International Euro-Par Conference. Proceedings Series: Lecture Notes in Computer Science, Vol. 3149. ISBN 3-540-22924-8. 2004.
- [8] *Integrating HPC and Grid Computing for 3D Structural Analysis of Large Buildings*. J.M. Alonso, C. Alfonso, G. Garcia and V. Hernandez. Proceedings of the Fourth International Conference on Engineering Computational Technology, Civil-Comp Press, 2004. Paper 92. ISBN 0-948749-98-9. 2004.
- [9] *Desarrollo de una Aplicación de Alta Productividad para el Cálculo Estructural de Edificios de Gran Dimensión*. J. M. Alonso, C. de Alfonso, G. García, V. Hernández. Actas de las XV Jornadas de Paralelismo, Universidad de Almería, págs. 277-282. ISBN: 84-8240-714-7. 2004.
- [10] *Desarrollo de un Servicio Grid para el Análisis 3D de Estructuras de Edificación*. J. M. Alonso, C. de Alfonso, V. Hernández. Anales de Ingeniería Mecánica. Revista de la Asociación Española de Ingeniería Mecánica. Año 15. Volumen No. 4. ISSN: 0212-5072. ISSN (Edición Digital) 1698 - 5990 Páginas, inicial: 2387 final: 2395. Diciembre 2004. León.
- [11] *Minimizando los Tiempos de Análisis 3D de Estructuras de Edificación de Gran Dimensión*. J. M. Alonso, G. García, V. Hernández. Anales de Ingeniería Mecánica. Revista de la Asociación Española de Ingeniería Mecánica. Año 15. Volumen No. 4. ISSN: 0212-5072. ISSN (Edición Digital) 1698-5990. Páginas 2407-2416. Diciembre 2004. León.
- [12] *Enabling High Level Access to Grid Computing Services*. J.M. Alonso, V. Hernández, G. Moltó. European Research Consortium for Informatics and Mathematics (ERCIM). ERCIM News. Special: Grids: The Next Generation, ISSN: 0926-4981. Volumen: 59. Páginas: 42-43. Año: 2004.
- [13] *Servicios ASP para el Proceso y Transmisión de Imágenes Médicas Tomográficas*. I.Blanquer, V. Hernandez, D. Segrelles. VII Congreso Nacional de Informàtica de la Salud (Inforsalud) ISBN 84-930487-8-X. 2004.
- [14] *A Middleware Grid for Storing, Retrieving and Processing DICOM Medical Images*. Ignacio Blanquer, Vicente Hernandez, Ferran Mas, Damià Segrelles. Workshop on Distributed Databases and Processing in Medical Image Computing (DIDAMIC). Working Notes of the Workshop on Distributed Databases and Processing in Medical Image Computing. Rennes (Francia) 2004.
- [15] *A P2P Platform for Sharing Radiological Images and Diagnoses*. Ignacio Blanquer, Vicente Hernández, Ferran Mas. DiDaMIC'04 (Rennes, Francia). 2004.

- [16] *Empleo de Técnicas Computacionales Avanzadas para la Aceleración de las Simulaciones Cardiacas*. J.M. Alonso, J.M. Ferrero (Jr.), V. Hernández, G. Moltó, M. Monserrat, J. Saiz. Comité organizador del CASEIB 2004. Libro de Actas del XXII Congreso Anual de la Sociedad Española de Ingeniería Biomédica. ISBN 84-688-9318-8. Páginas: 267-270. 2004.
- [17] *Globus-Based Grid Computing Simulations of Action Potential Propagation on Cardiac Tissues*. José M. Alonso, Vicente Hernández, Germán Moltó. Proceedings of the Euro-Par 2004 Conference. Lecture Notes in Computer Science. ISBN 3-540-22924-8, ISSN 0302-9743. Volumen: 3149. Páginas: 444-451. Año: 2004.
- [18] *Computación de Altas Prestaciones sobre Entornos Grid en la Simulación Eléctrica de Tejidos Cardiacos*. J.M. Alonso, J.M. Ferrero (Jr.), V. Hernández, G. Moltó, M. Monserrat, J. Saiz. Libro de Actas de las XV Jornadas de Paralelismo, ISBN: 84-8240-714-7. Páginas: 271-276. Año: 2004.
- [19] *Biomedical and Civil Engineering Experiences Using Grid Computing Technologies*. J.M. Alonso, C. de Alfonso, V. Hernández, G. Moltó. Proceedings of the Parallel Computing (ParCo 2005) Conference. 2005.
- [20] *Clinical Decision Support Systems (CDSS) in Grid Environments*. Ignacio Blanquer, Vicente Hernandez, Damià Segrelles, Montserrat Robles, Juan Miguel Garcia, Javier Vicente Robledo. From Grid to HealthGrid IOS Press ISBN 1-58603-510-X. 2005.
- [21] *A Grid-based Application for the 3D Structural Dynamic Simulation of High-Rise Buildings*. J. M. Alonso, C. de Alfonso, V. Hernández, G. Moltó. Proceedings of the Tenth International Conference on Civil, Structural and Environmental Engineering Computing. 2005.
- [22] *Creating Virtual Storages and Searching DICOM Medical Images through a Grid Middleware based in OGSA*. Ignacio Blanquer, Vicente Hernández, Damià Segrelles. Internacional Symposium on Cluster Computing and Grid (CCGRID05). ISBN 0-7803-9075-X. 2005.
- [23] *A Web-Based Application Service Provision Architecture for Enabling High-Performance Image Processing*. Carlos de Alfonso, Ignacio Blanquer, Vicente Hernandez, Damia Segrelles. Lecture Notes in Computer Science, ISBN 3-540-25424-2. ISSN: 0302-9743. 3402, 260-273. 2005.
- [24] *Large Medical Datasets on the Grid*. Carlos de Alfonso, Ignacio Blanquer, Vicente Hernández. Methods of Information in Medicine 2005; 44: 172-176. ISSN 0026-1270. 2005.
- [25] *The Grid as a Healthcare Provision Tool*. Vicente Hernández, Ignacio Blanquer. Methods of Information in Medicine 2005; 44: 144-148. ISSN 0026-1270. 2005.
- [26] *Three-Dimensional Cardiac Electrical Activity Simulation on Cluster and Grid Platforms*. J.M. Alonso, J.M. Ferrero (Jr.), V. Hernández, G. Moltó, M. Monserrat, J. Saiz. Springer-Verlag Lecture Notes in Computer Science. ISBN 3-540-25424-2, ISSN 0302-9743. Volumen: 3402. Páginas: 219-232. Año: 2005.

- [27] *An Object-Oriented View of Grid Computing Technologies to Abstract Remote Task Execution*. J.M. Alonso, V. Hernández, G Molto. IEEE CS Press Proceedings of the 13th Euromicro Conference on Parallel, Distributed and Network-based Processing. ISBN 0-7695-2280-7, ISSN 1066-6192. Páginas: 235-242. Año: 2005
- [28] *A Peer-to-Peer Environment to Share Medical Images and Diagnoses Providing Context-Based Searching*. Ignacio Blanquer, Vicente Hernández, Ferran Mas. 13th Euromicro Conference on Parallel, Distributed and Network-based Processing. Lugano (Suiza). 2005.
- [29] *Experiences on a Large Scale Grid Deployment with a Computationally Intensive Biomedical Application*. J.M. Alonso, V. Hernández, G Molto. 18th IEEE International Symposium on Computer-Based Medical Systems (CBMS 2005). Special Track: Grids for Biomedicine and Bioinformatics.
- [30] *Computación Paralela y Tecnologías en el Cálculo Dinámico de Estructuras de Edificación*. J.M. Alonso, V. Hernández, G Molto. Proceedings of the 12th International Congress on Computer Science Research (CHCC05). Pendiente de aceptación.
- [31] *Three-dimensional Structural Dynamic Analysis using Parallel Direct Time Integration Methods*. J.M. Alonso, V. Hernández. Proceedings of the Tenth International Conference on Civil, Structural and Environmental Engineering Computing (CC2005). 2005.
- [32] *A Grid-based Application of the 3D Dynamic Analysis of High-Rise Buildings*. J. M. Alonso, C. de Alfonso, V. Hernández, G. Moltó. Proceedings of the The Tenth International Conference on Civil, Structural and Environmental Engineering Computing (CC 2005). 2005. (Aceptado, pendiente de publicación).
- [33] *Grid Technology for Structural Analysis*. J. M. Alonso, C. Alfonso, G. García. V. Hernández. Computers & Structures. 2005. (Pendiente de aceptación).

Hybridizing Evolutionary Computation and Ant Colony Optimization. Applications to Fuzzy Rule Learning and Bioinformatics Problems.

TIC2003-00877

Oscar Cordon *

Dept. of Computer Science and Artificial Intelligence. E.T.S.I. Informática
University of Granada. 18071-Granada. Spain

Abstract

The goal of this research project is to analyze the possibilities of hybridization of different metaheuristics, namely Evolutionary Computation and Ant Colony Optimization, as well as other Soft Computing techniques such as fuzzy logic and fuzzy clustering. We aim at proposing new hybrid algorithms to solve some real-world applications, such as fuzzy rule learning and several Bioinformatics problems. More specifically, we will analyze the different components existing in the isolated metaheuristics to respectively explore and exploit the search space. This way, we will be able to develop hybrid algorithms combining complete instances of the individual metaheuristics, on the one hand, and to identify a set of exploration and exploitation components (useful for the search diversification and intensification, respectively), that can be appropriately combined in order to obtain better search and optimization algorithms, on the other hand.

Keywords: Metaheuristics, Evolutionary Computation, Ant Colony Optimization, Hybridization of Metaheuristics, Multiobjective and Multimodal Evolutionary Algorithms, Fuzzy Rule-Based Systems, Fuzzy Clustering, Bioinformatics.

1 Research project goals

The goal of the project is to study the chances for the hybridization of different metaheuristics [4], specifically, Evolutionary Computation (EC) [1] and Ant Colony Optimization (ACO) [2], proposing new hybrid algorithms with a better trade-off between search space exploration and exploitation, as well as between accuracy and efficiency, to get better solutions for both classical combinatorial optimization problems and real-world applications such as fuzzy modeling and Bioinformatics.

This global goal can be divided into five specific objectives:

1. Study of the different instances of EC and ACO, as well as other metaheuristics, and identification of their exploration and exploitation components.

*Email: ocordon@decsai.ugr.es

2. Proposal of specific hybrid algorithms from the study developed in the previous objective. Taking the theoretical and practical study of the comparison of the analyzed metaheuristics as a base, we aim at proposing second generation hybrid search and optimization algorithms allowing us to improve the results obtained by the isolated algorithms in classical combinatorial optimization problems. Namely, our aim is to work in three specific research lines:
 - (a) Incorporation of new evolutionary algorithm (EA) diversification components into ACO algorithms.
 - (b) Integration of parallel EA components into ACO algorithms.
 - (c) Integration of multiobjective and niching techniques into ACO algorithms.
3. Design of hybrid algorithms between EAs and ACO for the inductive learning of fuzzy rule-based system (FRBS) knowledge bases in real-world applications. This objective collects two different research lines:
 - (a) Automatic learning of fuzzy knowledge bases by means of the EC-ACO hybrid algorithms designed.
 - (b) Automatic learning of fuzzy knowledge bases by evolutionary meta-learning algorithms for the data base definition, and ACO algorithms for the rule base derivation.
4. Design of hybrid algorithms between EAs and fuzzy clustering, as well as between EAs and ACO to integrate genetic information of different kinds, obtained from different sources, for the modeling and prediction of genetic regulatory networks. This objective involves the following stages:
 - (a) Design of hybrid algorithms (multiobjective and multimodal EAs, hybridized with fuzzy clustering) for pattern discovery to detect protein regulatory sites and DNA sequence promoters.
 - (b) Design of inference algorithms of causal connections among groups of co-expressed and co-regulated genes between EC-ACO hybrid algorithms devoted to the genetic network derivation from DNA sequence analysis and oligonucleotide microarrays. Synthesis and description of the inferred pathways by means of FRBSs.
5. Validation of the proposed hybrid algorithms in the mentioned application areas.

The estimated schedule for the achievement of the previous goals is shown in Table 1. Notice that, each activity is associated with the corresponding described objective. Besides, its starting and end month is also reported. We should remind that the project is to be developed in a 36 month time.

2 Successful achievements of the project

The work developed till now in the research project mainly corresponds to the second, third and fourth objectives described in the previous section. The next three subsections are devoted to describe the results obtained in each of these three areas, together with the publications generated from them.

Table 1: Estimated research project schedule

Activities/Tasks	Start	End
O1. Study of existing bibliography	m1	m4
O1. Analysis of the exploration and exploitation components of EAs and ACO algorithms	m5	m8
O2.1. Hybridization of ACO algorithms and EA specific components	m9	m12
O2.2. Design of parallel ACO algorithms based on parallel EAs	m13	m16
O2.3. Integration of multiobjective and multimodal techniques into ACO	m17	m20
O3.1. Automatic learning of fuzzy knowledge bases by means of EC-ACO hybrids	m9	m22
O3.2. Automatic learning of fuzzy knowledge bases by meta-learning EC-ACO hybrids	m23	m28
O4.1. Design of multiobjective and multimodal EAs-fuzzy clustering hybrid algorithms to detect protein regulatory sites and DNA sequence promoters	m9	m22
O4.2. Design of inference algorithms of causal connections among groups of co-expressed and co-regulated genes between EC-ACO hybrid algorithms	m23	m28
O5. Validation of the proposed hybrid algorithms in the application areas considered	m29	m36

2.1 Proposal of specific hybrid algorithms between EAs and ACO

Three research lines have been developed within this first area, corresponding to the three subobjectives distinguished in the previous section.

2.1.1 Incorporation of new EA diversification components into ACO algorithms

We have gone on working with our proposal of the Best-Worst Ant System (BWAS), ACO algorithm which incorporates EA components to improve its performance. The relation between the three distinguished components of the algorithm has been analyzed in deep, in order to determine that they three are actually needed in order to get a proper performance. As a result of this work, two publications have been developed, corresponding to two book chapters:

1. S. Alonso, O. Cordon, I. Fernández de Viana, F. Herrera, La Metaheurística de Optimización Basada en Colonias de Hormigas: Modelos y Nuevos Enfoques. G. Joya, M.A. Atencia, A. Ochoa, S. Allende (Eds.), Optimizacion Inteligente: Técnicas de Inteligencia Computacional para Optimización, Servicio de Publicaciones de la Universidad de Málaga (2004) pp. 261-313.
2. S. Alonso, O. Cordon, I. Fernández de Viana, F. Herrera, Integrating Evolutionary Computation Components in Ant Colony Optimization Evolutionary Algorithms: An Experimental Study. Recent Developments in Biologically Inspired Computing, L. Nunes de Castro, F.J. Von Zuben (Eds.), Idea Group Publishing (2004) pp. 148-180.

On the other hand, the incorporation of Dr. Damas to the project team, opened new research lines within this area. Dr. Damas' area of expertise is the application of metaheuristics (mainly, EAs) to a real-world problem, the registration of 3D medical images, on which he made his PhD two years ago, advised by Dr. Cordon. Since 3D image registration can be formulated both as a combinatorial optimization problem and as a numerical optimization problem, it has become a very interesting benchmark for the application of hybrid metaheuristics. More specifically, up to now in this project we have worked with EAs establishing an appropriate trade-off between intensification and diversification, such as CHC [3], as well as with hybrid techniques considering the use of local search optimizers, namely Scatter Search

[5] and Iterated Local Search [6] paradigms. Seven accepted (two national conference contributions, four international ones, and one international journal paper) and three publications currently submitted to international journals have been originated from this work:

1. O. Cordón, S. Damas, J. Santamaría, Registrado de Imágenes 3D Mediante CHC, Actas del Tercer Congreso Español de Metaheurísticas, Algoritmos Evolutivos y Bioinspirados (MAEB'04), Córdoba (España) (2004) pp. 91-97.
2. O. Cordón, S. Damas, J. Santamaría, R. Martí, Una Aplicación de Scatter Search al Problema del Registrado de Imágenes Médicas, Actas del Cuarto Congreso Español de Metaheurísticas, Algoritmos Evolutivos y Bioinspirados (MAEB'05), Granada (España) (2005). In press.
3. O. Cordón, S. Damas, J. Santamaría, A Scatter Search Algorithm for the 3D Image Registration Problem, 8th International Conference on Parallel Problem Solving from Nature (PPSN VIII), Lecture Notes in Computer Science 3242, Birmingham (UK) (2004) pp. 471-480.
4. O. Cordón, S. Damas, J. Santamaría, Solving the 3D Image Registration Problem by Scatter Search with Path-Relinking, Proceedings of INFORMS Annual Meeting, Denver (USA) (2004) pp. 113.
5. O. Cordón, S. Damas, J. Santamaría, A Scatter Search-based Optimizer for the Registration of 3D Surfaces, Proc. IEEE Congress on Evolutionary Computation (CEC), Edinburgh (UK) (2005). In press.
6. O. Cordón, S. Damas, J. Santamaría, R. Martí, 3D Inter-Subject Medical Image Registration by Scatter Search, Proceedings of the Second International Workshop on Hybrid Metaheuristics (HM2005), Lecture Notes on Computer Science 3636, Barcelona (España) (2005) pp. 90-103.
7. O. Cordón, S. Damas, Image Registration with Iterated Local Search, Journal of Heuristics (2005). In press.
8. O., Cordón, S. Damas, J. Santamaría, R. Martí, Scatter Search for the 3D Point Matching Problem in Image Registration, Submitted to INFORMS Journal on Computing (June 2005).
9. O. Cordón, S. Damas, J. Santamaría, A Fast and Accurate Approach for 3D Image Registration using the Scatter Search Evolutionary Algorithm, Submitted to Pattern Recognition Letters (November 2004).
10. O. Cordón, S. Damas, J. Santamaría, Feature-based Image Registration by means of the CHC Evolutionary Algorithm, Submitted to Image Vision and Computing (June 2004).

We should remark that, the availability of a real-world optimization problem will allow us to study the performance of the EC-ACO hybrid techniques developed in the first stages of the research project in a significantly much more realistic way than when working with the classical traveling salesman problem (TSP) and quadratic assignment problem (those initially proposed to be considered in the project). This is one of the first tasks to be done in the second part of the project.

2.1.2 Integration of parallel EA components into ACO algorithms

The second part of the second publication mentioned in the previous subsection, was devoted to the proposal of different parallel ACO algorithms based on the usual parallelization strategies followed in EC. However, the results obtained were not promising at all. As a consequence, we decided to give up with this working line to focus more in the remaining project research lines.

2.1.3 Integration of multiobjective and niching techniques into ACO algorithms

We have specially focused on the design of multiobjective ACO algorithms. A review of the existing bibliography on the topic has been developed, finding a large number of existing proposals (around fifteen) of multiobjective ACO algorithms, but applied to very different problems. From them, those based on the Pareto-optimality definition have been selected as the most representative, resulting in ten different algorithms, which have been adapted to solve

the same benchmark, the bi-criteria TSP. Besides, their performance has been compared to two state-of-the-art multiobjective EAs, SPEA2 and NSGA-II. This way, a realistic comparison between the different approaches has been established for the very first time. Our conclusion is that some multiobjective ACO algorithms outperform SPEA2 and NSGA-II when no local optimizer is considered, but the latter EAs present the best performance when every algorithm is hybridized with a local search procedure. The results of the work developed comprise an international conference contribution and a paper submitted to an international journal:

1. C. García-Martínez, O. Cordon, F. Herrera, An Empirical Analysis of Multiple Objective Ant Colony Optimization Algorithms for the Bi-criteria TSP. 4th International Workshop, ANTS2004, Brussels (Bélgica). M. Dorigo y otros (Eds.), LNCS 3172 (2004) pp. 61-72.
2. C. García-Martínez, O. Cordon, F. Herrera, A Taxonomy and an Empirical Analysis of Multiple Objective Ant Colony Optimization Algorithms for the Bi-criteria TSP. Submitted to European Journal of Operational Research (August 2004).

2.2 Design of EC-ACO hybrid algorithms for the inductive learning of FRBS knowledge bases in real-world applications

The two research lines developed in this second area are described in the next two subsections.

2.2.1 Automatic learning of fuzzy knowledge bases by EC-ACO hybrid algorithms

The project team has worked on the design of FRBSs by the learning of their rule bases using the BWAS algorithm. Very promising results have been obtained both in fuzzy modeling applications (related to electrical distribution problems) and in mobile robotics. The latter application is very interesting as the interpretability of the generated knowledge base is an important requirement. This research line has concluded in four publications: a national conference contribution, an international conference one, a paper recently published in an international journal, and another one submitted to another journal:

1. M. Mucientes, J. Casillas, Generación eficiente de un controlador difuso con alta interpretabilidad para navegación de robots móviles, Actas del XII Congreso Español sobre Tecnología y Lógica Fuzzy (ESTYLF 2004), Jaén (2004) pp. 401-406.
2. M. Mucientes, J. Casillas, Obtaining a fuzzy controller with high interpretability in mobile robots navigation, Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2004), Budapest, Hungary (2004) pp. 1637-1642.
3. J. Casillas, O. Cordon, I. Fernández de Viana, F. Herrera, Learning Cooperative Linguistic Fuzzy Rules using the Best-worst Ant System Algorithm, Intl. Journal of Intelligent Systems 20 (2005) pp. 433-452.
4. M. Mucientes, J. Casillas, Fast Learning of Fuzzy Controllers with High Interpretability in Mobile Robotics. Submitted to IEEE Transactions on Fuzzy Systems (2005).

2.2.2 Automatic learning of fuzzy knowledge bases by means of meta-learning EC-ACO hybrid algorithms

On the one hand, we have worked both on the hybrid learning of fuzzy knowledge bases by means of the said methodology (a wrapping EA is considered to learn the FRBS data base, while the BWAS is applied to derive the fuzzy rules for each possible data base definition). On the other hand, the later adjustment of those knowledge bases using EAs has also been considered by either tuning the membership functions and selecting cooperative rule sets, or

jointly performing both tasks. Three publications have been obtained in this line: a national conference, an international conference, and an international journal paper:

1. J. Casillas, O. Cordón, F. Herrera, P. Villar, Aprendizaje Híbrido de la Base de Conocimiento de un Sistema Basado en Reglas Difusas mediante Algoritmos Genéticos y Colonias de Hormigas, XII Congreso Español sobre Tecnologías y Lógica Fuzzy (ESTYLF2004), Jaén (2004) pp. 357-362.
2. J. Casillas, O. Cordón, F. Herrera, P. Villar, A Hybrid Learning Process for the Knowledge Base of a Fuzzy Rule-Based System, Proc. Tenth International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU2004), Perugia, Italy (2004) pp. 2189-2196.
3. J. Casillas, O. Cordón, M.J. Del Jesus, F. Herrera, Genetic Tuning of Fuzzy Rule Deep Structures Preserving Interpretability and its Interaction with Fuzzy Rule Set Reduction, IEEE Transactions on Fuzzy Systems 13:1 (2005) pp. 13-29.

2.3 Design of EAs-fuzzy clustering and EC-ACO hybrid algorithms to integrate genetic information for the modeling and prediction of genetic regulatory networks

Finally, in this third area, the following two working lines are being developed.

2.3.1 Design of hybrid algorithms for pattern discovery to detect protein regulatory sites and DNA sequence promoters

To do so, we have considered the hybridization of the knowledge representation capability of fuzzy logic, used to model the uncertainty existing in the biological patterns to be uncovered, and the search ability of multiobjective and multimodal EAs, considered to extract a non dominated solution set for the problem. As a result of this work, four publications have been developed: an international conference contribution, an international journal paper, and two papers submitted to other two international journals:

1. R. Romero-Zaliz, O. Cordón, C. Rubio, I. Zwir, A Multiobjective Evolutionary Fuzzy System for Promoter Discovery in E. coli. Proceedings of the First International Workshop on Genetic Fuzzy Systems, Granada (2005) pp. 68-75.
2. V. Cotik, R. Romero-Zaliz, I. Zwir, A Hybrid Promoter Analysis Methodology for Prokaryotic Genomes, Fuzzy Sets and Systems 152 (2005) pp. 83-102.
3. I. Zwir, R. Romero-Zaliz, H. Huang, E.A. Groisman, Extracting Promoter Features from Genomic Datasets: Towards an Annotation of Genome Regulatory Regions. Submitted to Genome Research (May 2005).
4. I. Zwir, H. Huang, E.A. Groisman, Analysis of Differentially-Regulated Genes within a Regulatory Network by GPS Genome Navigation. Submitted to Bioinformatics (2005).

2.3.2 Design of EC-ACO hybrid inference algorithms of causal connections among groups of co-expressed genes for genetic network derivation

As this task is scheduled for the twenty third month of the project, the work developed on it is only preliminary. Specifically, we have worked on the selection of the most complete and biologically meaningful set of genes which become differentially expressed between control and treatment temporal microarray experiments in a real-world problem studying the host response over time to systemic inflammatory insults. Different statistical approaches have been applied to retrieve differentially expressed genes over treatment and time. These gene expression profiles are the seed from which we will be able to build genetic networks, that will

be validated by means of additional data taken from databases such as Gene Ontology, as well as with clinical data. By now, we have obtained an international conference publication in this line, as well as two papers submitted to international journals:

1. C. Rubio, O. Cordon, I. Zwir, Identifying Meaningful Temporal Gene Expression Patterns in the Inflammation and the Host Response to Injury. Proc. Affymetrix User Group Meeting, Edinburgh (2004).
2. C. Rubio-Escudero, R. Romero-Zaliz, O. Cordon, J.P. Cobb, I. Zwir, Identifying Gene Profiles by Reverse Problem Solving: From Grouping Gene Expressions to Combining Microarray Analysis Methods. Submitted to Bioinformatics (July, 2005).
3. R. Romero-Zaliz, O. Cordon, C. Rubio-Escudero, J.P. Cobb, I. Zwir, A Multi-objective Evolutionary Conceptual Clustering Methodology for Gene Annotation from Network Databases. Submitted to Journal of Machine Learning Research (July, 2005).

3 Indices to measure the quality of the obtained results

In summary, a total number of 28 publications has been developed in this first half of the research project, distributed as follows: 4 international journal papers, 9 papers submitted to international journals, 2 book chapters, and 13 (4 national and 9 international) conference contributions. All these publications are available in the web page of the Research Group “Soft Computing and Intelligent Information Systems” (SCI2S), to which the project team members belong (<http://sci2s.ugr.es>).

On the other hand, as regards the staff under formation, the project team members are advising three different PhD dissertations in the project topics:

- The PhD dissertation of Ms. Rocío C. Romero Zaliz, advised by Drs. Cordon and Zwir, has been finished yet, and will be presented next September, 14th, in the E.T.S.I. Informática of the University of Granada. It is entitled “Reconocimiento de Perfiles de Regulación Genética mediante Algoritmos Evolutivos Multiobjetivo” and is directly related to the objective 4.1 of the project, described in Section 1.
- Ms. Cristina Rubio, whose incorporation to the project team has been recently asked, is also being advised by Drs. Cordon and Zwir in the other project objective related to Bioinformatics, namely objective 4.2. Her work is related to biologically significant gene selection and genetic networks design for the systemic inflammatory insults response problem described in Section 2.3.2.
- Finally, Mr. Jose Santamaría is developing his PhD dissertation under the advise of Drs. Cordon and Damas on the application of hybrid metaheuristics to the 3D medical image registration problem described in Section 2.1.1 (project objective 2.1). The most of the publications mentioned in that section as regards this topic correspond to his PhD work.

We should also notice that we have got an additional grant from the Spanish Ministry related to the Bioinformatics research lines developed in the current project. It is within the “Acciones Complementarias del Programa Nacional de Cooperación Internacional en Ciencia y Tecnología del Ministerio de Educación y Ciencia” Action, in the “a7) Ayudas para el establecimiento de nuevas relaciones científicas en el ámbito de las prioridades geográficas de cooperación consignadas en el Plan Nacional” modality, and is entitled “Procesos de Regulación Genética sobre Microarrays usando Extracción de Conocimiento” (Ref. BIO2004-270-E). Its

aim is to encourage the Bioinformatics research lines with some US research groups on the topic. Two research project team members, Drs. Cerdón and Zwir, with the latter being the Grant Coordinator, as well as Ms. Cristina Rubio, whose incorporation to the project team has been recently asked, participate in the grant. We have started to work with the said research groups and four publications submitted to international journals have been already developed with them (see Sections 2.3.1 and 2.3.2).

Moreover, some collaborations with other research groups have arisen from the project work. The global project scope, the study of metaheuristics and their applications, is directly related to one of the Research Networks funded by the Spanish Ministry, the Spanish Network on Metaheuristic Procedures (HEUR) (Ref. TIC2002-10886E). The members of the project team are also members of the Network, which has generated different collaborations with other Spanish research groups. Namely, as regards the current project, we have worked with the Network Coordinator, Dr. Rafael Martí, from the University of Valencia. From this joint work, the following results have been obtained:

- Organization of a special session on “Scatter Search” in the Fourth Spanish Conference on Metaheuristics and Evolutionary and Bioinspired Algorithms (MAEB2005), within the First Spanish Computer Science Conference (CEDI2005), that will be held in Granada next September 2005.
- Development of three joint publications: a contribution to the previous session (national conference), an international conference contribution, and a paper submitted to an international journal, *INFORMS Journal on Computing*. All of them have been already mentioned in Section 2.

References

- [1] T. Bäck, D.B. Fogel, Z. Michalewicz (Eds.), *Handbook of Evolutionary Computation*, IOP Publishing Ltd and Oxford University Press, 1997.
- [2] M. Dorigo, T. Stützle, *Ant Colony Optimization*, MIT Press, Cambridge, MA, 2004.
- [3] L.J. Eshelman, The CHC Adaptive Search Algorithm: How to Safe Search when Engaging in Non Traditional Genetic Recombination, In *Foundations of Genetic Algorithms 1*, G.J.E. Rawlins (Ed.), Morgan Kaufmann Publishers, San Mateo, pp. 265-283, 1991.
- [4] F. Glover, G. Kochenberger (Eds.), *Handbook of Metaheuristics*, Kluwer Academic Publishers, 2003.
- [5] M. Laguna, R. Martí, *Scatter Search: Methodology and Implementations in C*. Kluwer Academic Publishers, Boston, 2003.
- [6] H.R. Lourenco, O.C. Martin, and T. Stützle. Iterated Local Search. In F. Glover and G. Kochenberger, Editors, *Handbook of Metaheuristics*, Kluwer Academic Publishers, 2003, pp. 321-353.

Development of High Level Software using Loosely Coupled Parallel Computers for the Optimal Design of Thermal Systems and Equipment.

TIC2003-07970

Asensio Oliva Llena*
Centre Tecnològic de Transferència de Calor (CTTC)
Universitat Politècnica de Catalunya (UPC)

Abstract

The goal of the project is to get deeper understanding of the fluid dynamics and heat transfer phenomena taking advantage of the new computing resources in parallel computation using loosely coupled parallel computers (Beowulf clusters). According to the most appropriate techniques for their parallelization, the codes/software to be developed within this project for the treatment of the different phenomena of interest have been grouped into three main areas: i) detailed resolution of turbulent incompressible flows based on DNS and LES methodologies, ii) resolution of flows with (approximately) parabolic spatial structure (cases of combustion, free surfaces and flow between plates will be of special interest); iii) resolution of flows with elliptic structure (turbulent flow in general with RANS models and moving surfaces). Although the project has essentially a basic/fundamental nature, the potential applications are addressed to the simulation of thermal systems and equipment of industrial and social interest in which the milestones achieved in this basic research line are applicable. Thus, the experience acquired will allow not only to maintain and improve the technology transfer to the companies with which the Group collaborates in the framework of research contracts and/or national and European research projects, but also to increment the possibilities of application to other companies and industrial and social sectors.

Keywords: heat and mass transfer, turbulence, DNS, LES, RANS, combustion, radiation, free surfaces, moving grids/boundaries, numerical methods, CFD&HT, HPCN, paralelization, Beowulf clusters.

1 Project objectives

1.1-Introduction

This project focuses on the continuation of the work that the Group is carrying out in the framework of its main research line: "Numerical simulation and experimental validation of heat and mass transfer phenomena; application to the design of thermal systems and equipment". Starting from the know-how and numerical facilities developed by the Research Group, and as a continuation of previous TIC projects, the goal of the project is to get deeper understanding of the

* Email: oliva@cttc.upc.edu

fluid dynamics and heat transfer phenomena taking advantage of the new computing resources in parallel computation using loosely coupled parallel computers: Beowulf clusters.

Even though the project has a fundamental character, and the main achievements will be in the academic area (publications, doctorate thesis...), it will be essential for the applied studies of the Group in the development of high performance software for the simulation of specific thermal system and equipment of industrial and social interest.

1.2-Strategy for parallel code development

According to the most appropriate techniques for their parallelization, the codes/software to be developed within this project for the treatment of the different phenomena of interest have been grouped into three main areas:

- 1) Detailed resolution of turbulent incompressible flows based on DNS and LES methodologies
- 2) Resolution of flows with (approximately) parabolic spatial structure (cases of combustion, free surfaces and flow between plates will be of special interest)
- 3) Resolution of flows with elliptic structure (turbulent flow in general with RANS models and moving surfaces).

The possibilities of using the most appropriate algorithms in every area without restrictions will compensate the large scope of the project in respect to the implementation, verification, validation and maintenance of the codes. The design of the algorithms and the structure of the three mentioned codes will assure their portability and efficient use in other architectural types, such as share memory and NUMA (Non-Uniform Memory Access) systems. The recent installation of the supercomputer Mare Nostrum, a distributed memory tightly coupled computer has confirmed that our initial strategy of total portability was appropriated.

1.3-Methodology

The general methodology of the work for the development of these algorithms and codes can be divided into the following levels (which present high inter-relationships):

- 1) Mathematical formulation of the physical phenomena.
- 2) Algorithms and techniques for the numerical resolution.
- 3) Development and verification of parallel software.
- 4) Resolution of different phenomena/configurations and analysis of the quality of the numerical solutions (discretization errors and convergence errors).
- 5) Validation of the mathematical formulations developed (modelization errors).

Although the project has essentially a basic/fundamental nature, the potential applications are addressed to the simulation of thermal systems and equipment of industrial and social interest in which the milestones achieved in this basic research line are applicable. Thus, the experience acquired will allow a technology transfer to the companies with which the Group collaborates in the framework of research contracts and/or national and European research projects. This will allow not only to maintain and improve these collaborations, but also to increment the possibilities of application to other companies and industrial and social sectors.

1.4-Project Development

The work is being carried out without differing significantly from the anticipated plan.

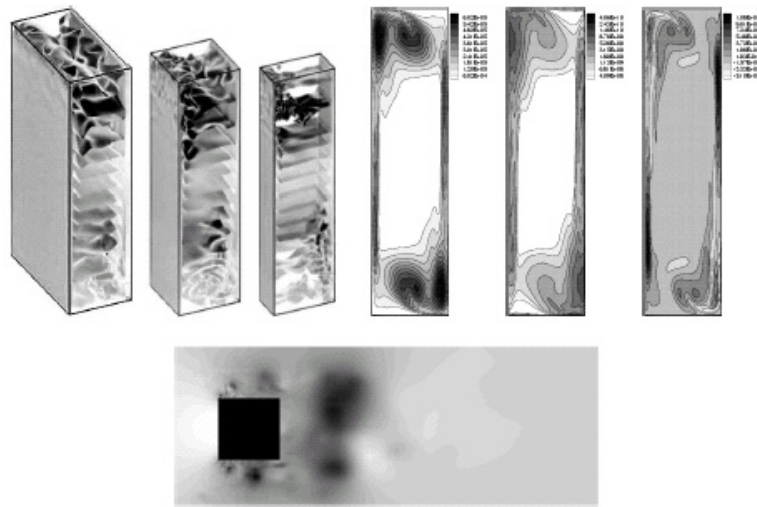
Concretely, the work is structured in three main areas of parallel CFD code development, plus a fourth area regarding know-how development concerning PC clusters.

[illegible]

1.4.1-Code for detailed resolution of incompressible turbulent flows by means of DBS-LES methods (**Code1**)

- The three-dimensional discretization of Verstapen and Veldman has been implemented in its second and fourth order versions.
- A parallel direct solver for second and fourth order Poisson equations, optimized for low-cost clusters, which permits the resolution of each Poisson equation with just one message, has been developed and implemented.
- Time average functions have been implemented.
- Work on implementation of a new turbulence model analog to LES models is being carried out.
- Direct simulations of natural convection in closed enclosures have been performed.

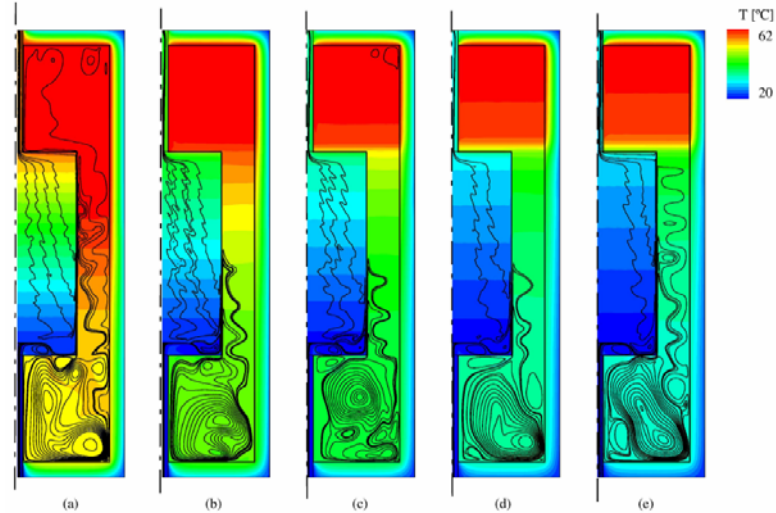
- The algorithms for the solution of the Poisson equation is being adapted to the new Mare Nostrum computer, in order to use effectively its major computing power. This task, which is of great interest, was not anticipated since neither this equipment existed nor the intentions of its installation were known in the beginning of the project.
- The code is still to extend to non-periodic situations in none of the axis, the implementation of LES models is to finish, and the code is to apply to more situations of interest.



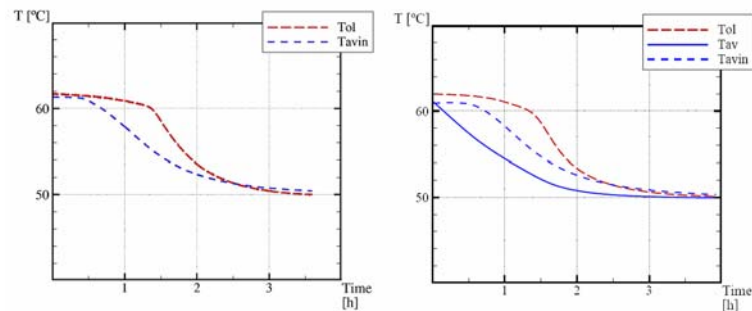
Illustrative results obtained with **Code1**. From left to right, three instantaneous temperature fields at $Ra_z = 6.4 \times 10^8$, 2×10^9 and 10^{10} , respectively, with $Pr = 0.71$ (air); the turbulent kinetic energy, the turbulent kinetic energy dissipation and the turbulent kinetic energy generation for $Ra_z = 6.4 \times 10^8$ and $Pr = 0.71$. Bottom: Instantaneous pressure field for flow past a square cylinder at $Re = 22,000$.

1.4.2-Code for the resolution of flows with approximately parabolic structure (**Code2**)

- Turbulence coupling formulations, generation ratios on the basis of Eddy break-up models and flamelet models have been carried out and implemented in the code.
- Radiation in participating media and its solution by means of the DOM technique and the Optically Thin Limit approximation have been formulated.
- Parallelization of DOM and its integration to the combustion code are pending.
- Regarding two-phase flows, the modeling has been carried out by means of a VOF model, implementation and application are performed in the design context of the absorption refrigeration systems using the parallel code for flows with parabolic structure.
- Regarding flows in channels of arbitrary geometry, the implementation of pressure and periodic boundary conditions, and the RANS models is completed.
- Coupling radiation in participating media (DOM) with combustion is pending. To do so, it is necessary to modelize the behavior of gas mixtures, implementing a multiband model.



Sample of Code2 results: CFD simulation of a Multifunctional thermal energy storage prototype (in black the streamlines).



Comparison of experimental results and CFD simulations using Code2. Testing of a Multifunctional thermal energy storage. Left: experimentally measured average and outlet temperature. Right: CFD results, also with average temperature (discontinuous blue line). Note (i) the good agreement (ii) that CFD simulations provide more data than experiments, the average temperature in this case.

1.4.3-Code for flows with elliptic structure (**Code3**). The following goals have been achieved:

- Implementation of two-equation turbulence models, introducing anisotropy and nonlinearity
- Implementation of Explicit Algebraic Reynolds Stress Models
- Comparative study of different RANS formulations
- Generalization of RANS models for turbulent flows with moving boundaries
- Studies on the discretization of compressible flow equations. Numerical implementation.
- Generalization of RANS for compressible flows (Favre-averaging)

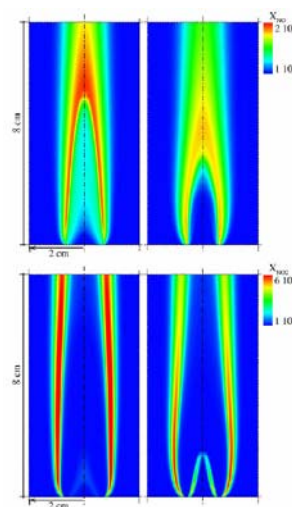
Sample results of Code3. Pollutant formation. Mole fraction isopleths for: nitrogen oxide, NO (top); nitrogen dioxide, NO₂ (bottom).

Implementation of the Differentially Reynolds Stress Models is pending.

Intensive work has been carried out on non-structured grids and actually a sequential prototype code (in C) to resolve the three-dimensional and transient Navier-Stokes equations is available. Other parallel C++ code is being developed, all the basic classes to contain the grids, communications, etc. have been written.

Finishing the sequential prototype for non-structured grids in C, transferring the code to the parallel C++ version, and implementing additional models in this new code, like the RANS-type models, are pending.

This is essentially what is anticipated in the project, except for the delay in the initial phase caused by the use of C++, but the result is expected to be more robust and modular.



1.4.4-Tools for the administration of the PCs

Different codes in order to facilitate the installation and management of the cluster are developed. The selection and implementation of software for parallel queue management is pending.

2 Indicators of Results

Information on different indicators that can show the relevance of the results obtained up to now has been included. These indicators consider at least: staff in formation, achieved publications, technology transfer, patents, participation in international projects (especially in the EU), and collaborations with national and foreign groups.

2.1 Participation in projects

A narrow collaboration is maintained with the companies that supported this project: Unidad Hermética of Electrolux Group (nowadays CUBIGEL S.A. (Unidad Hermética)); Radiadores Cerezo, Tefrinca. Below are the collaboration projects with the above-mentioned companies, for which the achievements of this project are fundamental:

- Research project nº C04749, Company: CUBIGEL, S.A. (Unidad Hermética), Period: 2002-2005, Funding: 365.000 Euro, Title: Study of the viability and development of hermetic compressors prototypes working with CO₂ as fluid refrigerant.
- CRAFT Programme (ref. CRAFT – 1999 – 71024 – “COMHEX”), European Commission, Directorate-General XII. The CTTC is the Research Project Co-ordinator. Period: 2002-2004; Funding corresponding to the CTTC: 351.060 Euro. Title: Optimisation of Gas-Liquid Compact Heat Exchangers Considering Their Interaction with the Rest of the System.
- Research project nº C05652, Company: CUBIGEL S.A. , Period: 2005 – 2006; Funding: 108.000,00 Euro. Title: Optimization of Hermetic Reciprocating Compressors Based on Numerical Simulation and Experimental Validation Methodologies. Detailed Study of the Flow through Valves and the Piston-Cylinder Leakages.

- CRAFT Programme (Contract No. COOP-CT-2004-513106-EFROST). The CTTC is the Research Project Co-ordinator. Period:2005-2006; Funding corresponding to the CTTC: 476.051,00 Euro. Title: Efficient Refrigerated Food Storage.

The Group maintains contacts and carries out collaborations with various centers/universities and companies, with the common link being the numerical resolution of heat and mass transfer and fluid dynamics phenomena. The majority of these projects is of European scope and performed under our leadership. The most relevant projects are given below:

- CRAFT Programme (ref. CRAFT-1999-70878-“STATIC II”)(ref. CTT n° E00694), European Commission, Directorate-General XII, Period: 2002-2004; Funding corresponding to the CTTC: 372.562 Euro, Title: Stagnation Proof Solar Collectors Transparently Insulated Flat Plate Collectors - II. CTTC is the research coordinator.
- CRAFT Programme (ref. CRAFT-1999-70967-“ASFIC”)(ref. CTT n° E00695), European Commission, Directorate-General XII, Period: 2002-2004, Funding corresponding to the CTTC: 396.592 Euro, Title: Advanced Solar Facades with Integrated Collectors-Accumulators for Domestic Hot Water and Space Heating Applications. CTTC is the research coordinator.
- CRAFT Programme (ref. CRAFT – 1999 – 71024 – “COMHEX”), European Commission, Directorate-General XII. The CTTC is the Research Project Co-ordinator. Period: 2002-2004; Funding corresponding to the CTTC: 351.060 Euro. Title: Optimisation of Gas-Liquid Compact Heat Exchangers Considering Their Interaction with the Rest of the System.
- CRAFT Programme (ref. CRAFT – 1999 – 70604 – “OPICS”), European Commission, Directorate-General XII. The CTTC is the Research Project Co-ordinator. Period: 2002-2004; Funding corresponding to the CTTC: 336.500 Euro. Title: Optimised Integrated Collector Storage: low-cost solar thermal systems for houses and offices.
- CRAFT Programme (Contract N°: ENK6 – CT – 2002 - 30033), GREENFACADE. The CTTC is the Research Project Co-ordinator. Period: 2003-2004; Funding: 311.200 Euro. Title: Vegetation for cooling advanced glazed facades: Arquitectural integration, gardening technology, numerical simulation.
- CRAFT Programme (Contract No. 1999 - 72475), HYDRA. The CTTC is the Research Project Co-ordinator. Period: 2003-2005; Funding 333.400 Euro. Title: Hybrid latent/sensible compact storage devices for combined thermal solar energy applications: refrigeration, heating and air – conditioning (HYDRA).
- CRAFT Programme (ref. CRAFT – 1999 – 71217 – “COMPACT”), European Commission, Directorate-General XII. The CTTC is the Research Project Co-ordinator. Period: 2003-2005; Funding corresponding to the CTTC: 244.168 Euro. Title: Low-cost compact solar heaters made of plastic materials and composites.
- CRAFT Programme (Contract No. COOP-CT-2004-513106-EFROST). The CTTC is the Research Project Co-ordinator. Period:2005-2006; Funding corresponding to the CTTC: 476.051,00 Euro. Title: Efficient Refrigerated Food Storage.

2.2 Publications

- J. Castro, L. Leal, C.D. Pérez-Segarra, P. Pozo, “Numerical study of the enhancement produced in absorption processes using surfactants”, *International Journal of Heat and Mass Transfer*, Vol. 47, p.3463 – 3476, 2004.
- G. Colomer, M. Costa, R. Cònsul, A. Oliva, “Three – dimensional numerical simulation of convection and radiation in a differentially heated cavity using the discrete ordinates methods”, *International Journal of Heat and Mass Transfer*, Vol. 7, n° 2, p.257 – 269, 2004.

- O. García-Valladares, C.D. Pérez-Segarra, and J. Rigola, “Numerical simulation of double-pipe condensers and evaporators”, *International Journal of Refrigeration*, Vol 27, p.656-670, 2004.
- M. Soria, X. Trias, C.D. Pérez-Segarra, A. Oliva, “Direct numerical simulation of the three-dimensional natural convection flow in a differentially heated cavity of aspect ratio 4”, *Numerical Heat Transfer. Part A: Applications*, Vol 45, 2004.
- K. Claramunt, R. Cònsul, C.D. Pérez-Segarra, A. Oliva, “Multidimensional Mathematical Modelling and Numerical Investigation of Co-flow Partially Premixed Methane/Air Laminar Flames”, *Combustion and Flame*, Vol. 137, p.444 – 457, 2004.
- J. Rigola, G. Raush, C.D. Pérez-Segarra, A. Oliva, “Detailed experimental validation of the thermal and fluid dynamics behavior of hermetic reciprocating compressors”, *Int. Journal of Heating, Ventilating, Air Conditioning and Refrigerating Research*, Vol. 10, p.291-306, 2004.
- R. Cònsul, I. Rodríguez, C.D. Pérez-Segarra, M. Soria, “Virtual prototyping of storage tanks by means of three-dimensional CFD and heat transfer numerical simulations”, *Solar Energy*, Vol. 77, p.179 – 191, 2004.
- G. Raush, J. Rigola, C.D. Pérez-Segarra, A. Oliva, “A novel Sigma-Delta ADC application oriented to test hermetic reciprocating compressors. *Measurements Science and Technology*, Vol. 15, p.2207-2214, 2004.
- C.D. Pérez-Segarra, J. Rigola, M. Soria, A. Oliva, “Detailed thermodynamic characterization of hermetic reciprocating compressors”, *International Journal of Refrigeration*, Vol. 28, No.4, p.579-593, 2005.
- J. Rigola, C.D. Pérez-Segarra, A. Oliva, “Parametric studies on hermetic reciprocating compressors”, *International Journal of Refrigeration*, Vol. 28, No.2, p.253-266, 2005.
- F.X. Trias, M. Soria, C.D. Pérez-Segarra, A. Oliva, “A direct Shur-Fouriere decomposition for the efficient solution of high-order Poisson equations on loosely coupled parallel computers”, *Numerical Linear Algebra with Applications*. In press.
- C.D. Pérez-Segarra, C. Farre, J. Cadafalch, A. Oliva, “Analysis of different numerical schemes for the resolution of convection-diffusion equations using finite volume methods on three-dimensional unstructured grids. Part I: Discretization schemes”, *Numerical Heat Transfer. Part B: Fundamentals*. In press.
- C. Farre, C.D. Pérez-Segarra, M. Soria, A. Oliva, “Analysis of different numerical schemes for the resolution of convection-diffusion equations using finite volume methods on three-dimensional unstructured grids. Part II: Numerical analysis”, *Numerical Heat Transfer. Part B: Fundamentals*. In press.

2.3 Doctoral thesis

Two doctorate thesis have been performed within the framework of the present TIC project and the last period of the previous one. The maximum qualification have been obtained in the two theses: Kilian Claramunt Altimira (February 2005), “*Numerical simulation of non-premixed laminar and turbulent flames by means of flamelet modelling approaches*”, and Jesús Castro González (May 2005), “*Simulation of heat and mass transfer phenomena in the critical elements of H₂O-LiBr absorption cooling machines. Experimental validation and application to design*”. In September, the thesis “*Development of experimental and numerical infrastructures for the study of compact heat exchangers and liquid overfeed refrigeration systems*” by Stoyan Viktorov Danov will be read. Three more thesis will be read within the next year.

Sistema de Apoyo a la Visión mediante una Retina Artificial Neuromórfica a través de Realidad Aumentada

TIC2003-09557

José Manuel Ferrández*
Universidad Politécnica de Cartagena

Eduardo Fernández**
Universidad Miguel Hernández

Abstract

The construction of a system for aiding some kind of blindness is proposed. It consists in a bioinspired retina model into a reconfigurable architecture which is connected to an augmented reality device based on see-through technology. This device will permit the integration of real world images to the ones obtained by the neuromorphic retina. This system will provide a local gain control, in addition of a bioinspired preprocessing, FPGAs plasticity, and the miniaturization of the components, including low power devices. These characteristics are highly desirable in some low vision supporting systems.

Keywords: Neural acquisition, retinal coding, cellular neural networks, augmented reality.

1 Project Objectives

The objectives defined within the project are structured in a workplan. The list of objectives, with the estimated time and the responsible unit is the following:

- **Objective M1:** Study of the processing and coding capabilities of population of ganglions cells in the visual system.
Estimated time: 20 months
Responsible Unit: Universidad Miguel Hernández
- **Objective M2:** Define mathematical algorithms and simulation tools.
Estimated time: 6 months
Responsible Unit: Universidad Miguel Hernández
- **Objective M3:** Cellular Neural Network development
Estimated time: 6 months
Responsible Unit: Universidad Miguel Hernández

* Email: jm.ferrandez@upct.es

** Email: e.fernandez@umh.es

- **Objective M4:** Cellular Neural Model VHDL implementation
Estimated time: 12 months
Responsible Unit: Universidad Politécnica de Cartagena
- **Objective M5:** Software/hardware system partition.
Estimated time: 18 months
Responsible Unit: Universidad Politécnica de Cartagena
- **Objective M6:** Reconfigurable hardware design.
Estimated time: 15 months
Responsible Unit: Universidad Politécnica de Cartagena
- **Objective M7:** System adaptation to low vision problems. Augmented Reality application
Estimated time: 8 months
Responsible Unit: Universidad Politécnica de Cartagena
- **Objective M8:** Meetings, Publications and Patents
Estimated time: 6 months
Responsible Unit: Universidad Politécnica de Cartagena and Universidad Miguel Hernández

The described Workplan Flow Chart is

Objectives	Estimated time (months)																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
M1																								
M2																								
M3																								
M4																								
M5																								
M6																								
M7																								
M8																								

Means to achieve the objectives

The group at Universidad Politécnica de Cartagena is devoted basically to the development of the neural network models and its electronic implementation, as the augmented reality real time application. The group at Universidad Miguel Hernández is devoted to the acquisition of multielectrode recordings from retinal ganglions cells, and analyzing the coding capabilities and strategies of the visual system.

The equipment used by the group at Universidad Politécnica de Cartagena is composed by Xilinx FPGA development system, a network of workstations for simulation under MATLAB ® , Sony Glasstron PLM-700 augmented reality glasses, as well as different simulation software both for WINDOWS and LINUX.

The equipment used by the group at Universidad Miguel Hernández is composed by a multielectrode recording setup which consists in a fixed workstation for amplification, data recording and storage running Cyberonics systems, a multielectrode array developed in Utah. The recordings analysis is made with WINDOWS and LINUX systems.

2 Level of success achieved

2.1 Neural acquisition and analysis

Understanding how information is coded in different sensory systems is one of the most interesting challenges in neuroscience today. Much work is focused on how various sensors encode different stimuli in order to transmit information to higher centers in the neural hierarchy in a robust and efficient manner. Technology is now available that allows one to acquire data with more accuracy both in the temporal and the spatial domains. However this process produces a huge neural database which requires new tools to extract the relevant information embedded in neural recordings. However, it allows one to postulate some rules about the way different neural systems do this coding.

The Utah microelectrode array was used for obtaining the extracellular recordings. The electrode arrays contain 100 electrodes that were built from silicon on a square grid with 400 microns pitch. The distal 50 microns of the needles, metallized with platinum, form the active site of each electrode. The array was mounted on a micromanipulator, and each electrode was connected to a 5000 gain bandpass (filtered from 250 to 7500 Hz) differential amplifier. The analog signals were digitized and stored in a computer. The pre-processing consisted of fixing a threshold (based on an estimate of the noise of the electrode) in order to extract the kinetics of electrical activity above this threshold. Spike waveform prototypes were separated by template matching. For each stimulus and for each electrode, the time of the first and second spike, the number of spikes, and the interspike interval during the flash interval was also computed. (Figure 1)

Different studies have used different analysis tools to approach the decoding objective. It has been applied a statistical analyzer to the neural data in order to estimate the characteristics of the stimulus, and also linear filters for the decoding task. Other approaches used to get insights into the coding process are discriminant analysis, principal component analysis and supervised and non-supervised neural networks. Finally information theory and synchronicity studies have been used for analysing the synergistic or redundant behaviour of the code, and for determining the relevance of the temporal and spatial parameter in the code and also to check if there exists a population code instead of an aisle cell code[1],[6]. This studies have been applied in refining the parameters of the cellular neural network.

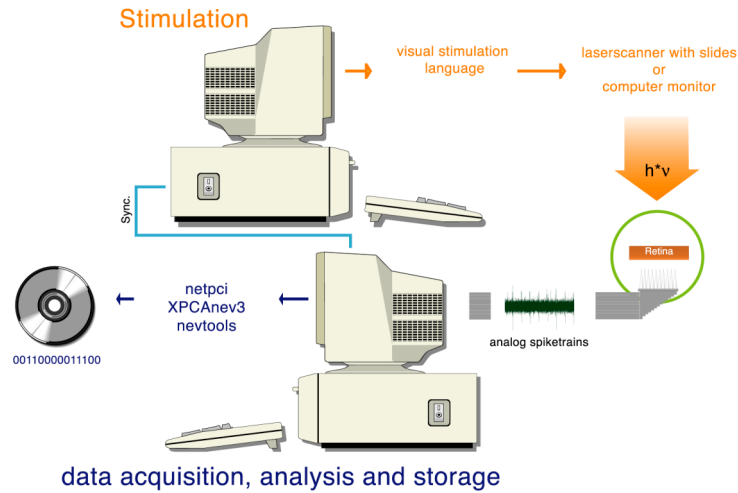


Fig 1. Recording method

2.2 Neural Tools developed

The number of laboratories using techniques that allow to acquire simultaneous recordings of as many units as possible is considerably increasing each day. However the development of tools used to analyze this multi-neuronal activity is generally lagging behind the development of the tools used to acquire this data. Moreover, the data exchange between research groups using different multielectrode acquisition systems is hindered by commercial constraints such as exclusive file structures, high priced licenses and hard policies on intellectual rights. It has been developed a free opensource software for the classification and management of neural ensemble data. The main goal is to provide a graphical user interface that links the experimental data to a basic set of routines for analysis, visualization and classification in a consistent framework. To facilitate the adaptation and extension as well as the addition of new routines, tools and algorithms for data analysis, the source code and documentation are freely available.

The main aim of this work was to provide a free open source software that allows an effective multielectrode data exchange between laboratories regardless the commercial system used for the recordings and that, at the same time, make available appropriate tools for an easy and fast pre-processing of the data [7],[8],[11],[14]. Data sharing requires agreement on techniques and formats, as well as methods for classification and selection therefore DATA-MEAns provide several visualization and analysis routines which permit to classify and reduce original datasets for further specific processing. Furthermore to facilitate the adaptation to any particular requirement as well as the addition of new routines, tools and algorithms for data analysis, the source code and documentation is freely available. We hope that it will help new contributors to add functions that are currently not included and that it could be useful for further analysis that allow integrate and compare findings from individual laboratories as well as test new hypothesis.

DATA-MEAns can read ASCII files created by popular data acquisition systems. These files can be easily created and edited using a word processor or any other program for editing text files. Although working with binary files allow to come up with more information, this approach would require huge and continuous efforts for putting together all different file formats, reading them,

and updating the “read modules” since the commercial acquisition systems often upgrade their native data files. Moreover, many laboratories use home-made acquisition systems, being excluded from using most of the commercially available software. Our goal is that this toolbox can be useful to the development of standards, all in a common format, which support multiplatform assembly, upload, annotation, search and acquisition, that will ease reanalysis of shared data by permitting data upload from within data acquisition applications, and data download directly to standard acquisition routines.

The design based on stand-alone applications offers great flexibility and independence from any other commercial software and represents an easy mechanism for adding new routines altering minimally the source code. This can be particularly useful when the researcher uses their own custom scientific programs. Thus it is possible to write the desired algorithm using a known language, make an executable and link it to the DATA-MEAns application by means of the ShellApi utility from Windows. Anyway we are working on a new multiplatform version of the software that will run under MacOSX and Linux. Finally, this application was intended only for a preliminary pre-processing of the data, hence only the most generally used routines are implemented, (Figure 2). We are aware that numerous functions currently in use, as well as advanced functions used in various labs working with multi-electrode data are not included. We hope that the facilities to extend the code and add more specific algorithms and routines can compensate the former and encourages the users to contribute providing ready-made functions useful to the scientific community. Thus, the program and source code are available free of charge upon request from the first author of this paper (email: p.bonomini@umh.es) or directly from the following URL: <http://cortivis.umh.es> (under the software menu), to academic researches interested in using or enhancing DATA-MEAns.

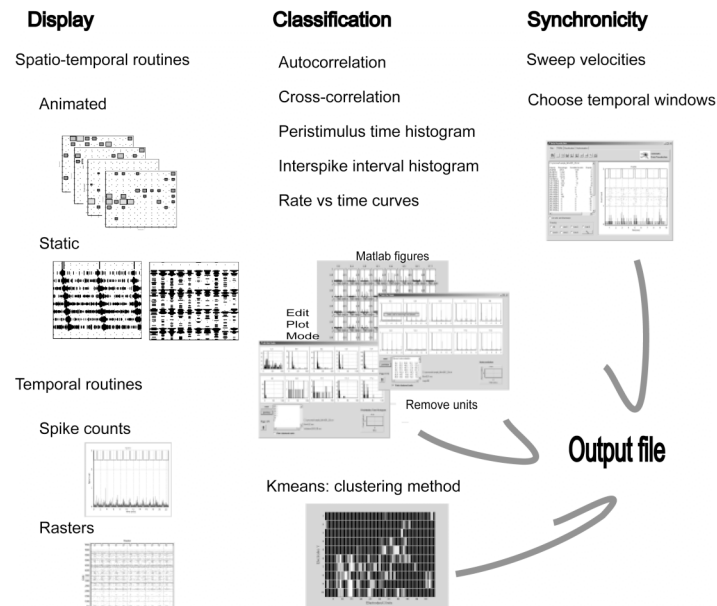


Fig 2. Functions implemented by DARA-MEAns

2.3 Model implementation

Cellular Neural networks (CNN) are a powerful analogue parallel computing paradigm. For software applications, which do not require a great capacity of processing, many practical applications have been implemented using general purpose microprocessor or DSP processor. However, for intensive computing applications such as real time signal processing, video processing or retinal emulation, a hardware implantation is necessary in order to achieve the maximum processor capacity required by these applications. In order to find the best approach in both cases, a new approach to Cellular Neural Networks model is used [16]. This approach is focused on CNN implementation on reconfigurable hardware architectures (FPGAs) and DSP microprocessors. The CNNs are analyzed from the perspective of systems theory, giving rise to an alternative model to those found in the literature available. Dynamic equations and their solutions, stability analysis and real-time implementation architecture are described as the most relevant points in the development of our model.

From the point of view of CNNs hardware implementations, local connectivity and intrinsic analogic parallel processing provided a first advantage to VLSI implementation. However nowadays, digital devices such as FPGA are playing an important role by competing with the VLSI circuits not only by its low price, but also by its features and versatility in the design. The new families of FPGA, which are emerging these last years, provide many hardware resources such as: optimized Block Select RAM, embedded multiplier accumulator (MACC), fast carry logic etc. These features and their capacity for parallel processing have increased the flexibility to map modular neural structures into FPGAs, although it is necessary to use hardware element as much as possible. That is why, we have proposed to analyze different digital models of CNN neuron [17], which arise from the original continuous model, to find the model which employs the minimum resources possible for its implementation, and is a good approximation of the continuous model as well (Figure 3). The main results, obtained from different implementations [18, 19], evidence the usefulness and functionality of the selected neuron model.

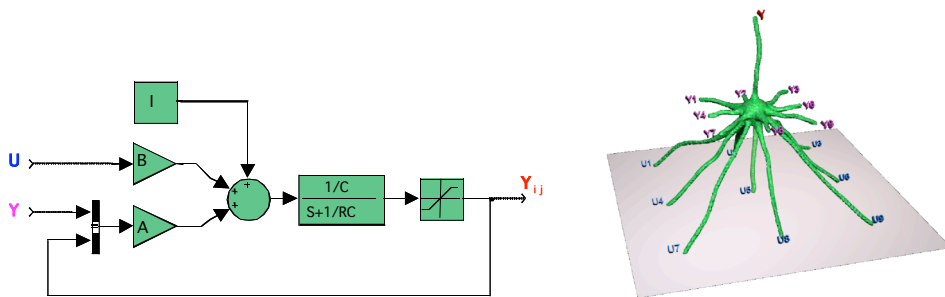


Fig. 3. Mathematical model and symbolic 3D representation of a single neuron

The most interesting result obtained with a single neuron in term of speed and area resources (Figure 4) has motivated the search of different solutions for the complete implementation of a great Cellular Neural Network with over 300.000 equivalents cloned neuron in a single FPGA. For this task we have proposed a novel architecture based on self-clocked approach which has given excellent result in other similar applications [20, 21] and that is very hopeful for the design of

CNN. Future works will focus on the design of multi layer CNNs as well as CNN with not cloned and CNN with nonlinear and variant in the time coefficients. Another aim is to connect the full neural network to co-design system [22] to achieve a greater versatility in the design and test of applications with CNN.

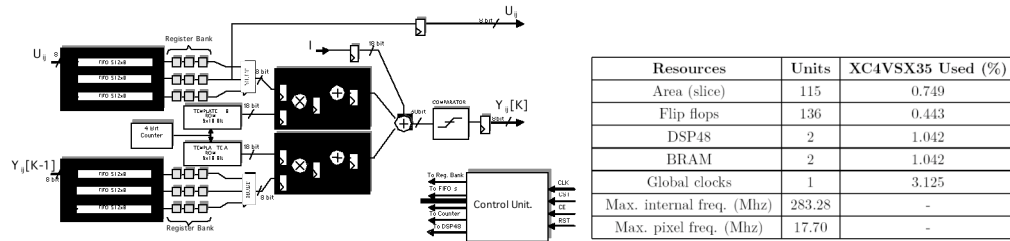


Fig. 4. Single neuron architecture and summary of used resources

2.4 Augmented Reality application

At present, an Augmented Reality (AR) system fully based on Field Programmable Gate Array (FPGA) devices is being developed. It aims for providing patients affected by a visual disorder known as tunnel vision with a help for their loss of peripheral vision. Tunnel vision is associated to eyes diseases such as retinitis pigmentosa and glaucoma, and it reduces considerably the patient's ability to localize objects or persons and navigate, and consequently, his relationship with people and the environment. To enhance the user's knowledge of the environment, we have adopted an AR approach where information considered as useful is presented to him thanks to a see-through Head Mounted Display (HMD). This information consists of the contours detected on the images captured from a camera worn by the user. An example of the view of the user through the HMD is shown in Figure 5. It can be seen how the contour information is superimposed on the user's own view.



Fig. 5. Simulation of patient's view through the HMD looking at a clothes stand situated on a corner and beside a shelf.

In order to achieve both the constraints of size and power imposed by the mobility of the system, and the requirements of high performance of the application, an FPGA-based platform is proposed. Reconfigurable hardware offers a trade-off between the very specialized solution of Application Specific Integrated Circuits (ASICs) and the inefficiency of General Purpose Processors (GPPs), combining the desirable high performance of ASICs with the characteristics of system flexibility of GPPs. Like ASICs, it involves hardware implementation and consequently parallelism and high performance. Like GPPs, it provides reconfigurability, and hence flexibility and rapid prototyping. So, it is possible to implement not only specific algorithms but also interfaces, controllers, glue logic, even microprocessors, and so to integrate the whole system in just one chip.

To extract the contour information two different approaches have been developed, implemented and evaluated: the Canny algorithm and a cellular neural network with the appropriate template. The first one is a specifically designed solution to extract edges from images. The second one is a more versatile solution, thanks to the possibility of modifying the processing through the change of the template set. The Canny algorithm is a well known image processing algorithm for extracting contour information. We have proposed [23] a novel FPGA-based architecture for its hardware implementation, optimized for Xilinx Virtex-II devices. Specific purpose resources included in this family of FPGAs, such as embedded 18 bits multipliers and internal BlockRAMs, have been used to achieve efficient implementation. It is able to process up to 1024×768 pixels image at 60 Hz, reaching the real time execution required for the application.

Nowadays, the advantages of the combination of massive parallelism and local interaction among cells that offered the CNN are well known, especially for image processing. Moreover, in our application with the use of CNNs we aim to increase the versatility of the system, making possible to tune the output to the desired and to perform different processing algorithms just by changing the template set. For CNN hardware implementation, VLSI design has been the most suitable choice to exploit their advantages up to now. For our application, we have described CNN discrete models focused on the implementation in reconfigurable hardware [24, 25, 26].

The system is shown in Figure 6. It consists basically of a camera to acquire images of the environment, a head mounted display to visualize the information that enhances the user's vision, a Xilinx Virtex-II FPGA as processor and controller unit and SRAM memory to store information. The camera is a monochrome module with a 384×288 pixels 8 bits digital output at 50 frames per second, based on a CMOS image sensor. The HMD used is the Sony Glasstron PLM-S700E. It is a binocular and high resolution device with adjustable see-through capability. Its design and light weight allows its use with correction glasses and makes it easy the movement of the head and the mobility of the person. VGA synchronization signals for the HMD are generated by the FPGA to show a 640×480 pixels image at 60 frames per second. The SRAM memory is necessary to store image from the camera and, in the CNN-based system, information between different stages. The memory interface for controlling the write and read operations is implemented on the own FPGA. For the Canny-based approach, a XC2V1000 Xilinx FPGA has been used. The implementation results and timing statistics are shown in [23, 26]. For the CNN-based approach a XC2V4000 Xilinx FPGA has been used. Initial implementation results are presented in [24] and improved results in [25, 27]. A comparison between the two approaches is carried out in [26].

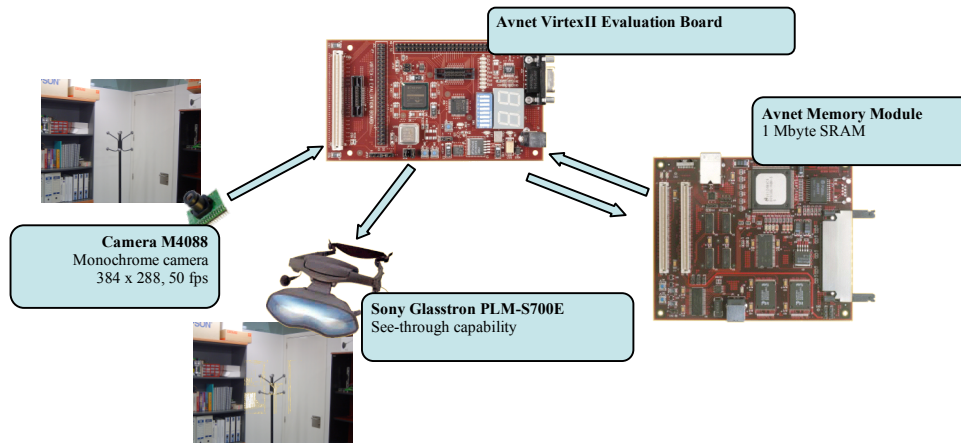


Figure 6. System Architecture.

3 Result markers

3.1 Personnel under training

Student working in the Project:

- Jose Javier Marínez Alvarez, Ph.D. student, is developing cellular neural networks on FPGAs, studying different architectures, and for different applications.
- Javier Toledo Moreo, Ph.D. student, is working on augmented reality systems on reconfigurable devices, for low vision problems.
- M. Paula Bonomini, Ph.D. student is working on acquiring and analysing neural data, mainly retinal ganglion cells activity and visual cortex

Jose Javier Marínez Alvarez, Javier Toledo Moreo and M. Paula Bonomini are working in the project since its startup and there exist three more graduate students collaborating with different tasks in the system.

3.2 Publications carried out fully or in part under the Project Sponsorship

Publications [1][6] have to do with the acquisition and analysis of the retinal code. The tools for analysing the neural data are described in [7][8][11][14]. The retina model and the cellular neural network implementation have been published in [12][15][16][17][18][19][20][21][22]. Publications [23][24][25][26][27] have to do with the real time augmented reality application for low vision patients. Finally, the application of the retinal model for a cortical visual neuroprostheses have been described in [2][3][4][5][9][10][13]

3.3 International Projects, Collaboration with national and foreign groups

The Miguel Hernández University group act as coordinator of a EC project (Contract N°: QLK6-CT-2001-00279) under the quality of life program, with an amount of 2.256.730 Euros.

Project Full Name: Cortical Visual Neuroprosthesis for the blind

Acronim: CORTIVIS

Coordinador: Dr. Eduardo Fernandez

Research Programme: Quality of Life

Tematic Priorities: QOL-2001-6.4, QOL-2001-6.1

Both research groups are working in a proposal to the next Future Emerging Technologies (FET) Program of the EC, oriented to e-Health. The following partners have been collaborating with both groups:

- Dr. Helga Kolb Department of Physiology, University of Utah
Salt Lake City, USA
- Dr. Josef Ammermüller, Department of Biology and Neurobiology
University of Oldenburg, Oldenburg, Germany
- Dr. Peter Ahnelt Department of Physiology, Medical School
University of Vienna, Vienna, Austria
- Dr. Dick Normann, Department of Bioengineering
University of Utah, Salt Lake City, USA
- Dr. Lyle Graham Centre National de la Recherche Scientifique UFR Biomédicale
Université René Descartes, Paris, France
- Dr Alvaro Pascual-Leone Beth Medical Hospital and University of Harvard
Boston, USA
- Dr. Pierre Rabischong, Faculte de Medicine
University of Montpellier, Montpellier, France
- Dr. Leonel Sousa , Instituto de Engenharia de Sistemas e Computadores
Signal processing systems, Lisboa, Portugal
- Dr. Wim Rutten Department of Biomedical Signal and Systems.
University of Twente, Netherlands
- Dr. N.F. de Rooij , Institute of Microtechnology
University of Neuchatel, Switzerland
- Francisco Pelayo, Departamento de Arquitectura de Computadores
Universidad de Granada, Granada
- Fernando Vargas, Departamento de Física
Universidad de Murcia, Murcia

3.4 Social collaborations

Both research groups have been invited by **Asociacion Nacional de Retinosis Pigmentaria**, to show their results on the day of this disease in Albacete. A prototype of the system was used by the patients, where they check the functionality of the device, and collaborate in future refinements.

4 References

- [1] Ortega, G.J., Bongard, M., Louis, E., Fernández, E. "Conditioned spikes: a simple a fast method to represent rates and temporal patterns in multielectrode recordings." *J. Neurosci Methods*. 133, 2004: 135-141.
- [2] Pelayo F.J., Romero S, Morillas C, Martínez A., Ros E., Fernández E. "Translating image sequences into spike patterns for cortical neuro-stimulation." *Neurocomputing*, vol. 58, 2004, pp. 885-892.
- [3] Fernández, E, Pelayo , F., Amermmuller, J., Ahnelt, P., Norman, R.A. "Cortical Visual Neuroprotheses for the Blind". *Restorative Neurology and Neuroscience*, 2004.
- [4] Romero, S., Pelayo, F., Morillas, C.A., Martínez, A., Fernández, E. "Reconfigurable retina-like preprocessing platform for cortical visual neuroprotheses." *Neuroengineering*. Wiley Press. 2004
- [5] Pelayo, F.J., Martínez, A., Romero, S., Morillas, C., Ros, E., Fernández, E. "Cortical visual neuroprotheses for the blind: retina-like software/hardware pre-procesor." *IEEE EMBS*, 2004
- [6] Ortega, G., Bongard, M., Louis, E., Fernandez, E. "Deterministic rate synchronicity in retinal ganglion cells." *J. Neurophysiol*. 2005. Submitted
- [7] Bonomini P., Ferrández J.M., Fernández E. "Dataset: an open source tool for the management of multielectrode array data", *31st Göttingen Neurobiology Conference*. Göttingen, Alemania 2005. German Society of Neuroscience Proceedings. George Thieme Verlag Stuttgart.
- [8] Bonomini, M.P., Ferrandez, J.M., Bolea, J.A., Fernández E. "DATAMeans: An open source tool for the classification and management of neural ensemble recordings." *J. Neurosci. Methods*, 2005. Accepted, in press.
- [9] Alfaro, A., Concepción, L., Fernández, E. " Visual Hallucinations in blind: from sensory deprivation to visual activity." *Ann. Neurol*. 2005. Submitted
- [10] Ferrandez J.M. , Bonomini M.P., Fernandez E. "A Multiplexed Current Source Portable Stimulator Architecture for a Visual Cortical Neuroprosthesis" *Lect. Not. Comp. Sci.* 3561, Springer, pp. 310-318, 2005
- [11] Bonomini M.P., Ferrandez J.M., Bolea J.A., Fernandez E. "Formulation and Validation of a method for classifying Neurons from multielectrode recordings" *Lect. Not. Comp. Sci.* 3561, Springer, pp. 68-76, 2005
- [12] Christian A. Morillas , Samuel F. Romero, Antonio Martínez ,Francisco J. Pelayo, Eduardo Ros , and Eduardo Fernández, "A Design Framework to Model Retinas." *International Workshop on Information Processing in Cells*, York, United Kingdom, 2005
- [13] S. Romero, C. Morillas, A. Martínez, F. Pelayo, E. Fernández, "A Research Platform for Visual Neuroprotheses" *Biosystems*. 2005. Accepted, in press
- [14] M. Bongard, D. Micol, E. Fernandez "An offline preprocessor for Neuronal Event Data" *J. Neurosci. Methods*, 2005. Submitted
- [15] Martínez-Alvarez J.J, Garrigós-Guerrero F.J, Toledo-Moreo F.J, Ferrández-Vicente J.M. "Modelado, Síntesis Digital y Descripción Hardware de Algoritmos y Sistemas Lineales con MATLAB®" *FPGAs: Computación & Aplicaciones JCR4'2004*", Barcelona, ISBN: 84-688-7667-4, pp. 193-200.
- [16] Martínez J.J, Toledo F.J., Ferrández J.M.: "New Emulated Discrete Model of CNN Architecture for FPGA and DSP Applications". 7th International Work-Conference on Artificial and Natural Neural Networks, IWANN 2003, LNCS 2687, June 2003, Maó, Menorca, Spain, , pp. 33-40
- [17] Martínez J.J, Toledo F.J., Ferrández J.M.: "Implementation of a Discrete Cellular Neuron Model (DT-CNN) Architecture on FPGA", *SPIE Microtechnologies for the New Millenium*,

- conf. on bioengineered and bioinspired systems, vol. 5839, May 2005, Sevilla, Spain, pp. 332-340
- [18] Toledo F.J., Martínez J.J, Garrigos F. J., Ferrández J.M.: "Image processing with CNN in a FPGA-based augmented reality system for visually impaired people", 8^o Int. Work-Conference on Artificial and Natural Neural Networks, IWANN 2005, LNCS 3512, June 2005, Vilanova i la Geltrú, Spain, pp. 906-912
 - [19] Toledo F.J., Martínez J.J, Garrigos F. J., Ferrández J.M.: "An augmented reality visual prosthesis for people affected by tunnelling vision", *1er Int Work-Conference on the interplay between Natural and Artificial Computation*, IWINAC 2005, LNCS 3561, June 2005, Spain, pp. 319-326
 - [20] Martínez J.J, Toledo F.J, Garrigos F. J., Ferrández J.M.: "FPGA Implementation of an Area-Time Efficient FIR Filter Core Using a Self-Clocked Approach", *15^o Int Conf. on Field Programmable Logic and Applications FPL 2005*, August 2005, Tampere, Finlandia, (accepted, in press)
 - [21] Martínez J.J, Toledo F.J, Garrigos F. J., Ferrández J.M.: "Filtro eficiente FIR-MAC para el diseño a alto nivel con Xilinx System Generator", *I Congreso Español de Informatica*, JCRA 05, CEDI 2005, Septiembre 2005, Granada, Spain, (accepted, in press)
 - [22] Martínez J.J, Legaz A., Toledo F.J, Garrigos F. J., Ferrández J.M.: "Sistema basado en PCI para la co-simulación de aplicaciones DSP y de imagen sobre FPGAs desde Matlab", *I Congreso Español de Informatica*, JCRA 05, CEDI 2005, Septiembre 2005, Granada, Spain, (accepted, in press)
 - [23] Toledo, F. J., Martínez, J. J., Garrigós, F. J., Ferrández, J. M.: "Augmented Reality System for Visually Impaired People based on FPGA", *4^o LASTED Int. Conf on Visualization, Imaging and Image Processing*, VIIP 2004, Sept. 2004, Marbella, Spain, pp. 707-712
 - [24] Toledo, F. J., Martínez, J. J., Garrigós, F. J., Ferrández, J. M.: "Image processing with CNN in a FPGA-based augmented reality system for visually impaired people", *8^o Int. Work-conference on Artificial Neural Networks, IWANN 2005*, LNCS 3512, June 2005, Vilanova i la Geltrú, Spain, pp. 906-912
 - [25] Toledo, F. J., Martínez, J. J., Garrigós, F. J., Ferrández, J. M.: "An augmented reality visual prosthesis for people affected by tunnelling vision", *1er Int. Work-conference on the Interplay between Natural and Artificial Computation, IWINAC 2005*, LNCS 3561, June 2005, Las Palmas de Gran Canaria, Spain, pp. 319-326.
 - [26] Toledo, F. J., Martínez, J. J., Garrigós, F. J., Ferrández, J. M.: "Reconfigurable hardware for an augmented reality application", *SPIE Microtechnologies for the New Millennium, conf. on bioengineered and bioinspired systems*, vol. 5839, May 2005, Sevilla, Spain, pp. 389-397.
 - [27] Toledo, F. J., Martínez, J. J., Garrigós, F. J., Ferrández, J. M.: "FPGA Implementation of an augmented reality application for visually impaired people", *15^o Int. Conf. on Field Programmable Logic and Applications, FPL 2005*, August 2005, Tampere, Finland (accepted, in press).

Virtual trainers using low-cost platforms TIC2003-08933-C02

Luis Pastor Pérez*
Universidad Rey Juan Carlos

Antonio Pérez Ambite**
Universidad Politécnica de Madrid

Abstract

Traditionally, the continuous improvement in the computers' price-performance ratio has been fostering the development and diffusion of new application areas, which has brought about new advances in price-performance ratios. Virtual reality (VR) applications are a good example of this situation: the impact that these technologies will have into the training procedures of operators who have to carry out critical operations will be strong, given the cost and safety advantages of these approaches. Before this can happen, it is still necessary to progress in the understanding of the person-computer interaction process, in the quality of the sensory feedback returned by VR systems (essentially visual, auditive and tactile), and in the price level of VR systems and facilities.

The aim of this project is to progress in the development of virtual reality systems showing both low cost and high performance, for which two research lines will be followed:

1. The reduction of the systems' computational requirements through the development of advanced algorithms for the efficient manipulation of the geometric structures associated to the generation of images and haptic responses of 3D objects, within the time restrictions associated to the simulation processes.
2. The reduction of the hardware's cost, through the development of basic techniques for implementing VR systems on clusters made out of off-the shelf components.

Finally, two demonstration applications will be developed within the medical and industrial domains. For this purpose it will be necessary to adapt the developments described above to the particular applications, modelling also the visual and haptic response of systems.

Keywords: Virtual reality, surgical training, 3D modelling, parallel processing

* Email: luis.pastor@urjc.es

** Email: aperez@fi.upm.es

1 Project objectives

1.1 Objectives

The main objective of the coordinated project “Virtual trainers using low-cost platforms” is to contribute to the widespread use of simulation, advanced visualization and virtual reality systems by means of the development of low-cost solutions. Two lines of work will be taken for this purpose:

1. To minimize the computational load associated to the image generation process through the improvement of the current 3D model techniques.
2. To reduce the price of the systems using clusters of low-cost computers.

The evaluation of the results achieved by these two lines of work will be performed while developing prototypes which cover different approaches in the operator training field, after considering their technical viability and social and economical relevance.

- An arthroscopy trainer, which includes a simple visualization process, but with the requirements to incorporate mechanisms of interaction between surgical instrumental and anatomical tissues and to supply force feedback through haptic devices.
- An industrial operator trainer based on immersive visualization with image projection on several walls of a reconfigurable visualization CAVE-like system or wraparound screens.

These objectives have been distributed between the two subprojects of the proposal in the following way:

- Subproject no. 1: Improvement of multiresolution representation techniques and development of demonstration applications. Within Subproject no. 1, algorithms for limiting the complexity of virtual objects and environments are being developed. Also, this subproject is in charge of developing the demonstration applications, which requires the development of a certain number of algorithms and techniques for several purposes, such as collision detection, deformation of non-rigid tissues, force feedback generation, etc.
- Subproject no. 2. Utilization of clusters for simulation, immersive visualization, alignment error detection and correction in projecting mosaics. The aim of this subproject is to create a basic experimentation environment for developing parallel and distributed versions of VR algorithms, particularly in the field of visualization.

1.2 Available resources

The currently available resources allow ensuring the correct development of the tasks programmed until the end of the project, having in mind to specifically make new acquisitions when needed. In the following, we show a short compilation of the most remarkable material resources.

Virtual reality and advanced visualization.

- A RAVE-like (Reconfigurable Advanced Visualization Environment) immersive projection system
- One portable 3D projection system.
- One laboratory of virtual reality equipped with virtual reality helmets, augmented reality helmets, - 1 Dataglove, several haptic devices (1 Phantom, several Phantom Omni and 2 Impulse Engine) and 1 VLI.
- A 3D laser scanner.

Computacional resources

- Access to a Cluster with 248 IBM Blade processors (no. 341 in the Top500 supercomputers ranking) in the CeSViMa consortium (Center of Supercomputation and Visualization from Madrid).
- Several Beowulf clusters, one configured with 40 processor nodes, another one with 20, another with 4, and the last one made up of 10 biprocessor nodes.
- Several Silicon Graphics multiprocessors: Prisma (16 processors), Altix (8 processors) Origin (8 processors), Challenge (10 processors).
- A SP2 IBM multiprocessor with 10 processors.
- A Sun multiprocessor with 10 processors.

1.3 Task schedule

We have carried out a hierarchical breakdown of the performed activities following the same approach presented in the memory application. According to that, the project “Virtual trainers using low-cost platforms” has been decomposed in two subprojects:

- Subproject no. 1 (TIC2003-08933-C02-01): Development of virtual trainers in surgical and industrial environments.

In this subproject, a new level of subdivision has been performed in order to remark the significance and difference between two programmed general tasks:

- Subproject no. 1a (S1a): Development of efficient techniques for representing and visualizing 3D objects.
- Subproject no. 1b (S1b): Development of demonstration applications.

- Subproject no. 2 (TIC2003-08933-C02-02): Development of immersive visualization systems using low-cost platforms.

Besides the tasks belonging to the whole project, General Coordination (*cogen*) and Configuration and documentation management (*gescd*), in the following sections we detail the activities assigned to each one of the subprojects:

- Tasks related to Subproject no. 1a (S1a): Coordination of the research group, Selection and study of the bibliography (*slaseb*), Market analysis and equipment acquisition (*slamae*), Analysis and development of new multiresolution representation techniques considering simultaneously

geometry and texture (*tmgf*), Multiresolution techniques through the use of intermediate representations (*tmri*), Development of simplification techniques for complex scenes through object hiding (*lseo*), Model representations in artificial environments (*rmea*).

- Tasks related with Subproject no. 1b (S1b): Research group coordination (*s1aco*), Study of existing systems (*s1besd*), Market Analysis and equipment acquisition (*s1bame*), Analysis of the surgical and industrial trainers (*fæ*), Design of the surgical and industrial trainers (*jde*), Implementation of a basic application for the surgical and industrial trainers (*jce*).

- Tasks related with Subproject no. 2 (S2): Research group coordination (*s2co*), Selection and study of the bibliography (*s2aseb*), Market analysis and equipment acquisition (*s2amae*), Visualization architecture's definition (*dav*), Design and implementation of the visualization architecture (*dvr*), Adaptation of the representation techniques (*atr*), Adjustment of the alignment among the projected images (*aip*), Nodes' synchronization (*sen*), Image synthesis parallelization (*par*).

Table 1 shows the initial planning of the project that highly fits the developments executed until now.

Table 1

Table 1

Task	First year												Second year												Third year											
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
<i>Cogen</i>	.			.						.	.				.	.						.	.				.						.	.		
<i>Gescd</i>	.			.						.	.				.	.						.	.				.						.	.		
<i>S1aco</i>	.			.						.	.				.	.						.	.				.						.	.		
<i>S1aseb</i>	.			.						.	.				.	.						.	.				.						.	.		
<i>Slame</i>	.			.						.	.				.	.						.	.				.						.	.		
<i>Dav</i>	.			.						.	.				.	.						.	.				.						.	.		
<i>Dcv</i>																																				
<i>Atr</i>	.			.						.	.				.	.						.	.				.						.	.		
<i>Sen</i>	.			.						.	.				.	.						.	.				.						.	.		
<i>Pars</i>	.			.						.	.				.	.						.	.				.						.	.		
<i>S1bco</i>	.			.						.	.				.	.						.	.				.						.	.		
<i>S1esd</i>	.			.						.	.				.	.						.	.				.						.	.		
<i>S1bame</i>	.			.						.	.				.	.						.	.				.						.	.		
<i>Fae</i>	.			.						.	.				.	.						.	.				.						.	.		
<i>Fde</i>	.			.						.	.				.	.						.	.				.						.	.		
<i>Fce</i>	.			.						.	.				.	.						.	.				.						.	.		
<i>Aip</i>																																				
<i>S2co</i>	.			.						.	.				.	.						.	.				.						.	.		
<i>S2aseb</i>	.			.						.	.				.	.						.	.				.						.	.		
<i>S2amae</i>	.			.						.	.				.	.						.	.				.						.	.		
<i>Tmgt</i>	.			.						.	.				.	.						.	.				.						.	.		
<i>Tmri</i>	.			.						.	.				.	.						.	.				.						.	.		

Task	First year											Second year											Third year											
	1	2	3	4	5	6	7	8	9	10	11	1	2	3	4	5	6	7	8	9	10	11	1	2	3	4	5	6	7	8	9	10	11	12
T_{seo}	.			.					.	.	.			.	.	.					.	.				.							.	.
R_{mea}	.			.					.	.	.			.	.	.					.	.				.						.	.	

2 Level of success achieved in the project

In the case of Subproject no. 1a, the Computational Geometry group has obtained results in several aspects, giving theoretical support to the development of computer applications oriented to the representation and visualization of three dimensional structures.

- Advances in the development of speculative parallelization techniques, which allow obtaining automatically parallel algorithms from sequential algorithms. It works especially well when it is applied to geometric problems. In particular there are interesting results in the computation of convex envelopes.

- Solutions for geometric problems with applications to the mechanic engineering and architecture fields have also been obtained.

- Problems related to supply planning delivery and transport planning have also been considered, where new Voronoi diagrams take part of the solution in a decisive way.

- New results related to visibility problems that can be applied in different areas of the project, as well as in other areas such as illumination and surveillance.

- Some combinatorial problems have also been solved.

- The study of some of the existing algorithms to perform frustum culling and occlusion culling has been dealt in great depth, both for the rendering of interior and exterior scenes. The studies are focused in the algorithms of portals combined with the use of leafy BSP's to render interiors, and multiresolution techniques for representing the ground geometry for the exteriors. In this context, the PhD Thesis "A multiresolution model for plane graphs representing geographic maps" has been finished.

- A 3D chat system has been developed, where the conversations among different users take place outwards, and both frustum culling and the occlusion culling are executed through BSP's implemented using genetic algorithms.

The developments carried out by the Computational Geometry group have been complemented with the work developed in Subproject 1b, applying some of the theoretical results to the development of VR trainers. During this process, also new algorithms have been developed which have resulted in new publications and achievements. In particular:

- There is an important advance in the development of a surgical trainer for arthroscopy operation. A preliminary version of the prototype has been presented in the last conference organized by the Spanish Arthroscopy Association in Sitges, from the 25th to the 27th of may 2005, achieving an outstanding success in its presentation and raising a great excitement about the possibilities of this tool in relation to the training of arthroscopy surgeons. As a result of the first developments, an agreement was reached with the company GMV in 2004. It was focused on achieving a commercial exploitation of the trainer (more details of this agreement can be found in the next Section).

- Many efforts have been devoted to the improvement of the techniques for representing 2D and 3D objects stated by the research group in previous projects and in the quest of new and more efficient techniques from a computational point of view. One example would be the consideration of new multiresolution techniques using intermediate representations.

- The techniques for content-based 2D image retrieval previously developed by some of the members of the research group have been extrapolated to video sequences, finishing the PhD Thesis "Content-based image and video retrieval techniques on parallel architectures", reported in the Section "PhD Thesis".

Next, we explain the development progress for the tasks assigned to Subproject no. 2. Some of them entail the adaptation of representation techniques and the parallelization of image synthesis procedures using the algorithms developed by the other partners. Because of this, it was decided that, in addition to keep working on load balancing issues, to set up a system (with its corresponding prototype) for testing parallel solutions. This strategic decision is implemented using as a base the free distribution Chromium software and porting it to our system, and adapting the Softgenlock free distribution software to add its functionality to the Chromium tool. At the present moment, the last version that has been distributed of this tool (cr 1.8) is being used, so the required changes are applied to introduce the realignment and synchronization mechanisms and part of the objectives of the subproject will be covered.

To complete the activities associated to the considered objectives, we have started to work on the adjustment problem that appears when several images are combined, tackling it in a global way with the objective of achieving an automatic alignment of the projectors, both before beginning and during the projection, and correcting also in real time the small failures that could happen during the projection of real images.

Also, as a result of the collaboration between the participants in Subprojects no. 1 and no. 2 and considering the importance that it has for these type of systems, a general line of research about clusters and load balancing has been kept active all the time. In this line the following work has been done:

- Scalability model for heterogeneous clusters: a new model for heterogeneous efficiency, necessary to obtain a model of heterogeneous isoefficiency has been developed. Analytic studies complemented with implementation works in deeply heterogeneous clusters have been executed.

- Parallel computation model: the model LogGP has been generalized so it can work with heterogeneous clusters leading to the new HlogGP model.

- Load balancing: new procedures have been proposed for measuring the computational power of a node in order to decide which is the most suitable one to execute one task in a given moment. A detailed analysis of all the phases of a load balancing algorithm has been made, proposing for each case new models of work and performance metrics and obtaining the best solutions for each one of the phases.

In this course of action, the Doctoral Thesis “Load balancing in heterogeneous clusters” also reported in the Section “PhD thesis”, has been recently finished.

3 Results’ indicators

The developed work has generated numerous publications, most of them of international diffusion in relevant journals and conferences (see Annexe I). Also, a technology transfer of the results to the industrial sector has already been produced. As a result of the first developments performed by the Modeling and Virtual Reality Group (GMRV) to design and build a surgical trainer, we have contacted the Spanish company GMV to study the possibility of exploiting these results from a commercial point of view. These initial contacts have resulted in a collaboration agreement among the company GMV and the URJC and UPM universities to commercialize the system. The collaboration agreement lasts up to the end of 2005, being already foreseen its extension when the present agreement concludes [DPD05]. Furthermore, the company GMV has applied for public finance to CDTI and PROFIT programmes to subsidize the mentioned agreement [CDTI05, PROFIT05]. In this course of action, GMV has invited GMRV to participate to in new collaborations to develop a laparoscopy trainer [INFUS05].

Taking into account the analysis carried out and the acquired experience in the development of the medical trainer, it will also be possible to apply to the industrial trainer some of the techniques of representation and animation used in the surgical trainer. Then, we can say that at the present moment there is a solid base to guarantee the success of the development of the industrial trainer and it is possible to transfer the results to companies interested in the project.

Moreover, the GMRV group from URJC and UPM has obtained some funds from the Autonomous Community of Madrid to continue the research in the virtual surgical trainer area [SARA05]. Basically, the funds obtained guarantee the incorporation of a Doctor with a profile based on the simulation deformable objects with tactile feedback to the research group.

Recently, we have applied for additional funds from the Autonomous Community of Madrid to coordinate the development of I+D programs among different research groups in the Community of Madrid, where the participants in the present project have played a basic role to agglutinate and boost the presented initiative.

In the international area, the collaboration initiated with the Dipartimento di Informatica e Scienze dell’Informazione from the Genova University still continues. Professor Enrico Puppo, external participant in the project, carried out a one week visit to the Alcala University and gave a conference in the URJC.

Shortly, a visit from the Canadian professor Godfried Toussaint, from the McGill University of Montreal is expected.

Furthermore, we have contacted the Australian University of Western Sydney to set a schedule to exchange visits of researchers from both groups. Within that context, Prof. George Bryan will stay at URJC for a month by late 2005.

Also we maintain contacts with researchers from the University of Dublin, Trinity College from Ireland in tasks related to content based retrieval systems and with the Uppsala University from Sweden in the area of 3D representation.

To conclude, statistic summary of the works published until now is shown:

- International publications with impact index: 15 publications. 5 of them have been published in journals or as chapters of books with impact index. The other 10 have been published as part of proceedings. They all appear at the ISI Web of Knowledge.
- Other international publications: 10. A few of them are so recent publications that they have not been collected yet in impact indexes.
- National publications: 10

4 References

- [DPD05] Título del contrato/proyecto: Desarrollo de un Prototipo-Demostrador de Herramienta de Simulación Avanzada y Entrenamiento de Cirugía Mínimamente Invasiva
Empresa/entidad financiadora: GMV S.A. Organismos participantes: Dpto. de Tecnología Fotónica, de la Facultad de Informática de la UPM y Dpto. de Informática, Estadística y Telemática de la URJC Investigadores responsables: D. Luis Pastor Pérez y D. Angel Rodríguez Martínez de Bartolomé Duración: Febrero 2004-Diciembre 2005
- [CDTI05] Título del proyecto: "Desarrollo de un Prototipo-Demostrador de Herramienta de Simulación Avanzada y Entrenamiento de Cirugía Mínimamente Invasiva – SIME(CMI)" (Ref. 04-0698) Entidad financiadora: CDTI-MITC Entidades participantes: GMV, Dpto. de Informática, Estadística y Telemática de la Univ. Rey Juan Carlos, Dpto. de Tecnología Fotónica de la U.P.M. y Hospital Severo Ochoa de Leganés. Investigador principal del proy.: D. Jorge Potti Duración: 2005
- [PROFIT05] Título del proyecto: "Desarrollo de un Prototipo-Demostrador de Herramienta de Simulación Avanzada y Entrenamiento de Cirugía Mínimamente Invasiva – SIME(CMI)" (PROFIT 300100-2004-21) Entidad financiadora: MEC-PROFIT Entidades participantes: GMV, Dpto. de Informática, Estadística y Telemática de la Universidad Rey Juan Carlos, Dpto. de Tecnología Fotónica de la U.P.M. y Hospital Severo Ochoa de Leganés. Investigador principal del proy.: D. Jorge Potti Duración: Jun. 2004 – Jun. 2005

[INFUS05] Título del proyecto: "Investigaciones clave hacia un futuro simulador de cirugía mínimamente invasiva de altas prestaciones" Entidad financiadora: MEC-PROFIT Entidades participantes: GMV, CCMI, CIDAUT, SoftMed, Dpto. de Informática, Estadística y Telemática de la Universidad Rey Juan Carlos y Dpto. de Tecnología Fotónica de la U.P.M. Investigador principal del proy.: D. Jorge Potti Duración: 2006-2007

[SARA05] Título del proyecto coordinado: "SARA: Simulador Artroscópico Avanzado." Entidad financiadora: Consejería de Educación de la Comunidad Autónoma de Madrid Entidades participantes: Dpto. de Informática, Estadística y Telemática de la Univ. Rey Juan Carlos, Dpto. de Tecnología Fotónica de la U.P.M. y Hospital Severo Ochoa de Leganés. Investigador principal del proy.: D. Luis Pastor Pérez Duración: 2005

Annex I: List of publications

- J. L. Bosque, L. Pastor. Theoretical analysis of scalability on heterogeneous clusters. Proceedings of the 4th IEEE/ACM International Symposium on Cluster and the Grid. Editorial: IEEE Computer Society. Abril 2004. pp 285 – 292. ISBN: 0-7803-8431-8.
- J. L. Bosque, L. Pastor. HLogGP: A new parallel computational model for heterogeneous clusters. Proceedings of the 4th IEEE/ACM International Symposium on Cluster and the Grid. IEEE Computer Society. Abril 2004. pp 403 – 410. ISBN: 0-7803-8431-8.
- J. L. Bosque, D. Gil, L. Pastor. Dynamic load balancing algorithms in heterogeneous clusters. Proceedings of the IASTED International Conference on Parallel and Distributed Computing and Networks. Editorial: ACTA Press. Febrero 2004. pp 37 – 42. ISBN: 0-88986-385-7. ISSN: 1027-2666.
- M. Beltrán, J. L. Bosque. Predicting the response time of a new task on a Beowulf cluster. Parallel Processing and Applied Mathematics: 5th International Conference, PPAM 2003. Revised Papers. Editorial: Lectures Notes in Computer Science. Springer-Verlag Heidelberg. Volumen: 3019 / 2004. pp 145 - 152. Abril 2004. ISBN: 3-540-21946-3. ISSN: 0302-9743
- M. Beltrán, A. Guzman, J. L. Bosque. Dynamic tasks assignment for real heterogeneous clusters. Parallel Processing and Applied Mathematics: 5th International Conference, PPAM 2003. Revised Papers. Editorial: Lectures Notes in Computer Science. Springer-Verlag Heidelberg. Volumen: 3019 / 2004. pp 888 – 895. Abril 2004. ISBN: 3-540-21946-3. ISSN: 0302-9743.
- M. Beltrán, J. L. Bosque. Modelo estático de predicción de tiempos de ejecución para tareas. Actas de las XIV Jornadas de Paralelismo. Madrid, Septiembre 2003. pp 15 - 20. ISBN: 84-89315-34-5
- M. Beltrán y J. L. Bosque. Estimating a workstation CPU assignment with the DYPAP monitor. Proceedings of the 3rd International Symposium on Parallel and Distributed Computing. Editorial IEEE Computer Society.
- M. Beltrán, J. L. Bosque. Predicción de la asignación de CPU en sistemas basados en clusters o tecnología Grid. Actas de las XV Jornadas de Paralelismo. Almería, Septiembre 2004.
- J. L. Bosque, L. Pastor. HLogGP: un nuevo modelo de computación paralela para clusters heterogéneos. Actas de las XV Jornadas de Paralelismo. Almería, Septiembre 2004.

- M. Beltrán, J. L. Bosque, A. Guzmán. Resource Dissemination policies on Grids. On the Move to Meaningful Internet Systems 2004: OTM 2004 Workshops: OTM Confederated International Workshops and Posters, GADA, JTRES, MIOS, WORM, WOSE, PhDS, and INTEROP 2004, Agia Napa, Cyprus, October 25-29, 2004. Proceedings. Editorial: Lectures Notes in Computer Science. Springer-Verlag Heidelberg. pp 135 – 144. ISBN: 3-540-23664-3
- Oscar Robles, José L. Bosque, Luis Pastor y Ángel Rodríguez. Performance Analysis of a CBIR System on Shared Memory and Heterogeneous Clusters. IEEE International Workshop on Computer Architectures for Machine Perception CAMP'2005, pp. 309-314. Palermo, (Italia), 4 – 6 Julio 2005. ISBN: 0-7695-2255-6
- Marta Beltrán, Jose L. Bosque y A. Guzmán. Resource Dissemination policies on Grids. 5th IEEE/ACM International Conference on Cluster Computing and Grids, (pendiente de publicación). Cardiff (Reino Unido), Mayo 2005.
- Marta Beltrán, Jose L. Bosque y A. Guzmán, Initiating Load Balancing Operations. 11th IEEE/ACM International Euro-Par Conference 2005, (pendiente de publicación). Lisboa (Portugal) 31 Agosto al 2 de Septiembre, 2005.
- Diego R. Llanos, David Orden y Belén Palop. “Meseta: A new scheduling strategy for speculative parallelization of randomized incremental algorithms” HPSEC-05 (7th Workshop on High Performance Scientific and Engineering Computing), hold in conjunction with ICPP-05, Noruega, Jun. 2.005, IEEE Press
- Arturo González-Escribano, Diego R. Llanos, David Orden, Belén Palop, "Parallelization alternatives and their performance for the Convex Hull Problem" Applied Mathematical Modelling, special issue in Parallel and Distributed Computing. Elsevier ISSN: 0307-904X.
- Raquel Viana, Paola Magillo, Enrico Puppo, Pedro A. Ramos, "Multi-VMaP: a Multi-Scale Model for Vector Maps", Geoinformática An International Journal on Advances of Computer Science for Geographic Information Systems. Springer.
- M. García, L. Pastor y A. Rodríguez., "An Adaptive Multiresolution Mass-spring Model" Visual Communications and Image Processing 2005, organizado por el SPIE y por el IEEE. Beijing, China, Jul. 2.005
- C. Illana, J. Potti, J. M. Fernández, L. Pastor, J. M. Espadero, S. Bayona, A. Rodríguez, G. Triviño, "Entrenador-Simulador Avanzado para el Entrenamiento de la Artroscopia insightMIST", Española de Investigaciones Quirúrgicas (Span. J. Surg. Res.) Vol. VII, N°: 4 pág. 217, 2004 ISSN: 1139-8246
- J.M. Fernández, L. Pastor, J.M. Espadero, S. Bayona y G. Triviño, “SARA: Simulador Artroscopico Avanzado” Revista Española de Investigaciones Quirúrgicas Vol.: VII N°: 4. Páginas: 4 Fecha: 2004 ISSN 1139-8264
- P. Toharia, O. Robles, L. Pastor y A. Rodríguez, "XML Specification for AVI Files in a Content-Based Video Retrieval System", Proceedings of the 4th IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP 2004), J. J. Villanueva, Ed., pág. 374-378, ACTA Press, ISBN 0-88986-454-3, Marbella, Spain, Sep. 2004
- O. Robles, P. Toharia, L. Pastor y A. Rodríguez, "Automatic Video Cut Detection using Adaptive Thresholds", Proceedings of the 4th IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP 2004), J. J. Villanueva, Ed., pág. 517-522, ACTA Press, ISBN 0-88986-454-3, Marbella, Spain, Sep. 2004
- Oscar D. Robles, P. Toharia, A. Rodríguez y L. Pastor, "Towards a Content-Based Video Retrieval System using Wavelet-Based Signatures", Proceedings of the 7th IASTED International Conference on Computer Graphics and Imaging (CGIM 2004), M. H. Hamza, Ed., pág. 344-349, ACTA Press, ISBN 0-88986-418-7, Kauai, Hawaii, USA, Ago. 2004

- Marcelo Cintra, Diego R. Llanos, Belén Palop, "Speculative parallelization of a Randomized Incremental Convex Hull Algorithm", ICCSA 2004, CGA workshop (Computational Geometry and Applications), LNCS 3045, pages 188-197, Springer-Verlag, ISSN 0302-9743, Perugia, Italia; May. 2004
- M. Abellanas C. Moreno Jiménez. Geometric Graphs Realization as Coin Graphs. 4th Proc. INTERNATIONAL WORKSHOP on COMPUTATIONAL GEOMETRY AND APPLICATIONS (CGA'04). Perugia Fecha: 2004
- M. Abellanas, F. Hurtado, B. Palop. Transportation network Voronoi Diagrams (.ps.gz) (Slides). International Symposium on Voronoi Diagrams in Science and Engineering, Tokio, Fecha: 13-15 September 2004.
- Manuel Abellanas, Ferran Hurtado, Alfredo Garcia Olaverri, David Rappaport, Javier Tejel "Moving Coins. Japan Conference on Discrete and Computational Geometry." Tokyo, , Fecha: 8-11 octubre, 2004.
- M.Abellanas, P.Bose, A. Garcia, F. Hurtado, P. Ramos, E. Rivera-Campo, J. Tejel, "On Local Transformations in Plane Geometric Graphs Embedded in Small Grids." INTERNATIONAL WORKSHOP on COMPUTATIONAL GEOMETRY AND APPLICATIONS (CGA'04), Perugia, 2004.
- M. Abellanas, M. Claverol, F. Hurtado, Point set stratification and minimum weight structures, 20th European Workshop on Computational Geometry, Sevilla, 2004
- Ferran Hurtado, Mercè Mora, Pedro A. Ramos and Carlos Seara, "Separability by two lines and by flat polygons". Discrete Applied Mathematics 144 (1-2), p. 110-122, 2004.
- Arturo González-Escribano, Diego R. Llanos, David Orden y Belén Palop "Ejecución paralela de algoritmos incrementales aleatorizados", XI Encuentros de Geometría Computacional, Santander, España; Jun. 2005
- M. García, L. Pastor y A. Rodríguez, "Estudio de un modelo adaptativo multirresolución basado en la dinámica masa-muelle", XI Encuentros de Geometría Computacional, Santander, España; Jun. 2005
- O. Robles, P. Toharia, A. Rodríguez, L. Pastor; "Using Adaptive Thresholds for Automatic Video Cut Detection"; Proc. of TRECVID 2004 Workshop, 10 pág. <http://www.itl.nist.gov/iaui/894.202/projects/tvpubs/tvpapers04/urjc.pdf>, Gaithersburg, Maryland, EEUU; Nov. 2004
- C. Illana, J. Potti, J. M. Fernández, L. Pastor, J. M. Espadero, S. Bayona, A. Rodríguez, G. Triviño, "Entrenador-Simulador Avanzado para el Entrenamiento de la Artroscopia insightMIST", 4º CAOS-España Computer Assisted Orthopedic Surgery, Valladolid, España, Nov. 2004
- Pedro Díaz, Diego R. Llanos, Belén Palop; "Parallelizing 2D-Convex Hulls on clusters: Sorting matters"; Proceedings XV Jornadas de Paralelismo, pág. 247-252, ISBN 84-8240-714-7, Almería, España; Sep. 2004
- Antonio Guzmán y Marta Beltrán "Design and implementation of image activity measure on FPGA" Proceedings of the DCIS 2004 (ISBN 2-9522971-0-X); Burdeos (Francia) Nov. 2004
- Antonio Guzmán y Marta Beltrán "Satellite On-Board Image Compression Adviser" Proceedings of the IEEE ISSPIT 2004, Roma (Italia) Diciembre del 2004
- M. Abellanas, S. Canales y G. Hernández. "Buena iluminación." Actas de las IV Jornadas de Matemática Discreta y Algorítmica, pp. 239-246; Cercedilla. Sep.2004.

Annex II: PhD Theses

Title: “Content-based image and video retrieval techniques on parallel architectures”

Author: Oscar D. Robles Sánchez

University: Politécnica de Madrid

Center: Facultad de Informática

Date: December 2004

Director: D. Luis Pastor and D. Ángel Rodríguez

Mark: A cum laude unanimously

Title: “A multiresolution model for plane graphs representing geographic maps”

Author: Raquel Viaña

University: Alcalá de Henares

Center: Informática

Date: November 2004

Director: D. Enrico Puppo and D. Pedro Ramos

Calificación: Sobresaliente cum laude por unanimidad

Title: “Equilibrio de carga en clusters heterogéneos”

Author: Marta Beltrán Pardo

University: Rey Juan Carlos

Center: Escuela Superior de Ciencias Experimentales y Tecnológicas

Date: July 2004

Director: D. José Luis Bosque Orero

IVEM: Interfaces de Visión para Entornos Médicos TIC2003-00654

Jordi Vitrià Marca *
Centro de Visión por Computador
Universitat Autònoma de Barcelona
08193 Cerdanyola, Barcelona

Abstract

Computer Vision will be in a near future a key technology in our living and/or working environment. This project explores two uses of computer vision in this context: (a) the recognition of people and things to determine the context and the intention of an interaction, and (b) the use of automatic image analysis techniques for advanced visualization. In order to design and validate the developed technologies from a user-centered point of view, we have chosen a specific scenario and use case. The scenario is located in a medical environment: the intervention room in the Hemodynamics department of a hospital. This scenario has two interesting characteristics: there is an intensive use of multimedia digital information (generated on-line and off-line), and users cannot access information using the classical metaphors.

Keywords: Computer Vision, Face Recognition, Visualization, Medical Imaging.

1 Project Objectives

IVEM is a 3 year project, with start date 2003-12-01 and end date 2006-11-30. The total project funding is 126.400 Euro. There are 11 people involved in the project (5 PhD and 6 PhD students).

The specific project objectives were grouped around two general objectives corresponding to the research lines of the groups involved in the project:

- To analyze and to understand which are the most important requirements for a computer vision system that is supporting an interaction task that is done by a user that is performing a complex task and has a serious limitation for using the classical interaction techniques, and to develop a set of prototypes based on vision interfaces that allow the access to information in a natural way,

*Email: jordi.vitria@uab.es

- To propose advanced methods to extract and fuse multimodal medical information, and to design and develop an advanced system to visualize information about the patient (during an intervention) to perceive it in a natural and intuitive way.

These proposed objectives were (following reviewers' recommendations) slightly modified at the beginning of the project. Basically, changes have affected the use and validation of the developed technologies in the real medical environment. We have continued our research in the first objective but we have limited its application in a complex scenery. On the other hand, we have concentrated our efforts on the second objective, where research has been done in close collaboration with users.

1.1 Face detection and recognition

A long-term goal in computer-human interaction research is to approach the naturalness of human-human interaction. Automatic face detection and recognition is a critical component in this domain. For example, person and face detection allow applications to be aware when there is a person near the computer. However, research in this field is still young and relatively few robust applications exist. In this project we propose to advance the understanding of this problem along two research lines:

- (Scientific objective) To develop new classification methods and architectures for robust real-time face detection and classification in non controlled environments.
- (Scientific objective) To explore new visual cues for person identification in friendly environments.
- (Technological objective) To develop demonstration prototypes and to test the methods following a strict methodology and using a large set of data.

1.2 Analysis of IVUS images for intraoperative control

IVUS is a powerful new imaging modality for analysis and diagnosis of coronary vessels. In this modality, we obtain a cross-sectional view of the artery by inserting a catheter into it. The structure and composition of atherosclerotic plaques play important roles in coronary artery disease study and in the outcomes of coronary interventions. In Figure 1(a) we can see an IVUS image with the catheter in its center represented by a circle with high intensity and a point (the tip) in its center. Around the catheter there is the blood in the lumen of the vessel. In this picture there are two kinds of biological structures: the adventitia tissue, represented by a big region of high intensity and fine texture, and two calcium plaques, represented by thin regions of intensity equal to or greater than the adventitia and with dark regions beyond it. We can also see in Figure 2.b the operation room where IVUS images are used to control interventions.

Our objectives regarding IVUS image analysis are:

- 3D reconstruction of the coronary vessel. Intravascular ultrasound images are acquired during a pullback of catheter through the vessel, giving an internal view of the vessel. In order to locate IVUS images in space we need a 3D reconstruction of the catheter trajectory by using an alternative imaging technique: X-ray images (angiograms).



Figure 1: IVUS image and hemodynamics operation room.

- IVUS and Angiography image registration. When observing an intravascular ultrasound stack of images, it is difficult to figure out the image position and extension with regard to the vessel parts and ramifications, and misclassification or misdiagnosis of lesions is possible. The objective is to develop a computer vision technique to fuse the information from angiogram and intravascular ultrasound images defining the correspondence of every ultrasound image with a corresponding point of the vessel in the reconstructed 3D angiogram.
- Automatic segmentation of IVUS images. The composition and structure of the vessel change with age, hypertension, diabetes mellitus and many other factors. Until this moment, it is feasible to discriminate different morphological structures of the vessel as calcium deposits, fatty, fatty fibrous and fibrous materials. Today, it is not completely clear what the vulnerable plaque is. Textural analysis is one of the closest related processes in computer vision to the physicians expertise when dealing with IVUS images; due to the fact that plaque discrimination is performed using, mostly, morphological issues. Visual textural analysis is a difficult, subjective and time-consuming process highly depending on the specialist. Therefore, there is an increasing interest of the medical community in developing automatic tissue characterization procedures of IVUS images.
- Content-based automatic retrieval of IVUS images. The increasing volume of image data in several applications makes it necessary to use efficient and meaningful retrieval systems of images. The poor performance of textual entries for representing this information leads to the necessity of using content-based image retrieval (CBIR) systems. We propose to apply this technique to a database of IntraVascular UltraSound (IVUS) images to be used in during case-based reasoning sessions.

2 Level of Success

In order to assure real conditions and fair evaluation of the developed techniques, we have been collaborating with two external institutions: "Germans Trias i Pujol" Hospital, that is a leading institution in the use of IVUS images, and la Caixa, a savings bank that is interested

in facial recognition for marketing purposes (not security!). In both cases, but specially in the first case, the collaboration has been excellent.

We have organized periodical working sessions with the hemodynamics department of the "Germans Trias i Pujol" Hospital, and we have got access to all their clinical data.

Our work has been presented to the world leading company in IVUS instrumentation: Boston Scientific (MA, USA). We have met two times with Dr. Tat-Jin, director of the Research Division of Boston Scientific in order to evaluate our work for technology transfer. We have signed a non-disclosure agreement and they are testing our software in order to evaluate potential benefits for their instruments. They have also provided, with no cost, the latest available technology to test our developments.

During the development of the project we have started two new collaborations: the *Hospital del Valle de Hebrón de Barcelona* and *Institut Cartogràfic de Catalunya*. In both cases the main interest is to use the developed technologies in their specific problems. In the first case we are analyzing wireless capsule endoscopy images in order to characterize intestinal motility. In the second case the objective is to develop a real time traffic sign recognition system for mobile mapping. In both cases we have signed a contract for technology transfer.

The main scientific achievements can be summarized in the following points:

- From a theoretical point of view, our main achievements are concentrated in the field of pattern recognition. We have proposed a new methodology, based on boosting, for nearest neighbor classification [3, 5, 17, 21, 35, 36, 37]. We have also continued our research on the use of Independent Component Analysis for classification [4, 7].
- We have continued our research on face detection, tracking and recognition. Our group has been invited to participate in the Face Recognition Grand Challenge ¹, organized by the National Institute of Standards of the EEUU. FRGC will develop new face recognition techniques and develop prototype systems while increasing performance by an order of magnitude. The FRGC is open to face recognition researchers and developers in companies, academia, and research institutions. We have presented our results in the FRGC Workshop [28] and our paper was finalist for the FRGC Award.
- We have proposed to the face recognition community a new research line that has not been explored until now: the use of external face features for face classification, that has been applied to face verification, face recognition and gender recognition [6, 16, 20, 27, 32, 33, 34].
- We have continued our research in the processing of IntraVascular UltraSound (IVUS) images. We have developed a new system to segment these images in order to automatically classify different tissues [22, 23, 24].
- We have developed a new method to analyze vascular dynamics in order to eliminate artifacts from longitudinal visualization and to do a 3D reconstruction and multimodal image fusion [8, 9, 12, 14, 38, 25, 39].
- We have added a new image modality (CT) to validate the fusion between IVUS and angiography [31].

¹<http://www.frvt.org/FRGC/>

- We have developed a content-based IVIS image retrieval system [11, 13, 15, 26, 40].
- We have successfully tested the intraoperative system in real interventions.
- Finally, we have applied the developed techniques to other problems: wireless capsule endoscopy [10, 29, 30], and real time detection of traffic signals [18, 19].

3 Results

During the last two years, three young researchers of our group have obtained the PhD degree: David Guillaumet, *Statistical Local Appearance Models for Object Recognition*; Debora Gil, *Geometric Differential Operators for Shape Modelling*; and Oriol Pujol, *A Semi-Supervised Statistical Framework and Generative Snakes for IVUS Analysis*.

In the last two years the scientific production of the group has been the following one: Books: 2; Journals: 15; Conference Proceedings: 23 (some of them as book chapters).

Our research group has organized during this period three scientific events:

- The 3rd International Conference on Functional Imaging and Modeling of the Heart, June 2-4, 2005, in Barcelona [2].
- The 2nd Advanced Pattern Recognition and Medical Imaging Workshop in Bellaterra (Barcelona), September, 2005.
- The 7th Catalan Conference on Artificial Intelligence (CCIA'2004), October, 21-22, 2004, in Barcelona [1].

We also have continued our research collaborations (that have resulted in joint publications in almost all the cases) with several groups:

- Dra. L. Kuncheva, School of Informatics, University of Wales. We have been working in multiple classifier systems. Two members of our group have visited her laboratory for several months.
- Dr. J. Salas, Instituto Politécnico Nacional (IPN) of México in Querétaro, México. We have been working in motion detection and analysis for wireless capsule endoscopy images. Dr. Salas is spending a sabbatical year in our laboratory.
- Dr. M. Bressan. Xerox Research Center Europe, Grenoble, France. We have been working in the development of face recognition systems.

In September 2004 we presented a Marie Curie RTN Proposal to the UE. This proposal (with 15 partners) was led by the University of Wales (UK), the Lincoln University (UK) and our university. The objective of the proposal, that was called INSIGHT, was to develop an excellence network to support research in the area of pattern recognition and medical imaging. The proposal was not accepted, but this year will be presented again after some changes.

References

- [1] J. Vitrià, P. Radeva and I. Aguiló (eds.). *Recent Advances in Artificial Intelligence Research and Development*, Volume 113, Frontiers in Artificial Intelligence and Applications, IOS Press , 2004. ISBN: 1 58603 466 9.
- [2] A. Frangi, P. Radeva, A. Santos and M. Hernandez (eds.), *Functional Imaging and Modeling of the Heart*, Springer Verlag, Berlin, LNCS-3504, ISSN 0302-9743, 2005.
- [3] D. Masip, L. I. Kuncheva and J. Vitrià, *Ensemble-based method for linear feature extraction for two class problems*, Accepted for publication in Pattern Analysis and Applications.
- [4] J.Vitrià, M.Bressan, P.Radeva, *Bayesian classification of cork stoppers using class-conditional independent component analysis*, Accepted for publication in IEEE Trans. on Systems, Man and Cybernetics (Part C).
- [5] D.Masip, J.Vitrià. *Boosted Discriminant Projections for Nearest Neighbour Classification*. Accepted for publication in Pattern Recognition.
- [6] D.Masip, M.Bressan, J.Vitrià. *On the Feature Extraction for Gender Recognition Using the Nearest Neighbour Approach*, Vol. 20, n. 5, pp. 561-576 (2005) International Journal of Intelligent Systems.
- [7] M.Bressan, D.Guillamet, J.Vitrià. *Multiclass Object Recognition using Class-Conditional Independent Component Analysis*, Cybernetics and Systems, Volume 35, Number 1 / January-February 2004, pp. 35 - 61.
- [8] Debora Gil, Petia Radeva. *A Regularized curvature flow designed for a Selective Shape Restoration*, IEEE Transactions on Image Processing Volume 13, Number 11, November, 2004, pp.1444-1458. (SCI: 2,642)
- [9] Debora Gil, Petia Radeva. *Shape Restoration via a Regularized Curvature Flow*, International Journal of Mathematical Imaging and Vision, Volume 21, pp.205-223, 2004. (SCI: 0,617)
- [10] F. Vilariño, L. Kuncheva and P. Radeva, *ROC curves and Video Analysis Optimization in Intestinal Capsule Endoscopy*, Pattern Recognition Letters, (accepted).
- [11] J. Amores and P. Radeva, *Registration and Retrieval of Highly Elastic Bodies using Contextual Information*, Pattern Recognition Letters, (accepted).
- [12] D. Gil and P. Radeva, *Extending Anisotropic Operators to Recover Smooth Shapes*, Computer Vision and Image Understanding (in press).
- [13] J. Amores and P. Radeva, *Elastic Matching and Retrieval of IVUS Images Using Contextual Information*, International Journal on Intelligent Systems (in press).
- [14] D. Gil, P. Radeva, *Inhibition of False LandMarks*, Pattern Recognition Letters (accepted).
- [15] J. Amores, N. Sebe, P. Radeva, *Boosting the distance estimation. Application to the K-Nearest Neighbor Classifier*, Pattern Recognition Letters (accepted).

- [16] D.Masip, M.Bressan, J.Vitrià. *Feature extraction for real time face detection and classification*, Accepted for publication in EURASIP Journal of Applied Signal Processing.
- [17] D.Masip, J.Vitrià. *Boosted Linear Projections for Discriminant Analysis*, In J. Vitrià, P. Radeva and I. Aguiló (eds.). Recent Advances in Artificial Intelligence Research and Development, IOS Press, Amsterdam, 2004.
- [18] A.Lapedriza, J.Vitrià. *Open N-Grams and Discriminant Features in Text World: An Empirical Study*, In J. Vitrià, P. Radeva and I. Aguiló (eds.). Recent Advances in Artificial Intelligence Research and Development, IOS Press, Amsterdam, 2004.
- [19] X.Baró, J.Vitrià. *Fast Traffic Sign Detection on greyscale images*, In J. Vitrià, P. Radeva and I. Aguiló (eds.). Recent Advances in Artificial Intelligence Research and Development, IOS Press, Amsterdam, 2004.
- [20] D.Masip, J.Vitrià. *On feature extraction for gender recognition using the nearest neighbor approach*, in I.Aguiló, Ll.Valverde, M.T.Escrig (eds), Artificial Intelligence Research and Development, IOS Press, Amsterdam, 2003, pp. 178-188.
- [21] P.Radeva, J.Vitrià. *Discriminant Projections Embedding for Nearest Neighbour Classification*, in Progress in Pattern recognition, Image Analysis and Applications, Lecture Notes in Computer Science 3287, Springer Verlag, pp. 312-319, 2004.
- [22] O.Pujol, P.Radeva, J.Vitrià, J.Mauri. *Adaboost to Classify Plaque Appearance in IVUS Images*, in Progress in Pattern recognition, Image Analysis and Applications, Lecture Notes in Computer Science 3287, Springer Verlag, pp. 312-319, 2004.
- [23] O. Pujol and P. Radeva, *Texture Segmentation by Statistic Deformable Models*, International Journal of Image and Graphics, Volume 4, Number 3, July, 2004, pp.433-452
- [24] D. Gil, A. Hernandez, A. Carol, O. Rodríguez and P. Radeva, *A Deterministic-Statistic Adventicia Detection in IVUS Images*, LNCS-3504, Springer Verlag, FIMH'2005, 2005, pp. 65-74.
- [25] M. Rosales, P. Radeva, O. Rodriguez, and D. Gil, *Supresión of IVUS Image Rotation. A Kinematic Approach*, LNCS-3504, Springer Verlag, FIMH'2005, 2005, pp. 359-368.
- [26] J. Amores, N. Sebe, P. Radeva, *Efficient object-class recognition by boosting contextual information*, In Editors: Jorge S. Marques, Nicolás Pérez de la Blanca, Pedro Pina, Pattern Recognition and Image Analysis, Lecture Notes in Computer Science, Publisher: Springer-Verlag GmbH, ISSN: 0302-9743, Volume 3522. 2005 .
- [27] A.Lapedriza, J.Vitrià, *The Contribution of External Features to Face Recognition*, In Editors: Jorge S. Marques, Nicolás Pérez de la Blanca, Pedro Pina, Pattern Recognition and Image Analysis, Springer-Verlag GmbH, ISSN: 0302-9743, Volume 3522 / 2005.
- [28] A. Lapedriza , D. Masip and J. Vitrià *Are External Face Features Useful for Automatic Face Classification?* . IEEE Workshop on Face Recognition Grand Challenge Experiments, 21 June 2005, In conjunction with CVPR 2005, San Diego, CA

- [29] P. Spyridonous, F. Vilariño, J. Vitrià and P. Radeva, *Identification of Intestinal Motility Events of Capsule Endoscopy Video Analysis*. Advanced Concepts for Intelligent Vision Systems Sept 20-23, 2005, University of Antwerp, Antwerp, Belgium (accepted).
- [30] P. Spyridonous, F. Vilariño, J. Vitrià and P. Radeva, *"Experiments with SVM and stratified sampling with an imbalanced problem: Detection of intestinal contractions*. ICAPR'2005, 3rd International Conference on Advances in Pattern Recognition, Bath, United Kingdom, 22-25 of August 2005 (accepted).
- [31] J. Garcia, F. Carreras, P. Radeva, *"A Physics-Based Validation Technique to Compare Global versus Local Tagged MRI Analysis*. Computers in Cardiology, September 26, 2005, Lyon, France (accepted).
- [32] A. Lapedriza, D. Masip, J. Vitrià. *The contribution of external features to face recognition*. Conference on Pattern Recognition and Image Analysis, June 7-9, 2005, Estoril, Portugal.
- [33] D.Masip, M.Bressan, J.Vitrià. *Classifier combination applied to real time face detection and classification*. AVR2004, Barcelona, February 3-4, 2004.
- [34] D.Masip, J.Vitrià. *Real Time Face Detection and Verification for Uncontrolled Environments*. Second COST 275 Workshop Biometrics on the Internet: Fundamentals, Advances and Applications, pp 55-58 . Vigo (Spain), March 25-26, 2004.
- [35] D.Masip, J.Vitrià. *Boosted Linear Projections for Discriminant Analysis*. Catalan Conference on Artificial Intelligence, 2004.
- [36] P.Radeva, J.Vitrià. *Discriminant Projections Embedding for Nearest Neighbor Classification*. Iberoamerican Congress on Pattern Recognition, October 26-29, 2004, National Institute of Astrophysics, Optics and Electronics (Puebla, Mexico).
- [37] D.Masip, J.Vitrià. *Object recognition using boosted adaptive features*. ECOVISION: Early Cognitive Vision Workshop, Sabhal Mor Ostaig, Scotland, 28.5. - 1.6.2004.
- [38] G. Langs, P. Radeva, D. Rotger, F. Carreras *Building and Registering Parameterized 3D Models of Vessel Trees for Visualization during Intervention*. International Conference on Pattern Recognition 2004, Cambridge, UK 2004.
- [39] M. Rosales, P. Radeva, J. Mauri, O. Pujol, *Simulation Model of Intravascular Ultrasound Images*. Medical Image Computing and Computer-Assisted Intervention (MICCAI, 2004), Springer Verlag, LNCS 3217, pp.200-208. ISBN: 3-540-22977-9 AÑO: 2004
- [40] J. Amores, N. Sebe, P. Radeva, Th. Gevers, and A. Smeulders, *Boosting Contextual Information in Content-based Image Retrieval* in Multimedia Information Retrieval, (MIR'04), ACM, pp.31-38. ISBN: 1-58113-940-3 AÑO: 2004.

PASSIVIM: Panoramic based Summarization and Semantic Interaction with Digital Video for Multimedia Messages TIC2003-06075

Xavier Binefa Valls *
Computer Science Department
Universitat Autònoma de Barcelona

Abstract

This project aims to summarize the events present in clips coming from sports or news sequence as an augmented image. The summarization will be adapted to the capabilities of the multimedia mobile devices and networks. The user shall be provided with capabilities to view the information as a current video clip but he also will be able to retrieve semantically related visual information. To reach the proposed objectives, this project shall start from a clip acquired with a fixed camera. The project is divided into three main tasks, which are : **a)** Clustering and identification of trajectories generated from the different moving regions within the scene, **b)** representation of the clip with a panoramic image including information from the dynamic regions and from the camera movements (pan, tilt, zoom) enabling the reconstruction of the video clip at the user side and **c)**, take advantage of the representation obtained in *b* in order to classify and structure a sports video sequence by events of interest (goals, etc.). Finally, the project will add user interacting capabilities using the representations mentioned above.

Keywords:

1 Project Aims

The main goal of the PASSIVIM project is to apply computer vision technologies to develop new ways of transmission of video sequences, resulting in the bandwidth saving by transforming the sequences into panoramic images while keeping track of all the information about the moving regions in the video. This allows to transmit the video as an image and a set of information about the motion of the camera and the objects, as well as permits also the inclusion of enhanced features, such as relevance-based highlight detection in videos and interactive capabilities. The main objectives of the project are next detailed :

*Email: xavier.binefa@uab.es

1.1 Clustering and identification of motions in the sequence

Motion analysis in video sequences is the cornerstone of the PASSIVIM project. The accuracy and correctness of the motion modelling of the sequence dominant and local motions is directly related to the quality of the augmented image. Therefore our very first aim is to achieve a reliable analysis of all the existing motions along the sequence. This first goal of the project is divided in several tasks.

1.1.1 Identification of different motions in a video sequence

Firstly, the different moving regions of the sequence that obey the same motion model must be identified and separated. In order to do so, algorithms regarding motion vector fields and non parametric clustering have to be developed or adapted to our problem. The frames of the video sequence must be processed in order to find motion descriptors of the regions of the sequence; these descriptors will provide the source data to perform the identification and clustering of the moving regions.

1.1.2 Motion modelling in video sequences

The identification of these regions allows us to infer the followed motion model from the motion descriptors (vectors) in terms of a set of parameters for each of the moving regions. These regions have to be tracked for a further use in the final transmission and representation of the video sequence so tracking and parameter estimation must be used and/or developed to extract an accurate representation of each region motion and trajectory.

1.1.3 Object tracking and trajectories

Each of the objects is analyzed throughout the sequence in order to extract the trajectory of the players during a clip. This extraction provides further features such as an improved video browsing or an enhanced sports analysis, but yields new and complex problems like collisions handling when occlusions occur which are not trivial and must be addressed.

1.2 Representation of the sequence using a panoramic image

Once the motion regions have been identified the dominant motion is assumed to be the camera motion, that is to say that the largest region of the frames that follow the same motion model is considered the background, static in most of the cases. The parametric representation obtained in the previous step helps inferring the transformation between each pair of frames in order to warp all the frames to build the panoramic image.

1.2.1 Action localization

One of the enhanced features sought to achieve by our project is to know the context of a sequence in each time instant. That means to know where in the match field the action is taking place. In order to do so, an algorithm to perform global mosaic image registration must

be developed so the location of each one of the frames of the sequence in the match field can be obtained.

1.3 Addition of semantic information to the processed sequence

Taking advantage of both the motion descriptors and the parametric models extracted in the previous task, context-related information such as relevance ratio can be included to the transmission to enable further enhanced features. The main aim of this task is to analyze the retrieved information in order to provide this semantic information to the sequence, using mainly the motion of the camera (analyzing when a zoom occurs, or when the camera is pointing to a goal position in the field, for example).

1.3.1 Abstract model

A very useful application of all the previous analysis (panoramic image construction, object trajectory extraction and contextual localization) is to generate an abstract model (a match field schema) and represent all the trajectories of the objects into it. It proves to be very useful for further applications, such as match analysis, but an abstract model also becomes a crucial tool when working in multicamera environments, because all the views can be synchronized using the generated artificial field.

1.4 Transmission and user interaction

Finally, when the sequence has been processed, our objective is to show the advantages of our work by transmitting it into a mobile device (mobile phone or palm) and to compare the bandwidth required when applying our approach and when transmitting normal video. Moreover, our aim is to provide a prototype where all the motion analysis performed can be exploited in form of enhanced capabilities, such as see the players trajectories or watch the highlights of a complete game automatically.

1.4.1 Optimized transmission

The usefulness of panoramic image from a video sequence in terms of bandwidth saving is far proven, as the MPEG-4 standard also takes advantage of this fact. The whole background of a video sequence can be transmitted as a single image, with only the moving regions and players remaining to be transmitted as video.

1.4.2 Multicamera environments

As mentioned above, having an abstract model of the field and player trajectories allows us to synchronize multiple views between them, using the positions in the artificial model as a nexus. The main aim of having multiple views synchronized is to offer, for example, interactive replays of the most interesting game moments, as well as a user-friendly reverse angle, providing

the user the possibility to switch between cameras when there are occlusions and multicamera availability.

1.4.3 Automatic Highlight detection and user-friendly video browsing

The processes mentioned above can help when detecting automatically the most interesting plays of the game. A slow motion replay, a camera zoom, or the proximity of the action to one of the goals are indicators of relevance. These indicators are all detected by the previous tasks and stored as a handful of parameters, easily findable in a standard database. One of the objectives of the project is to offer a highlight-based sequence summary to the end user, taking advantage of these relevance indexes.

1.5 Project Progress

Table 1 summarizes the status of the project. In general terms, most of the project goals regarding computer vision algorithmic have been achieved, while the end-user part remains to be done in 2006. It is obviously the last part of the project because it depends on the reliability of the tasks developed in the steps 1.1, 1.2 and 1.3.

Progress of the Project Objectives			
<i>Ref.</i>	Description	Status	Remarks and Schedule
1.1	Clustering and motion Identification	100%	
1.1.1	Identification of different motions in a video sequence	100%	Goal achieved
1.1.2	Motion modelling in video sequences	100%	Goal achieved
1.1.3	Object tracking and trajectories	100%	Goal achieved
1.2	Representation of the sequence using a panoramic image	100%	Goal achieved
1.2.1	Action localization	100%	Goal achieved
1.3	Addition of semantic information to the sequence	50%	Currently working
1.3.1	Abstract model	100%	Goal achieved
1.3.1	Relevance Index Extraction	30%	Researching
1.4	Transmission and user interection	0%	2006
1.4.1	Optimized transmission	0%	2006
1.4.2	Multicamera environments	0%	2006
1.4.3	Automatic Highlight detection and user-friendly video browsing	0%	2006

Table 1: Status of the project goals at August, 2005

2 Tasks and development

In this section we detail the progress of the tasks related to each of the project goals, as well as problems encountered, changes and a brief summary of the techniques used or developed.

2.1 Clustering and identification of motions in the sequence

The objectives of this part have been accomplished. We have had to change the initial formulation in order to achieve a more reliable one, but our group has developed efficient techniques regarding identification and clustering of moving regions.

The starting point in order to solve this task was the polynomial fiber models presented in [1], but the lack of reliability of this approach in singular cases that involved zooms forced us to find an alternate method to identify different moving regions or objects. We explored an alternative by combining the optical flow techniques and the robust parameter estimators. With these methods we are able to extract an accurate parametric description of the camera motion that is fundamental to obtain a correct panoramic image.

The motion vector field proves to be a valid descriptor in order to cluster different moving regions. We first employ a traditional robust estimator (*Least Median of Squares*, LMedS henceforth) to describe the dominant model by a handful of parameters. We use a reduced affine model, that takes 4 parameters to describe the dominant motion of the sequence. From that point, different regions of interest that follow different motions can be isolated by studying the residual space of the vectors that do not follow the dominant motion model in terms of the module and angle of its difference vector. This method was presented in [10],[7] and [9].

Once the different regions are identified, an accurate tracking of the objects is performed using an adaptation of a mean-shift based algorithm presented in [13]. Problems arise when we deal with sequences that present occlusions between players, so we improved the tracking algorithms with a sports sequence management method presented in [6], capable of handling collisions.

2.2 Representation of the sequence using a panoramic image

This part is the most important of the work because it allows all the further processing. The goals have been achieved, our group has published several methods to perform robust real-time panoramic images building.

It is very costly to implement a full mosaicing algorithm that relies on a dense motion vector field. Calculating the dense field and estimating the camera parameters for each frame is an overwhelming task that cannot be carried out at real time. For this reason, our group has become expert in dealing with compressed domain data. Compressed domain, or MPEG data provides a pre-calculated sparse motion vector field that allows avoiding the costly process of calculating a dense motion vector field and estimating a parameter model from thousands of samples, permitting to extract motion descriptions and build panoramic images at real time. Mentioned papers [10],[7] and [9] also contain parts referring the panoramic image construction technique.

Experimental results have shown that our latter approach outperforms by far the starting polynomial fiber method. In figure 2 results of the panoramic image building are shown.



Figure 1: Resulting panoramic images of the application of our robust statistics algorithm to sports sequences. The lines of the field are kept perfectly straight, proving that our mosaicing technique performs flawlessly in sports sequences.

We have developed improved mosaicing algorithms by taking advantage of highly robust parameter estimators such as the *vbMDPE*[14] and a novel idea based on imaginary straight lines tracking. Our methods have been published in several conferences ([8], [4]). Our work with robust estimators has also resulted in publications in major conferences [2].

2.3 Addition of semantic information to the processed sequence

We are currently working in this task.

Algorithms of modeling the match field using a synthetic model have been developed. We are now able to generate an artificial model of the field with all the objects' trajectories embedded. This work was presented in [6].

We are currently working on extracting relevance indexes of the plays to enable a further automatic selection of highlights.

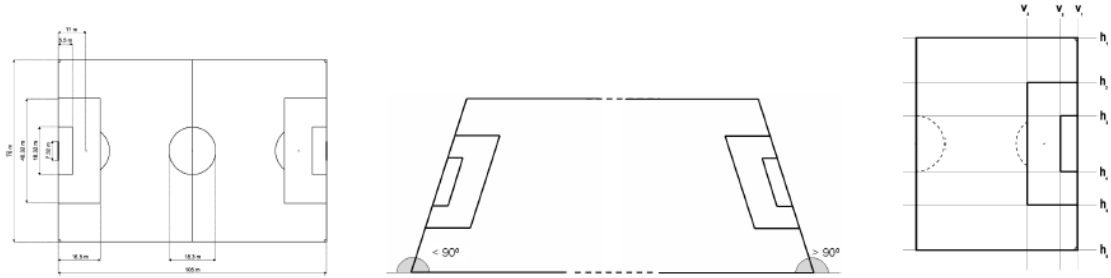


Figure 2: Schema of relationship between a match field and our abstract model.

2.4 Transmission and enhanced features of the processed sequence

This is the final part of the project, and it is scheduled for the next year. We depend on the results achieved in the previous part of the project, but the first results are promising and we think we may be able to accomplished the goals we propose :

- **An intelligent system for browsing the video sequences**, offering the possibility to overview a complete sequence by watching only the highlights, automatically extracted.
- **User interaction**. Our system will provide mechanisms for the user to choose the camera from where the action is being shown.
- **Optimized Transmission**. We aim to take advantadge of the process applied to the sequence by using the panoramic image to transmit video, with the consequent bandwidth saving.

3 Relevant scientific-technical results

3.1 Personnel involved in the project

Master Students

Luis Ferraz
 Sira Ferradans (Master Thesis at July, 2006)
 Brais Martínez (Master Thesis at July, 2006)

PhD. Students

Juan M. Sánchez (Ph.D November, 2003)
 Xavier Orriols (Ph.D February, 2004)

 Lluís Barcelo (Ph.D Thesis in December, 2005)
 Ramon Felip (Ph.D Thesis in December, 2006)

3.2 Publications

3.2.1 PhD. Thesis

- *Multiple Feature temporal models for the semantic characterization of video contents*, Juan Maria Sanchez Pujades, December, 2003
- *Generative Models for Video Analysis and 3D Range Data Applications*, Xavier Orriols i Majoral, February 2004

3.2.2 Master Thesis

- *Application of Robust Statistics and Mean Shift Analysis to Digital Video and LADAR Imaging*, Ramon Lluís Felip Rodríguez, September, 2004

3.2.3 Journal Papers

- R. Toledo, X. Orriols, J. Sanchez, X. Binefa Automatic Cataloguing of Advertisement in Magazines. *Multimedia Tools and Applications (In Press, from July 04)*
- L. Barceló, R. Felip, X. Binefa. Robust Dominant Motion Estimation using MPEG Information and a Helmholtz Principle based Parameter Estimator. *Pattern Recognition Journal (Submitted July 05)*

3.2.4 Conference Papers

- R.L. Felip, X. Binefa and J. Diaz. A New Parameter Estimator Based on The Helmholtz Principle. *IEEE Conference on Signal Processing (ICIP'05), Genova, September 2005 (In Press)*
 In this paper we present a novel parameter estimator that follows a general perception law as

an objective function, the Helmholtz principle. We present an interpretation of this postulate in statistical terms and apply it along with low order statistics in order to obtain a flexible parameter estimator with a high breakdown point, capable of suiting the specific needs of each problem. The performance of the algorithm is tested applying the estimator to find lines in synthetic data with up to 90% of noise and real LADAR images

- L. Barceló, R.L. Felip and X. Binefa. A new Approach for Real Time Motion Estimation Using Robust Statistics and MPEG Domain Applied to Mosaic Images Construction. *IEEE conference on Multimedia and Expo., Amsterdam July, 2005 (In Press)*
Dominant motion estimation in video sequence is a task that must be often be solved in Computer Vision problems but involves a high computational cost due to the overwhelming amount of data to be treated when working in image domain. In this paper we introduce a novel technique to perform motion analysis in video sequences taking advantage of the motion information of MPEG streams and its structure, using imaginary line tracking and robust statistics to overcome the noise present in compressed domain information. In order to demonstrate the reliability of our new approach, we also show the results of its application to mosaic image construction problem.
- J.M. Sanchez, R. L. Felip and X. Binefa Preatentional Filtering in Compressed Video. *IEEE conference on Multimedia and Expo, Amsterdam July 2005 (In Press)*
We propose the use of attentional cascades based on the DCT and motion information contained in an MPEG coded stream. An attentional cascade is a sequence of very efficient classifiers that reject a large number of negative candidate regions, while keeping all the positive candidates. Working directly on the compressed domain has two main advantages: computationally expensive features are already computed, and the stream is only partially decoded without the additional cost of full decompression, which will be reached by a very small number of the initial candidate regions. We have applied these concepts to skin color detection, as a pre-attentive filtering prior to face detection, and to text region detection with particular focus on license plates for vehicle identification. In both cases, a reduction of the number of candidate regions close to 95% is achieved, which turns into an enormous performance increase in video indexing processes.
- LL. Barceló, X. Binefa and J.R. Kender. Robust Methods and Representations for Soccer Player Tracking and Collision Resolution. *4th International Conference on Image and Video Retrieval (CIVR'05), Singapore, July 2005. (In Press in LNCS)*
We present a method of tracking multiple players in a soccer match using video taken from a single fixed camera with pan, tilt and zoom. We extract a single mosaic of the playing field and robustly derive its homography to a playing field model, based on color information, line extraction, and a Hausdorff distance measure. Players are identified by color and shape, and tracked in the image mosaic space using a Kalman filter. The frequent occlusions of multiple players are resolved using a novel representation acted on by a rule-based method, which recognizes differences between removable and intrinsic ambiguities. We test the methods with synthetic and real data.
- R. L. Felip, J. M. Sanchez and X. Binefa. Video summarization with moving objects in the compressed domain. *Proc. Of the IST/SPIE Conference Internet Imaging VI , Vol 5670, San Jose, pp. 143-154. January 2005*

The vast amount of video sequences available in digital format presents considerable challenges for descriptor extraction and information retrieval. The dominant motion in a video scene proves to be very important to characterize video sequences, but the cost to compute it is high when working in image domain. In this paper we present a method to extract an affine description of the global motion of a video sequence using a robust estimator and data from the compressed domain, where the motion vector field is already calculated. We also introduce a criterion to measure the reliability of the affine model estimated. Finally, we apply our approach to motion-based segmentation of video sequences and video summarization.

- LL. Barceló, X. Binefa. Contextual Soccer Detection Using Mosaicing Techniques. *4th Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA2005)*, pp. 77-84. Springer-Verlag LNCS 3522. June 2005, Estoril, Portugal

Sports Video understanding aims to select and summaries important video events that occur in only special fragments of the whole sports video. A key aspect to this objective is to determine the position in the match field where the action takes place, that is, the location context of the play. In this paper we present a method to localize where in the match field the play is taking place. We apply our method to soccer videos, although the method is extensive to other sports. The method is based on constructing the mosaic of the first sequence that we process: this new mosaic is used as a context mosaic. Using this mosaic we register the frames of the other sequences in order to put in correspondence all the frames with the context mosaic, that is, put in context any play. In order to construct the mosaics, we have developed a novel method to register the soccer sequences based on tracking imaginary straight lines using the Lucas-Kanade feature tracker and the vb-QMDPE robust estimator.

- R. L. Felip, X. Binefa. Affine Description of Motions in Compressed Domain. Pp: .93-100 In *Recent Advances in Artificial Intelligence Research and Development. Frontiers in Artificial Intelligence and Applications*, Ed. IOS Press. 2004.

This paper considers the problem of finding a robust solution to the parametric description of multiple motions in video sequences problem, using compressed domain data. Here, we present an effective way to recover an affine global motion description from the MPEG motion information using a robust parameter estimator. The error structure shown by the outliers of this first estimation allows us to cluster them into several local motions, using a non parametric estimator of the density (the mean shift) over a measure of the residuals of our estimation. Finally, results of our algorithm are shown in synthetic and real sequences.

- LL. Barceló, X. Binefa and J. R. Kender. Visual Events Detection in Well-Known Environments. Pp .217-224. In *Recent Advances in Artificial Intelligence Research and Development. Frontiers in Artificial Intelligence and Applications*, Ed. IOS Press. 2004

3.3 Collaboration with other research groups

Thanks to this project, collaboration with important international research groups has been possible. Our group has established a good relationship with Professor John Kender, from Columbia University. Our collaboration has resulted in several published papers as well as research stages of Juan M. Sanchez and Lluís Barceló in New York. Professor Kender has come several times to Barcelona to give lectures and to take part in technical committees and

thesis tribunals. Professor John Kender is a renowned researcher in the video analysis field.

We have also established collaboration links with Dr. Nicu Sebe, from the Amsterdam University. We have participated together in several Technical Committees of conferences as well as revisors for special issues of video analysis journals.

3.4 Technological Transfer

The PASSIVIM project has also brought us the opportunity to participate in other projects thanks to the prior work regarding robust statistics and object tracking performed. Specifically, we are also contributing in two more projects.

The first one is in collaboration with CIDA (*Centro de Investigación y Desarrollo de la Armada*) and it consist in automatic target detection and recognition in aerial 3D LADAR images. The algorithms of robust statistics developed in order to extract the motion parameters is our project fit perfectly in CIDA's project, due to the nature of source data and the model chosen to describe the scenes (a parametric model).

The second one aims to perform moving object tracking in infrared (IR) imagery. In this project we are working with a spanish company, *Tecnobit*. The tracking algorithms developed to follow the players in the PASSIVIM project were adapted to the IR imagery, yielding excellent results.

Both projects are in early stages but have provided some publications [11], [12], which are to be also credited to the PASSIVIM project.

References

- [1] X.Orriols, Ll. Barceló, X. Binefa. Polynomial Fiber Description of Motion for Video Mosaicing. *Proceedings of the International Conference on Image Processing, ICIP2001*
- [2] R.L. Felip, X. Binefa and J. Diaz. A New Parameter Estimator Based on The Helmholtz Principle. *IEEE Conference on Signal Processing (ICIP'05), Genova, September 2005 (In Press)*
- [3] LL. Barceló, X. Binefa and J. R. Kender. Visual Events Detection in Well-Known Environments. *Pp .217-224. In Recent Advances in Artificial Intelligence Research and Development. Frontiers in Artificial Intelligence and Applications, Ed. IOS Press. 2004*
- [4] L. Barceló, R.L. Felip and X. Binefa. A new Approach for Real Time Motion Estimation Using Robust Statistics and MPEG Domain Applied to Mosaic Images Contruction. *IEEE conference on Multimedia and Expo., Amsterdam July, 2005 (In Press)*
- [5] J.M. Sanchez, R. L. Felip and X. Binefa Preatentional Filtering in Compressed Video. *IEEE conference on Multimedia and Expo, Amsterdam July 2005 (In Press)*

- [6] LL. Barceló, X. Binefa and J.R. Kender. Robust Methods and Representations for Soccer Player Tracking and Collision Resolution. *4th International Conference on Image and Video Retrieval (CIVR'05), Singapore, July 2005. (In Press in LNCS)*
- [7] R. L. Felip, J. M. Sanchez and X. Binefa. Video summarization with moving objects in the compressed domain. *Proc. Of the IST/SPIE Conference Internet Imaging VI , Vol 5670, San Jose, pp. 143-154. January 2005*
- [8] LL. Barcelo, X. Binefa. Contextual Soccer Detection Using Mosaicing Techniques. *4th Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA2005), pp. 77-84. Springer-Verlag LNCS 3522. June 2005, Estoril, Portugal*
- [9] R. Felip. *Application of Robust Statistics and Mean Shift Analysis to Digital Video and LADAR Imaging*, Master Thesis, September, 2004.
- [10] R. L. Felip, X. Binefa. Affine Description of Motions in Compressed Domain. *Pp: .93-100 In Recent Advances in Artificial Intelligence Research and Development. Frontiers in Artificial Intelligence and Applications, Ed. IOS Press. 2004.*
- [11] R. Felip, S. Ferradans, J. Diaz-Caro and X. Binefa. Target detection in LADAR data using robust statistics. *In Proceedings of the SPIE Europe Symposium on Optics/Photonics in Security and Defence, Bruges, 2005*
- [12] B. Martínez, L. Ferraz, J. Diaz-Caro and X. Binefa. Optimization of the SSD multiple kernel tracking applied to IR video sequences. *In Proceedings of the SPIE Europe Symposium on Optics/Photonics in Security and Defence, Bruges, 2005*
- [13] D. Comaniciu, V.Ramesh and P. Meer, Kernel-based object tracking, *Transactions on Pattern Analysis and Machine Intelligence* 25, May 2003.
- [14] H. Wang and D. Suter. Variable Bandwidth QMDPE and Its Application in Robust Optical Flow Estimation, *In Proceedings ICCV03, International Conference on Computer Vision, Nice, France*

NIDO: Navegación e interacción con el usuario en el desarrollo de sistemas de información Web: Métodos, Técnicas y Herramientas

TIC2003-00369

Jesús Torres*
Coordinador

María José Escalona*
Participante

Antonia María Reina*
Participante

Abstract

NIDO (Navegación e Interacción con el usuario en el Desarrollo de sistemas de Información Web: Métodos, Técnicas y Herramientas) is a research project with two main lines of research. The first one is to continue with the research developed in the coordinated projects MENHIR (TIC97-0593-C05-03) and DOLMEN (TIC2000-1673-C06-03), while the second one is related to the opening of new research lines in order to cover specific aspects associated with the specification and development of web systems.

The rapide advance of the internet has made web systems become one of the main lines of work in companies which produce software, and also, an important research line in universities, research centres and companies' research departments. Therefore, one of the main aims of our research project is to define a methodological environment suitable to deal with navigational characteristics and user's interaction in web systems. This approach has to cover aspects of the whole life cycle, from requirements to implementation and testing.

Navigation and user interface are critical aspects in web systems. In order to make easier the application of our work, a second important goal is to give support to our approaches with tools that allow this kind of applications to be automated. Besides, in our work the definition of references and metrics to measure the grade of quality of the results is also a fundamental objective. Finally, an important characteristic of our project is the work with companies and our big interest in applying our results to real projects.

Keywords: Web engineering, advanced separation of concerns, user interfaces, user centered design

1 Project goals

When we asked for the NIDO Project, we tried to achieve mainly seven aims:

* E-mail: jtorres@lsi.us.es

* E-mail: mjescalona@us.es

* E-mail: reinaqu@lsi.us.es

- 1- To describe a modelling language to deal with the different aspects involved in web systems. For that, we thought that, on the one hand, it was necessary to study the current situation in order to evaluate which ones were the best proposals and, on the other hand, to detect the gaps in current web engineering approaches.
- 2- To define a web approach to deal with requirements in web systems. This work is a continuation of the definition of NDT (Navigational Development Techniques)[8] [9] [11] , a web proposal which started to be developed in the DOLMEN project and has to be completely defined in the NIDO project.
- 3- To define metrics to measure the quality of web systems. This work must start with an analysis of the current situation and, depending on their situation, some of them could be assumed, others could be adapted and, maybe, some new metrics should be defined to evaluate the results of the approach defined in the second goal.
- 4- To propose a framework to define the navigation aspect with the precondition of keeping the separation of concepts to decrease the complexity of the system.
- 5- To define a language to describe information in order to prototype and implement graphical user interfaces.
- 6- To develop a set of tools to support the application of our results. These tools should give support to all the life cycle of a web project.
- 7- To apply our results to real projects, mainly in projects related to e-learning and digital libraries environments, because they are complex systems which have web interfaces.

With these objectives in mind, we have proposed a methodology of work based on the division of tasks among the people in the project. We have divided the tasks into three main groups and each one is composed of different tasks. The structure of this division is showed in table 1.

BLOCK 1: Definition of the web modelling language	
	Task 1.1. Definition of our proposal NDT
	Task 1.2. Definition of the NDT process, its phases and techniques
	Task 1.3. Enrichment of NDT with quality metrics
BLOCK 2: Development of the navigation and the user interfaces	
	Task 2.1. Comparative studies of approaches to deal with aspects
	Task 2.2. Implementation of the navigation
	Task 2.3. Definition of a language to represent interfaces
BLOCK 3: Development of tools	
	Task 3.1. Development of the NDT-Tool to support NDT
	Task 3.2. Development of a tool to define and implement the navigation aspect
	Task 3.3. Development of a tool to generate user interfaces
	Task 3.4. Application of results

Table 1. Tasks in the project

With these tasks, we can cover all our objectives. Thus, table 2 shows a traceability matrix [11] which resumes the fulfilled objectives for each task. Each row represents an objective while each column corresponds to one task. An “X” expresses that the objective is partially or completely gotten with the task.

Furthermore, each task has been assigned to several researches in the group, and a schedule for all the project has been planned. So that, table 3 shows a chronogram representing the task calendar for the three years of our project.

OBJ/Task	1.1	1.2	1.3	2.1	2.2	2.3	3.1	3.2	3.3	3.4
1				X						
2	X	X					X			
3		X	X							
4				X	X					
5				X		X				
6							X	X	X	
7										X

Table 2. Traceability matrix

Task	First year												Second year												Third year																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
1.1																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								

Table 3. Chronogram of tasks during the project

2 Level of gotten aims

During the first year we have worked following the chronogram shown in table 3. For this reason, in this section we are going to present each group of tasks and to analyse how they have been developed.

2.1 Definition of the web modelling language

The first group of tasks was mainly focused on the definition of a web proposal whose development started in the Dolmen project. This proposal has been named NDT (Navigational Development Techniques) [11] and it is one of the most important results of our work.

When the project started, NDT was analysed and some comparative studies [7] to analyse the situation in web engineering were made by some of our researches. With these previous surveys, we found out that an important and very few studied aspect in web engineering was web requirements. NDT was created to give support to development teams in the requirement and analysis phases. However, previous studies were too abstract and in order to focus our work environment, we

needed to centre more the fact. At the beginning of 2004, a new more concrete survey was published in the Journal of Web Engineering [7]. This survey, developed in collaboration with Dr. Nora Koch of the University of Munich, was the start point of our work in tasks 1.1 and 1.2.

During the first year of the project, NDT has been completely defined. We have finished tasks 1.1 and 1.2 and its results were presented as part of the thesis of Dr. María José Escalona [11]. This thesis was submitted in October 2004 and it got the European mention. The development of NDT and also the European thesis opened several collaborations with companies and universities, that are going to be analysed in the following sections.

NDT is a web proposal focused mainly in the first phases of the life cycle and it covers an important gap in the web environment, the treatment of requirements. The treatment of requirements of NDT is very detailed, because it is based on a group of patterns. A pattern is a special template with specific fields for each kind of requirement.

The structured way of defining requirements in NDT allows us to apply systematic processes for the generation of models. In this sense, NDT is a powerful approach to give support to the development team. Furthermore, its processes and techniques can be applied using NDT-Tool, a tool-CASE that supports completely NDT and that allows us to obtain some results. The structured way of dealing with requirements and the support in the development process are the best features of NDT. For this reason, it has been accepted by several companies and it was applied to several real projects during these years.

2.2 Development of navigation and user interface

In order to accomplish task 2.1, a survey of the different proposals for separating concerns at the implementation level [16] has been made. Afterwards, a need for raising the level of abstraction was detected, and as result, another survey of current aspect modelling languages [19] was made. As a consequence, it was concluded that current aspects modelling languages were too close to AspectJ constructs (AspectJ is a general-purpose extension to Java for implementing aspects) and they were valuable for platforms that are able to separate concerns, but we need specify models in a platform independent way.

Therefore, in [18] the use of MDA [15] and domain specific languages for specifying concerns at the Platform Independent Modelling Level was proposed. Finally, the architecture of a framework for developing applications with such features was proposed in [17]. These previous steps are pursuing the fulfilling of task 2.2.

Also, it is important to highlight the work in the usability of user interfaces field to accomplish task 2.3. This research work has as a starting point the different references that, since 1990, stand up for the use of models and simulation to support the development of user interfaces. Concretely, the research line in our group is oriented to the application of modelling and simulation techniques to the User Centre Design with the aim of improving the usability of interfaces. Thus, the product is intended to be enhanced from the improvement of the process, using modelling and simulation as tools. Therefore, some future goals of this line are: firstly, the obtaining of suitable models to improve the understanding and the application of such processes; secondly, the experimentation with the effects that a decision related to the usability can cause during the software development, and, finally, the evaluation of the application of User Centre Design processes in organizations which develop interactive systems. In this area the publications [12] and [13] should be emphasized.

Finally, we should stick out the presentation of the Phd Thesis titled “Simulación Realista Del Comportamiento Mecánico De Telas Utilizando El Método De Elementos Finitos” by Dr. Juan Manuel Cordero in April 2005. This thesis develops a cloth simulator where the NDT

methodology has been proved. The configuration environment of the simulation is complex with multiple design and visualization aspects. This environment allows us to check and redefine some patterns of the generation of interfaces language.

With all these results, we can conclude that the tasks planned for this block has been covered. Obviously, this group of tasks is quite theoretic and it is not a finished work. Its results are the base of the next block, overall for tasks 3.2 and 3.3 that are going to be analysed in the next section.

2.3 Development of tools

The planning of the last group has changed a little, mainly the one related to task 3.4. The fulfilling of these tasks depends on the development of the previous tasks, thus, the obtaining of results depends completely on the rest of the work.

The first task of this block, the development of NDT-Tool, is a consequence of the NDT. NDT-Tool is a tool case to apply NDT. During the definition of NDT, a first version of NDT-Tool has been introduced [11] [10]. This first version had some problems because it was developed using Visual Basic and Access. Nowadays, and after checking it by means of its application to real projects, we have started a new release developed with JSP and MySQL. This new version is currently at the end of the development process and, now, about the 75% of the tool is available.

Related to task 3.2 and the advanced separation of concerns line, the framework to deal with different concerns has just started to be developed. Task 3.3 has a similar situation.

As we have mentioned previously, the planning of task 3.4 has changed a little. At the beginning, we had proposed starting with the application of results in the middle of the project. However, we have obtained several previous results, mainly in block 1. So, we have applied our results and tools in some real projects as it was shown in previous sections.

3 Results indicators

In this section, we are going to try to introduce a group of indicators to analyse the aims that have been achieved during this first period. We are going to group them depending on their nature.

3.1 Staff in training

In our project Staff in training can be divided into two main groups. The first one is composed of researchers who are working directly in the NIDO project. At the beginning, In NIDO there were four doctors: Dr. Jesús Torres, Dr. Manuel Mejías, Dr. Juan Antonio Ortega, Dr. Mariano González. The project started with four doctoral students, but nowadays, two of them, Dr. María José Escalona and Dr. Juan Manuel Cordero, have presented their thesis. Besides, during this period, two new doctoral students have been included in the project, Javier Gutierrez and Juan Antonio Álvarez.

The second group is made up of students who are working in some of the lines of the NIDO project, but they are not included in the project. In this situation, we have several final degree project students that are applying NDT to real projects.

3.2 Publications

During our first period of work, we have published several papers and results in different research forums. In table 4 we summarise these results, and in order to obtain a better understanding, they have been grouped according to their type.

Phd. Thesis	Journals	Books	International Conferences	National Conferences	Others
2	5	0	10	6	1

Some of the most relevant publications can be consulted in the reference section.

3.3 Relations with companies and patents

The high grade of applicability of our work has caused us to have good relations with companies. The first important collaboration has been with the government of Andalusia. During this year, we have applied NDT in two small projects involving the Andalusian Institute of Historical Patrimony. These projects, available on the web, deal with the management of information about the Andalusian historical patrimony via web. The first one manages the historical thesaurus, while second one deals with information about historical authors in Andalusia.

Recently, a new collaboration with the Andalusian Government, concretely with the Consejería de Cultura, has been signed. In this case, we are going to apply our results to a big project, named Mosaico, that has just started.

Another important landmark is our parallel PETRI project. This project as a main goal the development of a system to measure the grade of handicap of some patients and it is being developed with Alcer Foundation and Visual Informática S.L. In this project, NDT has been the used methodology to define requirements and to analyse the system.

Also, NDT is being assumed to be applied to web projects by several companies, like DMR Consulting.

In other line, NDT and NDT-Tool have been registered in the Intellectual Property Register in the last June.

3.4 Relations with universities and international projects

Our group, as it has been mentioned in previous sections, has several and important relations with other universities. As international collaboration, we have some important works with the University of Nice, concretely with the group of Dr. Jean Luoís Cavarero. Nowadays, we have several publications together oriented to the requirements treatment.

Another important relation is the one with the University of Munich, concretely with Dr. Nora Koch. Our recent collaborations are oriented to work in a new requirement proposal that comprises NDT and UWE, her web approach.

These previous works are also related to the collaborations with the Politechnical University of Milan, concretely with Dr. Piero Fraternali and his group.

As national collaborations, we have a good relations with the Politechnical University of Valencia, concretely with the group of Dr. Oscar Pastor and Alicante, specially with Dr. Cristina Cachero.

During 2004, several meetings have been celebrated with these groups. In February 2004, Dr. María José was working with Dr. Oscar Pastor's group in Valencia, and, in the same month, a

meeting between members of the universities of Seville, Nice and Valencia, was celebrated in Seville.

From this meeting, we started to work in the definition of a proposal for an European Project in conjunction with other five European universities: Seville, Nice, Valencia, Geneva and Florence and the University of Casablanca. This project has a main aim the development of an e-learning platform for the University of Casablanca. In October 2004, a meeting between these universities was celebrated in Casablanca. The project was presented last January and, nowadays, we are waiting for the answer.

Currently, we are developing a new European collaboration between the universities of Geneva, Nice and Seville. This project is more oriented to companies. In December, a meeting in Geneva was celebrated with ten companies from Monaco, France, Switzerland and Spain, in order to introduce the project aims. Furthermore, a new meeting with other ten companies is going to be hold in September. The idea key is to select the most interesting companies and to develop an approach to deal with the maintenance of software projects.

4 References

- [1] L.Cavarero, M.J. Escalona "Metrics for Dinamics: How to improve the Behaviour of an Object Information System". Proceedings of the 6th International Conference on Enterprise Information System (ICEIS'2004).ISBN: 972-8865-00-7.Vol. 3. pp. 344-349. Porto, Portugal. May, 2004
- [2] J.M. Cordero, J.J. Perea, J. Torres, J. A. Ortega . Simulación del comportamiento dinámico de telas usando el método de los elementos finitos. . XIV Congreso Español de Informática Gráfica. ISBN: 84-688-6998-8. Sevilla. 2004.
- [3] J.M. Cordero. Realistic Cloth Animation Using The Mass Spring Model. Game Programming Gems 5. ISBN: 1-58540-352-1.
- [4] J.J. Gutierrez, M.J. Escalona, M. Mejías, J. Torres, J.A. Álvarez .Comparative Analysis of Methodological Proposes to Systematic Generation of System Test Cases from System Requirements.Proceedings of the 3rd International Workshop on System Testing and Validation.(SV'2004). ISBN: 3-8167-6677. pp. 151-160. Paris, France. December, 2004.
- [5] J.J. Gutierrez, M.J. Escalona, M. Mejías, J. Torres. Aplicando técnicas de testing en sistemas para la difusión Patrimonial. Actas del V Congreso Nacional de Turismo y Tecnologías de la Información y las comunicaciones (TURITEC'2004). ISBN: 84-608-0181-0. pp. 237-252. Málaga, Spain. October, 2004.
- [6] M.J. Escalona, J.L.Cavarero."Building Class Diagrams Systematically". To be published in the Proceedings of the 7th International Conference on Enterprise Information Systems (ICEIS 2005). Miami, USA. May, 2005.
- [7] M.J. Escalona, N. Koch. Requirements Enginnering for Web Applications: a Comparative Study. Journal of Web Engineering. ISSN: 1540-9589.Vol. 2. pp. 193-212. Ed. Rinton Press. February, 2004
- [8] M.J. Escalona, M. Mejías, J. Torres. Developing Systems with NDT & NDT-Tool. Proceedings of the 13th International Conference on Information System Development (ISD'2004).ISBN: 9986-05-762-0.pp. 149-159. Vilnius, Lithuania. September, 2004.
- [9] M.J. Escalona, A. M. Reina, J. Torres, M. Mejías. NDT: A methodology to deal with the navigational aspect at the requirements phase. Proceedings of the Workshop in Aspect-Oriented Requirements Engineering and Architecture Design. Vancouver, Canada. October, 2004.

- [10] M.J. Escalona, J. Torres, M. Mejías, A.M. Reina NDT-Tool: A CASE Tool to deal with Requirements in Web Information Systems. Proceeding of the International Conference on Web Engineering (ICWE 2003). Lecture Notes in Computer Science (LNCS 2722). Spain, 2003.
- [11] M.J. Escalona. Modelos y técnicas para la especificación y el análisis de la navegación en sistemas software. Ph. European Thesis. Department of Computer Languages and Systems. University of Seville. Spain, 2004.
- [12] N. Hurtado, M. Ruiz, J. Torres, J. M. González. Revisión de lenguajes declarativos para la descripción de interfaces de usuario independientes del dispositivo. *Interacción Persona-Ordenador*, nº 5. 2004.
- [13] N. Hurtado, M. Ruiz, J. Torres. Towards Interactive Systems Usability Improvement through Simulation Modeling. 6th International Workshop on Software Process Simulation and Modeling. ProSim 2005. co-located with ICSE 2005. St Louis, MO. USA. ISBN 3-8167-6761-3
- [14] N. Koch. Software Engineering for Adaptive Hypermedia Applications. Ph. Thesis. FAST Reihe Softwaretechnik Vol(12). Germany, 2001.
- [15] OMG: MDA Guide, OMG doc.ab/2003-05-01.
- [16] A. M. Reina, J. Torres. "Components+Aspects: A General Overview". *Revista Colombiana de Computación*. Publicado por la Universidad Autónoma de Bucaramanga. Vol 5(1). Junio, 2004. pp 77-95. ISBN: 1657-2831.
- [17] A. M. Reina, J. Torres, M. Toro, J. A. Álvarez. "Separación de conceptos y MDA: Arquitectura de un framework". I Taller sobre Desarrollo de Software Dirigido por Modelos, MDA y Aplicaciones (DSDM'04). Celebrado junto con las Jornadas de Ingeniería del Software y Bases de Datos (2004). Málaga, España. November, 2004. Publicado por la Universidad Politécnica de Valencia. ISBN: 84-688-8870-2. Eds. Vicente Pelechano, Antonio Vallecillo, Javier Muñoz, Joan Fons.
- [18] A. M. Reina, J. Torres, M. Toro, M. J. Escalona. "Modelando aspectos con lenguajes específicos de dominio". Segundo Taller de Desarrollo de Software Orientado a Aspectos (DSOA04). Celebrado junto con las Jornadas de Ingeniería del Software y Bases de Datos (2004). Málaga, España. November, 2004. Publicado por la Universidad de Extremadura como Informe Técnico TR-23/04. pp. 91-98. ISBN: 84-688-8889-3. Eds. Lidia Fuentes, Ana Moreira, Juan Manuel Murillo.
- [19] A. M. Reina, J. Torres, M. Toro. "Towards developing generic solutions with aspects". The 5th Aspect-Oriented Modeling Workshop (AOM). Workshop held in Conjunction with the UML 2004 Conference. Lisbon, Portugal. November, 2004.

Content-Based Music Processing

(Procesamiento de Música Basado en su Contenido)

TIC2003-07776-C02

Xavier Serra Casals*
Institut Universitari de l'Audiovisual
Universitat Pompeu Fabra

Rafael Ramirez Melendez
Institut Universitari de l'Audiovisual
Universitat Pompeu Fabra

Abstract

From the beginning, the aim of the project has been to work on different aspects of content-based music processing. So far, we have studied and developed tools for musical content extraction, modeling, and processing. Specifically, we have been working on (1) the analysis and manipulation of low-level audio descriptors by spectral modeling analysis and synthesis techniques, (2) the extraction of high-level musical attributes (musical structure, transitions and articulation) from these low-level audio descriptors and (3) the use of artificial intelligence techniques such as case-based reasoning and inductive logic programming for content-based melody processing. We have also been working on the design and implementation of a special-purpose concurrent real-time programming language for music data processing

1 Project Objectives

Content-based processing is a research area which includes diverse aspects such as analysis, indexing, search and transformations of signals produced by an audiovisual source. In a musical context, content-based signal processing is a fruitful research area which has already become one of the main research areas in computer music. In the area of video and image, there are a number of active research groups, both in Spain and in Europe. In contrast, in the audio area there is much less activity. In fact the IUA coordinator group of this project is one of the few research groups which has contributed to the Mpeg-7 standard.

In this context, this project has proposed to work on different aspects of content-based music processing and develop tools for the extraction, modeling and processing of these contents. We investigate topics such as the analysis and manipulation of low-level audio descriptions, as well as

* Email: xserra@iua.upf.es

methods for extracting high-level musical attributes from the low-level attributes and the use of artificial intelligence techniques for musical analysis and transformation.

In the past, the two research groups involved in this Project have developed a tool for the analysis of musical descriptions. The tool is able to automatically extract a XML description from an audio signal input. We have also developed a synthesis tool which is able to interpret XML data in order to obtain a melody which is synthesized and stored in a file or directly sent to the sound card of a computer. This tool is based on previous work on a saxophone synthesizer. Currently, we are developing a tool for the expressive transformation of monophonic Jazz melodies.

The experience and results obtained in past collaborations of the IUA and the IIIA, place us in a privileged position for accomplishing the project objectives.

The objective of this Project is to work on different aspects of content-based audio processing. The project investigates and develops tools for musical content extraction, modeling and processing. In particular, it investigates:

1. The analysis and manipulation of low-level audio descriptors. We use spectral models analysis and synthesis techniques previously developed by members of the Music Technology Group in the IUA.
2. The extraction of high-level musical attributes (transitions, articulation and musical structure) from the low-level descriptors.
3. The use of musical knowledge for the analysis and transformation of sound. This musical knowledge is modeled using artificial intelligence techniques such as case-based reasoning and inductive logic programming. In the case of the case-based reasoning, we investigate new techniques for the tasks of adaptation and maintenance.
4. The design and implementation of a concurrent real-time programming language for modeling musical processes and to be used in audio processing applications.

On the one hand, The IUA has successful experience in sound manipulation techniques (analysis and synthesis) using spectral modeling. It also has wide experience in high-level programming environments in the area of audio processing as well as in concurrent programming languages design. On the other hand, the IIIA has successful experience in musical knowledge representation and case-based reasoning. The technologies from both groups complement each other and provide an effective approach for content-based music processing.

In order to describe the musical content of a sound source we rely on the spectral modeling and techniques developed by the music technology group of the IUA. These techniques allow the extraction of numerical values for temporal, dynamic, timbric and melodic descriptors of a

musical phrase. We have used some of these descriptors for content description and search in musical notes databases. The solid base that these techniques provide allows the development of a new level of descriptions whose utility can be maximized by the knowledge induced by manipulating a data set of examples using case-based reasoning and other artificial intelligence techniques.

2 Project Progress

IUA

At the IUA, we have pursued two research lines: on the one hand, we have investigated the extraction of different symbolic attributes from monophonic audio recordings and the intelligent synthesis of audio from symbolic descriptions. On the other hand, we have explored different machine learning techniques in the context of expressive music performance modeling. In parallel to these two main research lines, we have worked on the design and implementation of a concurrent programming language based on first order logic for the specification, execution and verification of musical processes. The different research lines have generated a number of publications (see IUA publications in next section).

Regarding the extraction of different symbolic attributes from monophonic audio recordings, we have worked on algorithms and tools for the extraction of two types of attributes: note-level attributes and intra-note attributes. The former ones characterise properties of note in a melody, e.g. pitch, duration, onset and mean energy, while the later ones characterise internal properties of a note, e.g. attack, sustain, and release. We have used both types of attributes for automatically inducing expressive performance models. We have also explored different approaches to synthesising audio from symbolic descriptions, the most promising being using samples together with a predictive model determining the most suitable note (and transition) to be played in a determinate musical context.

Regarding the application of different machine learning techniques to the task of expressive music performance modeling, our approach has been the following one: Instead of manually modelling expressive performance and testing the model on real musical data, we let a computer use machine learning techniques to automatically discover regularities and performance principles from real performance data. We have applied both inductive regression techniques, e.g. model trees, support vector machines, and rule-based classification techniques, e.g. inductive logic programming, in order to induce expressive music performance models. We have devised and implemented machine learning algorithms to fit our needs.

On the one hand, regression methods are considered to be black-box in the sense that it is very difficult (or impossible) to understand the predictions they produce. Black-box statistical

approaches may be good at deriving predictions from data, but formulating understandable rules from the analysis of data is something entirely different from formulating predictive models from that data. On the other hand, rule-based classification methods are good at explaining the predictions they provide but are restricted to a set of discrete classes as prediction space. Our problem at hand is one that requires the prediction precision of regression methods for generating accurate solutions (i.e. expressive performances) but at the same time it is highly desirable to be able to explain the system predictions. Having this in mind, we have not limited ourselves to anyone of the two techniques and instead we have explored both approaches applying a representative sample of both regression and classification methods.

IIIA

Similarity Measures Task:

Taking as starting point the results achieved in the Tabasco project (TIC 2000-1094-C02), we proposed new different similarity measures for comparing musical sequences (scores and performances) at different abstraction levels. We compared their behaviour in a database of jazz standards and determined the possible usage of each measure. The research on this has produced the following publications: [4] [5] [6] [7] [8] (see IIIA publications in next section). This task finalized successfully.

Adaptation Task:

The goal of the task is to propose new adaptation mechanisms for CBR systems that construct solution to problems that require a complex solution structure (such as design problems). We have proposed a formalism based on components (see [2]) that can be incorporated into the Case-Based Reasoning inference cycle. This formalism is based on our previous work on the “Constructive Adaptation” algorithm. We are still working on this task and the plan is to improve the current components and to perform experiments in the musical domain.

Case-Based Reasoning Maintenance Task:

The maintenance tool has as a goal to guarantee or increase the quality of the solutions a case-based system is proposing. In problems like our musical domain, where the representation of the problem requires complex structures using different abstraction levels, the maintenance is a critical and open issue. This task was fired just few months ago and started analysing the current state of the art. We have produced a first result (see [1]) where we show how an improvement on the “Competence Groups” approach can be applied for analysing in a more powerful way the case bases. We provide a new measure, based on analysing the quality of solutions, which allows a better characterization of the case base structure. Using this measure we have developed a visualization tool that allows a 3D navigation through the case base space. We are still working in this task and we are investigating how to develop a self-model of the reasoning cycle for automatically analyse the quality of solutions proposed and fire the appropriate maintenance mechanisms.

3 Results Indicators

During its first year, several researchers have joined the project:

- Emilia Gomez, Engineer in telecommunications and PhD student (obtained her Diploma de Estudios Avanzados (DEA) in 2004) in the computing and digital communication program of the UPF. She worked on research tasks such as signal processing and analysis.
- Esteban Maestre, Engineer in telecommunications, is currently registered as PhD student (first year) in the computing and digital communication program of the UPF. His research is devoted to audio analysis, description and synthesis using signal processing techniques.
- Amaury Hazan, Engineer in computing, is currently registered as PhD student (first year) in the computing and digital communication program of the UPF. His research is devoted to the design and implementation of artificial intelligence techniques..
- Marc Vinyes, Engineer in telecommunications and mathematician, is currently registered as PhD student (1st year) in the computing and digital communication program of the UPF. He has been working part-time on audio analysis and description.
- Maarten Grachten, Computer Scientist and PhD student (obtained his Diploma de Estudios Avanzados (DEA) in 2004). His Phd work is devoted on the research on developing a Case-Based Reasoning system for generating expressive tempo-transformations of jazz melody performances. His work proposal received the “ICCBR’03 Best Paper Award” at the International Conference on Case-Based Reasoning.
- Alejandro García, Computer Scientist and PhD student (first year). His research is be devoted on the use of Maintenance polices for improving the performance of Case-Based Reasoning systems.
- Jose María Abasolo. The Phd work of Jose María Abásolo (he obtained his Diploma de Estudios Avanzados (DEA) in 2004) is devoted on the research on a Component-Based framework for developing Case-Based Reasoning Applications. The research is specially focused on providing methodological assistance on solving analytical problems such as the use of Case-Based Reasoning techniques for generating expressive tempo-transformations.

Emilia Gomez is no longer working directly in this project, she is currently working in SIMAC, a EU project.

So far, the project has generated the following publications:

IUA

- [1] Rafael Ramirez, Amaury Hazan, (2005) ``A Learning Scheme for Generating Expressive Music Performances of Jazz Standards``, Proceedings International Joint Conference on Artificial Intelligence IJCAI 2005.
- [2] Rafael Ramirez, Andrew Santosa (2005), ``Formal Verification of Concurrent and Distributed Constraint-based Java Programs``, Proceedings IEEE International Conference on the Engineering of Complex Computer Systems, IEEE Press.
- [3] Rafael Ramirez, Amaury Hazan (2005), ``An Approach to Expressive Music Performance Modeling``, Proceedings of 118th Audio Engineering Society Convention.
- [4] Maestre, E. Gómez, E. (2005) 'Automatic characterization of dynamics and articulation of expressive monophonic recordings' Proceedings of 118th Audio Engineering Society Convention.
- [5] Rafael Ramirez, Amaury Hazan, (2005), `` Modeling Expressive Music Performance in Jazz``, Proceedings International Florida Artificial Intelligence Research Society Conference, AAAI Press.
- [6] Rafael Ramirez, Amaury Hazan, (2005) , `` Understanding Expressive Music Performance Using Genetic Algorithms``, Proceedings European Workshop on Evolutionary Music and Art, Springer-Verlag LNCS.
- [7] Rafael Ramirez, (2004), ``Inductive Logic Programming and Music``, Proceedings International Computer Music Conference, ICMA Press.
- [8] Rafael Ramirez, Amaury Hazan, (2004), ``Learning Expressive Performance Rules in Jazz``, Proceedings International Computer Music Conference, ICMA Press.
- [9] Rafael Ramirez, Amaury Hazan, Emilia Gomez, Esteban Maestre, (2004), ``Understanding Expressive Transformations in Saxophone Jazz Performances Using Inductive Machine Learning``, Proceedings Sound and Music Computing International Conference.
- [10] Rafael Ramirez, Amaury Hazan, (2004), ``Rule induction for expressive music performance modeling``, ECML Workshop Advances in Inductive Rule Learning.
- [11] Rafael Ramirez, Juanjo Martinez, Andrew Santosa, (2004), ``Constraint-based Concurrent and Distributed programming in Java``, Proceedings International Symposium on Distributed Computing and Applications to Business, Engineering and Science, Wuhan University Press.
- [12] Rafael Ramirez, Amaury Hazan, Emilia Gomez, Esteban Maestre, (2004) ``A Machine Learning Approach to Expressive Performance in Jazz Standards``, Proceedings International Workshop on Multimedia Data Mining (in conjunction with ACM SIGKDD International Conference on Knowledge Discovery & Data Mining).

[13] Rafael Ramirez, (2004), ``Concurrent and Distributed Programming Using Constraint Logic Programs``, Proceedings ACM Symposium on Applied Computing, ACM Press.

[14] Rafael Ramirez, (2003), ``Inductive Logic Programming for Learning Musical Rules``, Proceedings International Conference on Applied Artificial Intelligence.

IIIA

[1] M. Grachten, F.A. García, J. L. Arcos, (2005), Navigating through Case Base Competence. 6th International Conference on Case-Based Reasoning ICCBR'05. (in press)

[2] E. Plaza (2005), Cooperative Reuse for Compositional Cases in Multi-Agent Systems. 6th International Conference on Case-Based Reasoning ICCBR'05. (in press)

[3] J. L. Arcos (2004), Improving the Quality of Solutions in Domain Evolving Environments. 7th European Conference on Case-Based Reasoning ECCBR'04. Peter Funk and Pedro A. Conzález-Calero eds. Lecture Notes in Artificial Intelligence, n° 3155. Springer-Verlag, pp. 464-475.

[4] M. Grachten, J. L. Arcos, R. López de Mántaras (2004), TempoExpress, a CBR Approach to Musical Tempo Transformations. 7th European Conference on Case-Based Reasoning ECCBR'04. Peter Funk and Pedro A. Conzález-Calero eds. Lecture Notes in Artificial Intelligence, n° 3155. Springer-Verlag, pp. 601-615 .

[5] M. Grachten, J. L. Arcos (2004), Using the Implication/Realization Model for Measuring Melodic Similarity. 16th European Conference on Artificial Intelligence ECAI'04. Ramon López de Mántaras and Lorenza Saitta eds. IOS Press, pp. 1023-1024.

[6] M. Grachten, J. L. Arcos, R. López de Mántaras (2004), Melodic Similarity: Looking for a Good Abstraction Level. 5th International Conference on Music Information Retrieval ISMIR'04. pp 210-215.

[7] M. Grachten, J. L. Arcos (2004), Music Performance Generation as Time Series Prediction. ECCBR'04 Workshop on Applying CBR to Time Series Prediction. Pablo Gravás and Kaylan Moy Gupta eds. Tech.R. 142-04 Dep. Sistemas Informáticos y Programación (UCM), pp. 329-336.

[8] M. Grachten, J. L. Arcos, R. López de Mántaras (2004), Evolutionary Optimization of Music Performance Annotation. Computer Music Modeling and Retrieval CMMR'04. Lecture Notes in Computer Science. Springer-Verlag

The creation of the Start-Up MusicStrands, S.A. has motivated the collaboration with the project. We have a research contract devoted on recommendation algorithms for musical recommendation. The goal of this research was motivated from our work in the ProMusic project on similarity measures for musical sequences. We are using our research experience for developing similarity measures to be used in a collaborative filtering framework for recommending songs from user play-lists.

TIC2003-07776-C02

We have also established collaborations with academic institutions. We have visited and have visitors from both Queens University Belfast and the University of Alicante. We are currently working closely with member of these two universities on musical analysis and automatic transcription, respectively.

MEDICI: Inpatient/Outpatient Monitoring for Diagnosis and Medical Research in Ischaemic Cardiopathy TIC2003-09400-C04

Roque Marín Morales ^{*}
Universidad de Murcia

Ramón Ruiz Merino [†]
Universidad Politécnica de Cartagena

Jesus María Rodríguez Presedo [‡]
Universidad de Santiago de Compostela

Alfonso Palma Álvarez [§]
Hospital Universitario Virgen Macarena
Universidad de Sevilla

Abstract

The aim of this project (MEDICI: Inpatient/Outpatient Monitoring for Diagnosis and Medical Research in Ischaemic Cardiopathy: Tools for Acquisition, Visualization, Integration and Knowledge Discovery) is the design, development and implementation of a set of tools for inpatient and outpatient monitoring of patient suffering Ischaemic Cardiopathy. These tools will provide support for both medical care and clinical research tasks. The work is organized around three development lines: a) Subsystems for data collection in different scenarios, including clinical history data, monitored electric and hemodynamic signals and data related to remotely monitored signals outside the hospital. b) Techniques for integrating the collected information and presenting it as a unique virtual clinical history associated to the patient; and c) tools for the extraction and validation of medical knowledge.

Keywords: Temporal Data Mining, Electronic Patient Record, Semantic Web Based Clinical Information Integration, Acquisition and Processing of Biomedical Signals, Telemedicine.

^{*}Email: rmarin@dif.um.es

[†]Email: ramon.ruiz@upct.es

[‡]Email: suso@dec.usc.es

[§]Email: apalma@hvm.sas.junta-andalucia.es

1 Project Goals

1.1 Introduction

MEDICI (Inpatient/Outpatient Monitoring for Diagnosis and Medical Research in Ischaemic Cardiopathy: Tools for Acquisition, Visualization, Integration and Knowledge Discovery) is an interdisciplinary project organized in four subprojects (<http://perseo.dif.um.es/~medici/>):

- **MEDICI-TDM:** Temporal Data Mining Techniques for Medical Research (University of Murcia).
- **MEDICI-MRF:** Phonocardiogram Remote Monitoring by means of an Intelligent Signal Interpretation System (Technical University of Cartagena).
- **MEDICI-ISI:** Information System Integration (University of Sevilla).
- **MEDICI-SICAU:** Cardiology Information System with Automatic Extraction and Acquisition of Medical Knowledge and Ubiquitous Access (University of Santiago de Compostela).

There are different hospitals involved in the project: Hospital Clínico Universitario of Elche, Hospital General Universitario Reina Sofía of Murcia y Hospital Virgen de la Macarena of Sevilla. Moreover, due to the project, collaborations are being carried at different levels with other hospitals, which will be cited in section 3.

Table 1 summarizes the human resources allocated in MEDICI project. Three predoctoral student has to be added to the total final count which corresponds to the three grant associated to the University of Murcia Subproject.

	UMU	UPCT	US	USC	TOTAL
EDP	6.5	8.5	3	6	24
Doctorate Researchers	5	8	1	7	21
Total Researchers	9	12	4	11	25

Table 1: Human resources involved in MEDICI

1.2 Project Goals

The **aim** of this project is the design, development and implementation of a set of tools for inpatient and outpatient monitoring of patient suffering from Ischaemic Cardiopathy. These tools will provide support for both medical care and clinical research tasks.

The developed tools cover two **application domains**, both focused on Ischaemic Cardiopathy:

- Patients who have been admitted at an Intensive Care Unit (ICU) suffering an Acute Myocardial Infarction (AMI), with or without associated complications. Two local scenarios are considered: a) Intensive Care Unit (ICU) and b) home remote monitoring after his or her discharge from hospital
- Renal patients with risk of Ischaemic Cardiopathy and subjected to dialysis. The working scenarios are: a) Peripheral Dialysis Center b) External Consultations and c) Hospital Admission

The work is organized around three development lines: a) Subsystems for data collection in different scenarios, including clinical history data, monitored electric and hemodynamic signals and data related to remotely monitored signals; b) Tools for the extraction and validation of medical knowledge; and c)

Techniques for integrating the collected information and for presenting it as an unique virtual clinical history associated to the patient.

The **first line** provide our hospitals with a technological infrastructure that may improve asistential tasks and, at the same time, may enable the application of Data Mining techniques to provide support to medical research. This infrastructure is also expanded to the outpatient sceneries in order to follow-up patients for a longer time.

The **second line** provides models, algorithms and tools for Temporal Data Mining. The main idea is to discover patterns described by temporal association of clinical history data and/or signal event and episodes.

The **third line** is composed of a set of models and tools to integrate the whole infrastructure around a general information system based on a multiagent architecture, which will provide access to data and visualization capabilities to the different data bases in the project. This philosophy is in the line of other emerging proposals in the field of VHCR (Virtual Healthcare Record). Furthermore, and in order to facilitate the technology transfer process of the project outcomes, an especial emphasis will be paid to interoperability issues related to the integration of projects results into conventional hospital information systems.

The two first lines are being applied to the first of the two applications domains (Acute Myocardial Infarction in ICU). The second domain (Ischaemic Cardiopathy in Nephrology) is used as a testbed for the VHCR developments and database interoperability through web services. These techniques will be applied to the first domain in the final stages of the project.

The **specific goals** of this project are:

1. Design, development and implementation of tools oriented to the follow-up of patients with Ischaemic Cardiopathy (inside and outside the hospital), that have been admitted in an ICU after suffering an Acute Myocardial Infarction.
2. Integration of the developed tools around a virtual information system, by means of ontology servers, VHCR techniques, standards for representation and sharing of clinical data, and semantic web techniques.
3. To provide tools for ubiquitous monitoring of biomedical signals, based on PDA-like devices, and in the bidirectional transmission of information through mobile telephony resources (GPRS), so that the patient follow-up from his or her stay in the ICU and the weeks following his or her discharge from hospital.
4. Phonocardiographical signal processing focused on the detection and tracking of heart murmurs patients suffering in heart valve related pathologies. This signal is included in the set of signals to be remotely monitored at home.
5. Research and development of temporal data mining techniques to intensively analyze the collected data, with the goal of detecting and discovering temporal patterns with possible clinical interest.
6. Clinical validation and evaluation of the developed modules.

1.3 Project Schedule

Until the elaboration of this document, the execution of the project fulfils the working plan (Figure 1). The incidences have had minimal importance and have not affected the achievement of the objectives. At the moment of writing this report, the project is located in the middle point of its development (18 months).

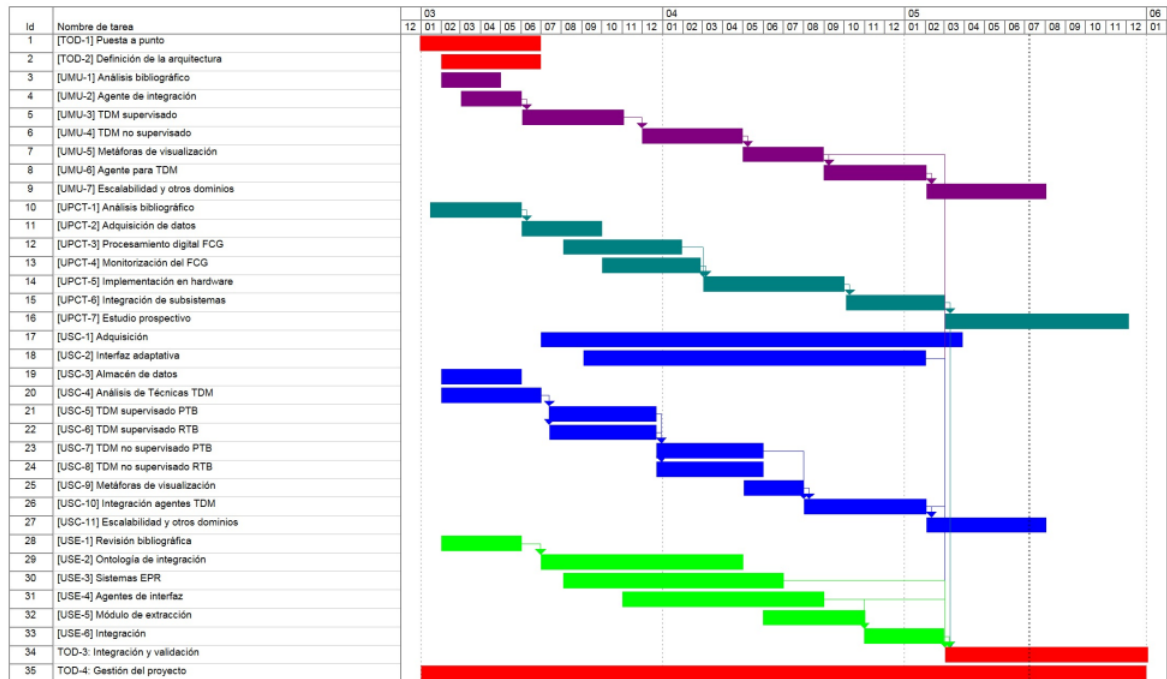


Figure 1: Project Tasks and Scheduling.

2 Level of success reached in the project

Next, we will describe the technical and scientific results achieved during the project, structured in the three development lines described in section 1.2.

2.1 Information Collection Tools Level

2.1.1 Inpatient infrastructure

- **CH4:** Information System for the management of Patient Clinical Histories that allows: the registry of all possible incidents during the stay of the patient at ICU, the automatic generation of all reports and documents legally required by the hospital administration, as well as management procedures and remote query (taking into account security measures). The validation of CH4 has been done at the ICU of the Hospital General Universitario of Elche. Furthermore, there is an agreement for the implantation of CH4 at the ICU of the Hospital Universitario of Getafe.
- **Nursing Sheets Management Tool:** Developed for nursery staff, this tool allows the collection of physiological and hemodynamical parameters, clinical analysis information, treatment, and feeding. Other facilities, such as remote access via PDA, have also been developed.

2.1.2 Outpatient Infrastructure

- **Home Outpatient Monitoring Prototype based on PDA and mobile phone:** [17]. We have developed the data acquisition drivers and the algorithms required for real time monitoring

of different physiological signals, basically ECG and oxygen saturation. It is worthy to mention here the detection and morphological classification of heart beats, the detection of the main pathological arrhythmia, and the extraction of several parameters from the ECG signal, such as cardiac frequency, width and height of the different waves that conforms the cardiac cycle. Furthermore, all this information has been integrated together with the data from the oxygen saturation sensor. An encoding algorithm has been implemented as well as the data-transmission algorithm through mobile phone connected to a PDA via Bluetooth. The validation of the system has been done at The ICU)of the Hospital General Universitario of Elche.

2.1.3 Tools to objectivate auscultation findings: e-PCG

- Tools for the acquisition and management of phonocardiographic and echocardiographic information:** A graphical environment (running on PC and PDA) has been implemented to acquire all the information related with the auscultation process, in order to objectivate and assist decisions involving the phonocardiographic signal (PCG). Then, it is necessary not only to acquire PCG and auxiliary (peripheral pulse) signals, but also to register additional information (physiological data of the patient, auscultation area, maneuvers, etc.), which have great influence in the analysis of PCG. These tools include a multimedia database for storing, querying and visualization of annotated PCG signals, which uses visual metaphors to represent the cardiac cycle. Methods for guaranteeing security and confidentiality of patients have been implemented. Also, a web service enables the remote access to the patient data and the associated signals[7]. Besides, a tool for the analysis of echocardiographic images has been implemented on a high-performance DSP (*Digital Signal Processors*) platform[25, 26, 27]. This tool uses algorithms to remove the speckle noise and to track the endocardium through active contour model.
- ASEPTIC: Processing algorithms for the PCG signal:** The analysis of PCG signal has been implemented using a hierarchical structure based on four abstraction levels (Envelopes, Events, Sounds, and Diagnosis), and several tasks specialists in each level. The modularity of this structure allows a high degree of independence in the implementation of the processing blocks, and it establishes a growing up abstraction level as we climb through the hierarchy. This approach tries to mimic the interpretation process of the cardiac auscultation carried out by the physician, which uses also a hierarchical approach to search and characterize the cardiac events [14]. The advantage of this method is that it allows to refine the search of events, even inside other events, following a knowledge-based progressive depth approach.
- Continuous monitoring techniques for PCG:** Due to slow evolution characteristics of this signal, it is enough to perform a periodical analysis of PCG segments rather than a continuous monitoring, which makes easier the implementation of specific purpose modules on reconfigurable hardware circuits. An efficient method for the compression of PCG based on the discrete wavelet transform (DWT) [12, 13] has also been developed. This method improves the storage and transmission of PCG, and it has achieved higher compression rates than the main audio compression methods used nowadays, for a given quality level.
- Prototype of an FPGA-based intelligent stethoscope:** The general architecture of the prototype of intelligent stethoscope has been defined, and many of the time-consuming tasks has been implemented on high performance FPGAs (*Field Programmable Gate Array*). The processing blocks implemented on FPGA are the following ones: digital filters, using a technique that improves commercial tools, and whose size and operating frequency are independent of the number of coefficients of the filter [4, 10]; the computation of the instantaneous frequency [3]; the FFT (*Fast Fourier Transform*); the envelope computation (using digital filters); and the DWT (*Discrete Wavelet Transform*).

Systems related to the processing of PCG are being used at the cardiology service of the Hospital Universitario Reina Sofia of Murcia.

2.2 Knowledge Discovery Level

- **Temporal Abstraction Agent:** It can be considered the first step for the information treatment in order to feed the TDM process. Its goal is to reduce the event set granularity and to allow a higher level of the abstraction. Moreover, this agent provides a summary of the patient evolution (CH4 tool incorporates this feature) and allows to discover hypotheses at different levels of abstraction. It is based on a Fuzzy Temporal Constraint Network model (FTCN), and it is domain independent. This agent takes into account different contexts, that could change the result of the abstraction, and enables the composition of mechanism of temporal abstraction.
- **Algorithms and tools for Sequences and Temporal Episodes mining:**
 - A theoretical model of Fuzzy Temporal Diagnostic Patterns (FTDP)[5, 6, 8, 16] has been adapted as representational framework for supervised temporal episode mining. The mined hypotheses, that could not be completely specified, are introduced by the user using *CATEKAT – 2*[15], an adapted version of a previous tool for causal and temporal medical knowledge acquisition.
 - We have also developed and implemented two algorithms for temporal sequences mining: *TSET* and *TSET^{MAX}*. The first one is an Apriori-like algorithm that uses a single data structure (set-enumeration tree) to store the detected sequences. The second algorithm, *TSET^{MAX}*, improves the computational efficiency of *TSET*, by means of using a lookahead technique. This algorithm implements a model based on Dempster-Shafer Theory to generalize frequent sequences into temporal episodes, in order to describe, in a compact way, all the possible temporal orders in which frequent events may appear. This latter can be applied as a post-processing algorithm for the sequences discovered by *TSET* and *TSET^{MAX}*. The evaluation proves that both algorithms are scalable -concerning to the data-base size.
 - Finally, a model for temporal episode mining has been proposed, using a FTCN as representational framework. This model allows the use of metric temporal constraints, and not only qualitative ones. To this end, an Apriori-like method -based on brute force- has been designed. This method mines the temporal relations instead the events. This algorithm has been partially implemented and some theoretical problems, concerning the consistency of detected episode constraints, will be solved in the next step of the project.
- **Algorithms and tools for mining Multivariable Fuzzy Temporal Profiles (MFTP):**
 - **TRACE:** [19] This is a new version of a previous tool for the acquisition and recognition of signal profiles. It is based on Multivariable Fuzzy Temporal Profiles (MFTP)[18, 20] and is oriented to the representation of hypotheses using signal finding patterns. The TRACE tool also allows the definition of signal morphology patterns based on generic templates and refine them using the occurrence detected by the recognition algorithms. Furthermore, it is possible to select signal fragments that can be relevant for the physician, disposing a graphical representation of a pattern that generalizes the morphology of the given fragment. The TRACE Tool has proved to be useful at the ICU of the Hospital General Universitario of Elche for discovering several patterns related to carotid sinus sensitivity from the electrocardiographic records. A more detailed clinical evaluation of this system, that eventually could permit a novel research line on migraine, is ongoing. Visual metaphors has also been integrated using profiles highlighting and a predefined

color coding scheme. This system facilitates the understanding and the labelling task of the recorded and processed signals, as well as the result of the mining techniques used in this project.

- **Algorithms for Mining Fuzzy Temporal Rules (FTR):** The starting point of this line has been the application to biomedical signal of a FTR model already validated in control and landmark identification in mobile robotic [1]. A FTR defines conditions on signal values, their occurrence times (instants and intervals), and the relations between different occurrence times. Currently, we are testing supervised and non supervised learning algorithms. The algorithm output is a FTR set, which may point out relevant clinical situations that could be missed otherwise. In the biomedical signal field, the goal of the FTR is also to provide some diagnosis (totally or partially), to verify a previously established diagnosis, or too detect possible failures or complications.

2.3 Clinical history integration and presentation level

- **Integration Ontologies..** Despite several stable versions of different ontologies for the integration of the information within the clinical history are available, it must be taken into account that the clinical information system knowledge model (ontology) cannot be considered completed because the concepts can be changed and extended. This model is based on the separation of concepts related to the system engineer (reference model) from those concepts closer to clinical staff (archetypes). The advantage of this paradigm is to permit that clinical staff could define new concepts without knowing in depth the information system, and to have a flexible and interchangeable representation of concepts, avoiding semantic conflicts. To this end, the following ontologies have been developed:
 - **Reference Model Ontology:** A federated ontology based on the openEHR reference model. These reference models have been constrained to those aspects related to the representation of archetypes. Besides, mechanisms for extending the reference model and archetype specification have been defined. The work developed in this line can be extrapolated to any health system architecture. OWL, which is a normalized language, has been used as specification language.
 - **Demographic Model Ontology.** As practical case of higher level concept specification (archetypes), we have paid special attention to demographic information, as the first agents implemented have been those related to the identification and retrieval of patient's personal data. To this end, we have implemented a federated ontology based on the GPICs of CEN and the reference model of openEHR. Furthermore, other local ontologies about demographic concepts have been developed in order to perform integration tests of different domains with different semantics.
- **Agent Based Integration Platform based** Since designing a system as open as possible is one of the main goals of this line, we have considered the use of the CEN guidelines, in particular its HISA (Healthcare Information System Architecture) norm which describe the use of the ODP for systems development. Taken into account the main goals of the system, the design of the basic architecture of services is based on the service decomposition of the health environment as described in: CORBAmed, HISA[4] and PICNIC. In parallel, the typical clinical sessions of each scenario of MEDICI project has been analyzed; this analysis will allow the establishment of protocolized workflows that show the interaction between the agent implied in each session [24]. In current version of the architecture the following agents can be found: clinical observation access agent [9], laboratory access agent [21], security Agent [22], identification and integration of demographic data agent [24], and integration mapping and ontology mapping agents[23].

- **Clinical Case Visualization Tool** This tool allows to visualize temporal and atemporal data, using time line diagrams (Gantt Diagrams), graphs for causal relation description and conventional visualization techniques. Thanks to the use of techniques of weight and selection of the importance of different data, together with the selection of keywords in free text data, this tool allows to select the very relevant parts of the clinical case that could be lately used to apply mining algorithms. The current version of this tool is adapted to the particular case of an Intensive Care Unit.
- **Prototype of data warehouse for monitoring** In the project there is a prototype of datawarehouse for store all information related to generated signal monitoring, in such a way that it can be queried from both user interfaces and mining tool applications. Associated to the data warehouse, an adaptive interface has been implemented to query patient monitoring information, as well as other data generated by different data mining tools. This interface has a secure communication with the datawarehouse, and allows the access to the information depending on the user profile. Moreover, a user management tool has been implemented for registering new users and devices in the system.

3 Achievement indicators

3.1 Publications

Table 2 shows the number of publications related to the project. Because of the space limitation, only the most significant ones are cited in the References section. A full relation can be found at the project website (<http://perseo.dif.um.es/~medici/>).

International Journals	5
International Book Chapters	3
National Journals	1
National Book Chapters	2
International Conferences	34
National Conferences	9
Published Manuals	1
Degree Projects	23

Table 2: Project publications

3.2 Training accomplishments

Currently, there are 3 people under research training, all of them associated to the subproject in Murcia: a contract, a FPI grant, and a FPU grant. One more person, coming from the Udine University (Italy), will enroll next October, as he has obtained a Juan de la Cierva spanish MEC grant.

3.3 PhD Theses in progress

Table 3 summarizes ongoing PhD theses associated to the project. A more detailed description can be found in the project web page (<http://perseo.dif.um.es/~medici/>).

Universidad de Murcia	3
Universidad Politécnica de Cartagena	3
Universidad de Sevilla	1
Universidad de Santiago de Compostela	3

Table 3: Ongoing Doctoral Theses Associated to the Project

3.4 Collaborations with International and National Groups

Several groups, hospitals and services have been contacted during the development of the project. The goals and the current state of these collaborations follow briefly:

- Servicio de Anestesia, Reanimación y Unidad del Dolor del Hospital Universitario Morales Meseguer of Murcia. As a first goal, we have proposed the development of a tool for detecting patient profiles in anaesthesia. Possibly, a few residents in the service, currently developing doctoral theses, will be enrolled in this collaboration. We will also ask for contracting a programmer, and we are now looking for research funding.
- Servicio de Aparato Digestivo at Hospital Universitario Morales Meseguer of Murcia. The contacts with this service, besides technology transfer, are focused on the elaboration of joint projects oriented to the development of applications for clinical research and decision support systems.
- Departamento de Neumología at Complejo Hospitalario Universitario of Santiago de Compostela (Dr. Carlos Zamarrón Sanz). As a result of this collaboration, the group of the University of Santiago de Compostela has applied for and obtained fund support from the Xunta de Galicia for the development of a research project related to the topics of this project, but applied to data from patients who suffer from Síndrome de Apnea del Sueño (SAS). This project has been entitled “Minería de datos en polisomnografías de pacientes con alteraciones cardiopulmonares del sueño” (PGIDIT04SIN206003PR).
- Grupo de Investigación Bioimagen y Modelos de Visión del Instituto de Óptica “Daza de Valdés”, which belongs to the CSIC (Spanish National Research Council). With this group we have recently started a collaboration for the application of disorder measures developed in that group to time-frequency representations of the phonocardiographical signal. This collaboration will be fixed in short time by means of an agreement between the Technical University of Cartagena and the CSIC Institute.
- Department of Medical Informatics of the Amsterdam University (Dr. Ameen Abu-Hanna). D. Manuel Campos Martínez, member of the University of Murcia group, spent four months visiting that department, working on Temporal Data Mining. Dr. Abu-Hanna currently coordinates a research project which has some goals in common with MEDICI. Dr. Abu-Hanna has also visited University of Murcia, taking part in a doctoral programme with ANECA Quality Mention.
- Departamento de Informatica de la Universidad del Piamonte Oriental Amadeo Avogadro, Italia (Dr. Luigi Portinale). From august to December 2005, D. José Manuel Juárez Herrero (University of Murcia), will carry out a stay in that department to work in the field of the temporal Case-Based Reasoning applied to medicine. Dr. Portinale will be invited to take part in the doctoral programme given next year by the University of Murcia group.
- In the field of models for the clinical history, the University of Sevilla group has collaborated with the openEHR workgroup. They have shown the interest in the development of archetypes using the OWL language. In this moment a CR (Change Request) is being elaborated in order to modify the demographic reference model of openEHR and make it closer to the CEN standards.

As a profit from this relation, in May 2005 the Sevilla group invited the responsible of the group, Dr. Thomas Beale, with whom possible collaboration lines were defined.

- Fraunhofer Institute for Integrated Circuits IIS. It is one of the most well-known institutions in Europe for the microelectronic design. The Technical University of Cartagena group has started contacts with it, and there is an agreement for collaborating in the development and implementation of efficient low-consumption architectures for analog-digital converters, for the integration of the signal acquisition in a future mixed-signal ASIC.

3.5 Patents and Technology Transfers

- CH4 has been registered in the Registro Territorial de la Propiedad Intelectual (register number: 08/2005/347, with date 2005.05.24 and title CH4-UCI)
- Successful contacts: We have reached an agreement to computerize the ICU Service of the Hospital Universitario of Getafe, adapting the CH4 environment to the service. Currently an agreement has been elaborated, pending from approval at the University of Murcia.
- Contacts in progress: We have contacted with ASISA and Virgen de la Vega private hospital in Murcia through the OTRI of the University of Murcia, and we are trying to reach an agreement to develop a project centered around the tools to objectivate auscultation findings (e-PCG).

3.6 Other diffusion activities

- Some informative brochures related to the technological offer of the group have been edited (some of them are available at <http://perseo.dif.um.es/~aike/htm/resultado.htm>), which were sent to the OTRI (Oficina de Transferencia de Resultados de Investigación) of the University of Murcia). These brochures allowed us to contact ASISA (Asistencia Sanitaria Interprovincial de Seguros, S.A.).
- Organization of an invited session at IWINAC-2005 ("Advances on Intelligent Methods and Applications for Health-related Processes"). Project results were presented and discussed with others researchers working on similar research lines (16th Jun 2005), as for example Dr. Abu-Hanna.
- Courses related to the project are given in two Phd Studies programs with Quality Mention of ANECA (Murcia and Santiago)
- High Mountain expedition UMU8000 to Broad Peak (K3) in Pakistan. Sponsored by the University of Murcia, one of the scientific goals are related to telemedicine experience oriented to telemonitoring PCG signals in high altitude conditions. Data transmission is done via satellite telecommunication technology. The records are evaluated at the Servicio de Cardiología of the Hospital Universitario Reina Sofia of Murcia. This study aims to study high altitude heart adaptation. One member of the project team (Dr. Félix C. Gómez de León Híjes) is in the climbing team. He keeps a blog (<http://www.um.es/broadpeak/>) from the base camp, at 5,900 m.

References

- [1] Purificación Cariñena Amigo, Carlos Vázquez Regueiro, Abraham Otero Quintana, Alberto J. Bugarín Diz, and Senén Barro Ameneiro. Landmark detection in mobile robotics using fuzzy temporal rules. *IEEE Transactions on Fuzzy Systems*, 12(4) (Special Issue on Advances of Fuzzy Systems in Robotics):423–435, 2004.

- [2] M. A. Cárdenas-Viedma. A sound and complete Fuzzy Temporal Constraint Logic. *IEEE Trans. Syst., Man, Cybern.* (to be published) 2005.
- [3] Andrés Hernández-Esteban, F. Javier Toledo-Moreo, Juan Martínez-Alajarín, J. Javier Martínez-Álvarez, and Ramón Ruiz-Merino. FPGA-based implementation of the instantaneous frequency estimation of phonocardiographic signals. In *Proceedings of IFAC Workshop on Programmable Devices and Systems (PDS 2004)*, pages 423–428, Cracovia, Polonia, 18–19 de Noviembre 2004.
- [4] J. J. Martínez-Álvarez, J. Garrigós-Guerrero, J. Toledo-Moreo, and J. M. Ferrández-Vicente. *FP-GAs: Computación & Aplicaciones.*, chapter Modelado, síntesis digital y descripción hardware de algoritmos y sistemas lineales con Matlab, pages 193–200. Edicions Universitat Autònoma de Barcelona, 2004.
- [5] J. Palma, J.M. Juárez, M. Campos, and R. Marin. Fuzzy approach to temporal model-based diagnosis for intensive care units. In Iospress, editor, *Proceedings of the XVI European Congress of Artificial Intelligence 2004 (ECAI)*, pages 868–872, 2004.
- [6] J. Palma, J.M. Juárez, M. Campos, and R. Marin. A fuzzy model for temporal model-based diagnosis. *Artificial Intelligence in Medicine (to be published)*, 2005.
- [7] J. Palma, M. Breis, K. Herzog, A. Fritz, and J.L. Lopez. Multimedia infrastructure for intelligent diagnosis and processing of phonocardiographic signals. In R. Moreno, A. Quesada, and J. C. Rodríguez, editors, *Proceedings of 10th International Conference on Computer Aided Systems Theory, EUROCAST 2005*, pages 136–138, 2005.
- [8] J. M. Juárez, J. Palma, M. Campos, A. Morales, J. Salort, and R. Marin. A model-based architecture for fuzzy temporal diagnosis. In *LNCS (to be published)*. Springer-Verlag, 2005.
- [9] J. Álamo, I. Román, L.M. Roa, and G. Madinabeitia. Desarrollo de un servidor coas con tecnología de servicios web. In *Actas del CASEIB 2004. XXII Congreso anual de la Sociedad Española de Ingeniería Biomédica*, pages 301–304, 2004.
- [10] J. J. Martínez, A. Legaz, F. J. Toledo, F. J. Garrigós, and J. M. Ferrández. Sistema basado en PCI para la co-simulación de aplicaciones DSP y de imagen sobre FPGAs desde matlab. In *V Jornadas de Computación Reconfigurable y Aplicaciones (JCRA-05)*, Granada (España), Septiembre 2005 2005. (Aceptado para publicación).
- [11] Juan Martínez-Alajarín and Ramón Ruiz-Merino. Compresión de fonocardiogramas mediante las transformadas wavelet y wavelet packet. In *Actas del XXII Congreso de la Sociedad Española de Ingeniería Biomédica (CASEIB 2004)*, pages 181–184, Santiago de Compostela, (España), 10–13 de Noviembre 2004.
- [12] Juan Martínez-Alajarín and Ramón Ruiz-Merino. Estructura jerárquica de un sistema de diagnóstico basado en la señal fonocardiográfica. In *Actas del XXII Congreso de la Sociedad Española de Ingeniería Biomédica (CASEIB 2004)*, pages 177–180, Santiago de Compostela, (España), 10–13 de Noviembre 2004.
- [13] Juan Martínez-Alajarín and Ramón Ruiz-Merino. Wavelet and wavelet packet compression of phonocardiograms. *Electronics Letters*, 40(17):1040–1041, 2004.
- [14] Juan Martínez-Alajarín and Ramón Ruiz-Merino. Efficient method for events detection in phonocardiographic signals. In *Proceedings of SPIE*, volume 5839, 2005. (Aceptado para publicación).
- [15] J. Palma, M. Campos, J.M. Juárez, and A. Morales. Acquisition of causal and temporal knowledge in medical domains. a web-based approach. In *Engineering Knowledge in the Age of the Semantic Web (EKAW2004)*, volume 3257 of *LNAI*, page 513 – 514. Springer-Verlag, 2004.
- [16] J. Palma, R. Marín, J. L. Sánchez, and M. A. Cárdenas. Multiagent blackboard based architecture for real-time knowledge-based systems. In J. Cuenca, Y. Damazeau, A. García-Serrano, and J. Treur, editors, *Knowledge Engineering and Agent Technology*, pages 67–88. Iospress, 2004.

- [17] Jesús María Rodríguez Presedo, Daniel Castro Pereiro, Manuel Fernández Delgado, and Senén Barro Ameneiro. Telemonitorización electrocardiográfica de pacientes mediante el uso de dispositivos de computación móvil. *Telemedicina, Aplicaciones y Nuevas Tecnologías*, pages 143–155, 2004.
- [18] Abraham Otero Quintana, Paulo Félix Lamas, and Senén Barro Ameneiro. MFTP: a model to represent hierarchies of abstraction defined over multiple parameters. In *Proceedings 16th European Conference on Artificial Intelligence (ECAI)*, pages 95–100, España, 2004.
- [19] Abraham Otero Quintana, Paulo Félix Lamas, Senén Barro Ameneiro, and Francisco Palacios Ortega. A tool for the analysis and synthesis of alarms in patient monitoring. In *Proceedings Seventh International Conference on Information Fusion*, volume 1, pages 951–958, Suecia, 2004.
- [20] Abraham Otero Quintana, Carlos Vázquez Regueiro, Miguel Ángel Rodríguez González, José Luis Correa Pombo, and Paulo Félix Lamas. A fuzzy constraint satisfaction approach for landmark recognition in mobile robotics. In *Proceedings International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU 2004)*, volume 1, pages 183–190, Italia, 2004.
- [21] I. Román, R. García, L.M.Roa, and G. Madinabeitia. Open laboratory information system based on cen data types and loinc codes. In *Proceedings of the EMBEC'05 3rd European Medical & Biological Engineering Conference IFMBE European Conference on Biomedical Engineering*, 2005.
- [22] I. Román, C. Paz, G. Madinabeitia, and L.M. Roa. Consideraciones sobre el control de acceso a recursos en una arquitectura de historia clínica. In *Actas del CASEIB 2004. XXII Congreso anual de la Sociedad Española de Ingeniería Biomédica*, pages 305–308, 2004.
- [23] I. Román, L.M. Roa, and G. Madinabeitia. Ontology server for the integration of demographic servers. In *EMBEC'05 3rd European Medical & Biological Engineering Conference IFMBE European Conference on Biomedical Engineering*, 2005.
- [24] I. Román, L.M. Roa, L.J. Reina, and G. Madinabeitia. *Standardization of demographic service for a federated healthcare environment*. Studies in health technology and informatics. IOS Press, 2005.
- [25] J. Zapata and R. Ruiz. A DSP-based active-contour model. In *1st. Int. Conf. Informatics in Control, Automation and Robotics*, pages 425–429, Setúbal (Portugal), Agosto 2004 2004.
- [26] J. Zapata and R. Ruiz. Reduction of the speckle noise in echocardiographic images by a cubic spline filter. In *4th IASTED Int. Conf. Visualization, Imaging and Image Processing*, pages 772–777, Marbella (España), Septiembre 2004 2004.
- [27] J. F. Zapata and R. Ruiz. Implementation on DSP of an active contour algorithm for endocardium tracking in echocardiographic images. In *Proceedings of the SPIE*, volume 5839, 2005. (Aceptado para publicación).

The Application of Artificial Intelligence Techniques to Automatic Angiography Analysis using PACS to establish a Score for the effect of Coronary Stenosis TIC2003-07593

Alejandro Pazos [*]	Javier Pereira, Julián Dorado, Francisco
Head Researcher	Nóvoa, Alfonso Castro, Sabela
Department of Information Technologies	Fernández, Bernardino Arcay
and Communications. IMEDIR	Researchers
University of A Coruña	University of A Coruña

Abstract

The aim of this project is to develop a system making it possible to detect and establish a quantification pattern for the effect of coronary stenosis. In doing so it was first necessary to set up a PACS (Picture Archiving and Communication System) developed ad hoc by the research team for the Hemodynamics Service of the Juan Canalejo University Hospital Complex in A Coruña, Spain, in order to be able to acquire and treat the angiography studies. In this project, a system is being developed to help with the decision made by arterial models for patients, using image segmenting techniques and arterial labelling. Afterwards, algorithms are applied for detecting stenosis in the arterial tree. The precise quantification of this ratio makes it possible to make a much more precise diagnosis, avoiding having to carry out additional tests on the patient, with the subsequent saving of hospital resources.

Keywords: Angiography, Digital Image, Stenosis, Digital Image Processing, Digital Medical Images, PACS

1 Aims of the Project

General Aim

Cardiac illnesses are pathologies of major social importance that represent an ever-increasing segment of health costs. Angiographies are commonly used in the diagnosis and prognosis of these illnesses, carried out by hemodynamics units.

Most hospitals use angiographs with a single arm that generate 2D images from different perspectives. The Cardiologist then uses this to calculate a risk score depending on the 'post-stenosis' cardiac zone using a method that considers arteries that are proximal and distal to the

^{*} Email: apazos@udc.es

stenosis. This score is an estimate, which depends on the anatomy observed in the arterial tree and the position of the stenosis in the tree.

This project takes as its starting point that it is possible to develop an automatic system to support the decision-making process in the field of medical diagnoses for ischemic cardiopathies.

Specific Aims

1. Implantation of the PACS developed by the research team (as mentioned in the Introduction) in the hospital's communications network in order to acquire and store DICOM hemodynamics studies.
2. Segmentation of the coronary arterial tree using different digital image processing techniques. Using the infrastructure referred to in paragraph 1 above, a database of patients will be created with the angiographic studies and segmented images.
3. Using this information, the aim is to create an arterial model for each study making it possible to adjust the diagnosis to their particular anatomical features. Analysis will be made using this model in order to locate stenoses.
4. A knowledge-based system will then be developed capable of calculating the risk of coronary accident. This is carried out via the assisted quantification how severely the cardiac muscle is affected in the "post-stenosis" zone. Here the aim is to automatically emulate the system for quantifying stenosis currently used in clinical practice. This system may be used to support the doctor's decision in order to make a more objective and independent evaluation than that currently being made, for which Artificial Intelligence techniques will be used.
5. Developing new risk score patterns based on the information taken from the database of images developed. The systems currently in use for making risk evaluations are based on the visual analysis of images and evaluation of the parts of the arteries with and without stenosis. In this project, new quantification systems will be sought based on the use of image processing techniques and the status of the arterial tree (position of the stenosis, degree of occlusion, etc.).
6. Finally, the development of this system will facilitate access to clinical records, including angiographic images, by other hospitals, thereby assisting in inter-hospital consultations and the process of obtaining second opinions. Making data available over the Internet opens the way to implanting telemedicine consultations while only making minimum changes to the existing system.

2 Level of success achieved by the project

In line with the aims defined above, the project has developed in two main directions:

2.1 Information System (PACS)

Development of the system for acquiring, storing and accessing the hemodynamics studies produced at the Juan Canalejo Hospital. This platform serves as the foundation over which specific tools are being developed to achieve the second goal of the project. This is now almost complete and operating at test stage in the hemodynamics unit at the hospital. One system is in use that is already storing all of the angiographic studies produced by the two rooms in the unit. The other room creates analogical studies, and work is underway to define a protocol that easily digitalizes and standardizes these studies to DICOM for their subsequent integration.

This system gives doctors access via an interface to all of the hemodynamics studies produced at the hospital for their study and diagnosis. This makes it possible to eliminate the present conventional storage system, which stores each patient's data on a CD-ROM, leading to problems both in the time taken to access studies as well as with storage space. This manual system is currently being replaced by access to a PACS, thereby overcoming the deficiencies that exist at present. The tool that has been designed makes it possible to access the PACS and to create new sequences of frames, and to create diagnostic reports associated with the patient's study. It also makes it possible to modify the frames' adjustments, facilitating the diagnosis of the cardiopathy made by the specialist [1][2]. Figure 1 shows an example of the interface currently in use:

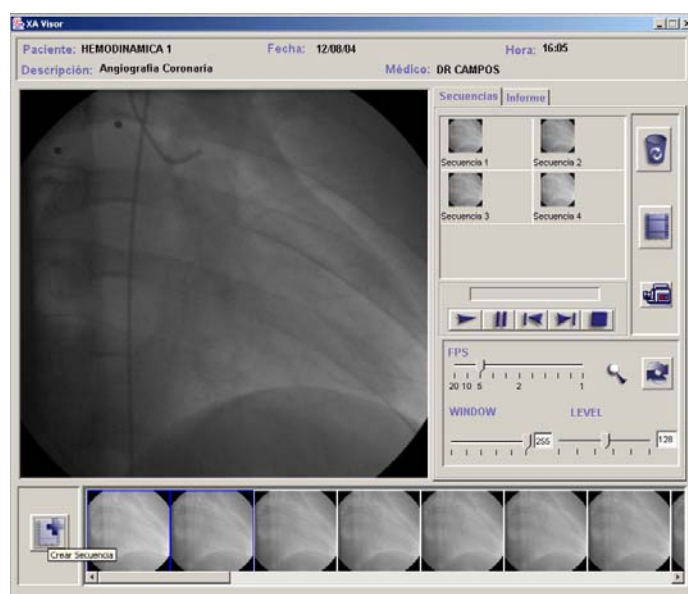


Figure 1. Interface used for access to hemodynamics studies

Study to assess the dimensions of the storage system for studies carried out at the Juan Canalejo Hospital Complex

Medical personnel were questioned about the number of studies made on a yearly basis in their unit, how long they lasted, the estimated size of these studies, and the number of series each contained. This information may be summarized as follows:

- The load on the exploration rooms is approximately equal, with around 50% of the studies being carried out in each.
- The number of studies carried out yearly, including diagnoses and angioplasties, is approximately 3700 (2600-2700 diagnoses and 1000 angioplasties). This number is subject to a degree of random variation, depending on breakdowns in the system, tending to rise each year.
- The study lasts 25 seconds for a diagnosis and another 25 seconds if an angioplasty is also carried out.
- The size of the information produced in each study is unknown by the specialists.
- The number of series in each study varies from case to case, as more than one is taken, varying parameters until the necessary information is obtained to make a correct diagnosis.

- Each series lasts between 4 to 5 seconds for diagnostic cases, and around 6 for those from orthographies.
- The studies have been stored on CD-Rom since the end of 2000.

30 CD-ROMs were then randomly selected in order to estimate the average size and duration of the studies. The results obtained are given below_

- Average space occupied per study (including DICOM file viewer): 169.46 Mb
- Average duration of each sequence recorded at 12.5 images per second: 3.888 s.
- Average duration of each sequence recorded at 25 images per second: 4.098 s.
- Average total duration of a patient's study: 72.093 s.

Taking this information into account, it may be deduced that the idea of the specialists regarding the duration of each sequence is very approximate. Also, the total duration of the studies is clearly underestimated. As for the total amount of information created, it may be concluded that some 488 Gigabytes are produced per year [3].

2.2. Processing tools

Involving the development of tools for processing angiographic images to degree how seriously the cardiac muscle is affected by stenosis. Vascular segmentation and the evaluation of lesions is a field in which continuous research is carried out, and in which there is room for major improvements such as combining techniques, working with various images, labelling, etc., providing more complete information and helping to improve the segmentation process by automating the detection of stenosis or locating it in the coronary tree. For this reason, this project is currently focusing on the implementation of different digital image processing techniques (classic algorithms, artificial neurone networks, deformable models, etc.), in order to determine which solution or combination of solutions is the most effective.

Different techniques and vascular segmentation algorithms have been tested, which may be categorized as follows:

- "pattern-matching" techniques
- model-based methods
- tracking-based methods
- methods based on AI
- methods based on artificial neurone networks
- methods for detecting miscellaneous objects with tubular structures

The algorithm that has provided the best results and which has been most widely used to date is within the techniques of pattern matching (based on the skeleton and the growth of different regions), as well as techniques based on tracking [4].

The trials carried out to date, using a technique designed for the segmentation of coronary angiographies, allow experts to initiate and guide the semiautomatic detection of the borders of vessels. Once satisfied with the segmentation, the location of the stenosis is marked on the image, applying an algorithm that will evaluate its impact depending on its location, helping to calculate the risk of cardiovascular problems associated with the situation of the stenosis in the coronary tree. The results from semi-automatic detection have generally been satisfactory in images showing a visible structure of connected vessels. The evaluation made with the expert, comparing the results of evaluating three images using this technique and three without using it, has shown that the cases considered as most serious by the expert corresponded with the highest cardiac disease values obtained by the algorithm, and equally for the most serious lesions. These results suggest the suitability of cardiac disease as a diagnostic tool and the impact this represents, and therefore the continuation of this line of work.

This said, the results obtained to date have presented a series of difficulties:

Difficulty of generalization: correction vs. usability: Visually analyzing the result of segmentation, we have been able to observe the extraction of background structures, such as submillimetric vessels or random structures present in the background of the image.

Computational difficulties derived from the type of image: Errors in extraction are mainly due to the lack of local contrast in the image and the lack of a priori anatomical knowledge, making it impossible to distinguish the vascular segments from different areas of the background of the image (such as bones or muscle tissue) when viewed locally; this means what at first appears simple becomes immensely complex when implementing it within the context of a computer model.

Final results strongly dependent on segmentation: The computerized techniques for evaluating coronary illnesses using X-Ray images have to guarantee a high level of accuracy and the ability to be reproduced, necessary in order to evaluate the progress of the illness or to decide if surgery is necessary. In this case, the values for cardiac disease are obtained from a series of calculations based on the segmented image, only depending on the correction and complexity of the segmentation and the diagnostic value of the image selected [5].

The series of images below shows how the processing technique currently in use works:

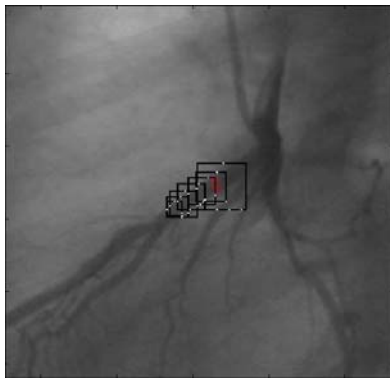


Figure 2. Automatic initiation of coronary vessel tracking

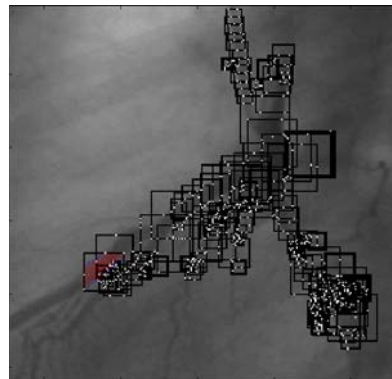


Figure 3. Result of automatically tracking the coronary tree

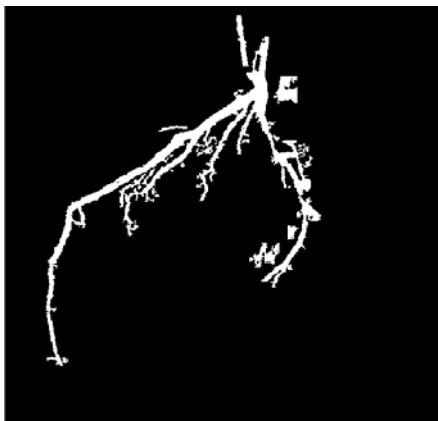


Figure 4. Mask of the tree structure detected

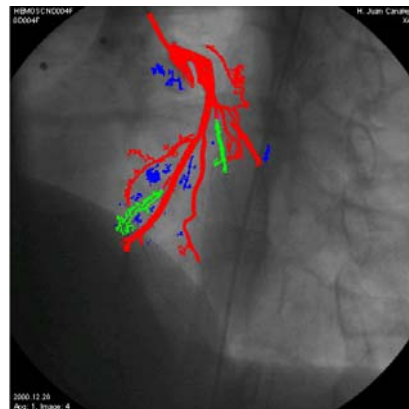


Figure 5. Result of automatic tracking.

automatically

When the system completes the automatic detection of the coronary tree (figure 2 and figure 3), the system presents a mask of the information obtained (figure 4). This mask is then placed over the original image, where the specialist must manually mark the point where the stenosis is located (figure 5). Once the specialist has located the stenosis on the segmented image, the algorithm quantifies its risk by calculating the percentage of pixels affected (in green) from the total number of pixels in the tree (those in green plus those in red). The blue pixels are segmented but not connected to the tree, and therefore are not included in the calculation. The score is provided for each stenosis and the degree to which all of those present are affected, which in this case, without some being contained in others, corresponds to the total of their scores (Table 1):

Table 1. Results of analysis with two stenoses

Stenosis	Risk	Total affected: 11.25%
1	6.92	
2	4.33	

3 Indicators of Results

Personnel being trained:

- a) Presentation of the Doctoral Thesis entitled “**A new perspective on developing an information system for medical images. A tool to support making clinical decisions**”
June 2004
- b) Presentation of the Doctoral Thesis entitled “**Study and validation of advanced segmentation algorithms in biomedical images. Integration in a computer system**”
June 2004

Training has also been given to numerous graduates in Computer Engineering or Technical Computer Engineering by project researchers for their degree projects, in which the main subject is directly related to the project and their results applied to it:

- Development of a system for handling medical images (PACS) based on the DICOM standard and its implantation in the San Rafael Medical Institute of Surgery in A Coruña. January 2005. Definition of an architecture for the integration of digital image processing algorithms with different computing requirements in Web systems. January 2005.
- Study, implementation and testing of safety mechanisms necessary for the transmission of clinical data in DICOM format, in line with Spanish legislation on the protection of personal data. January 2005
- Development of an information system for patients with colorectal cancer, integrated in the global Colon Cancer Family Registry of the National Cancer Institute in the USA. January 2005
- Web environment for developing medical research studies, based on computer measurement of radiological images integrated in a hospital PACS. Example of a study for the implantation of a knee prosthesis. July 2004
- Development of a technique for evaluating coronary stenosis with hemodynamic images. Integration in an Information System for medical images. July 2004.
- Integrated environment in a PACS for editing hemodynamics studies in DICOM format on digital video. September 2004

Another important aspect is the mobility of the team's researchers, visiting other research centres, or welcoming researchers from other centres.

- Visits: Prof. Alejandro Pazos. Jun 2004 and Jun-Jul 2005. Georgia Institute of Tecnology. USA
- Visits from other researchers: Prof. Ezquerro. Georgia Institute of Tecnology. Sept-Oct 2004 and Sept-Dec 2005

The head researcher involved in this project is the coordinator of the project "**A pilot study of Colorectal Cancer in Galicia, Spain**" lasting two years and fully funded by the NIH-USA for a total of \$162,440. This pilot project is taking place as part of Galicia's involvement in a worldwide project with five participants that has been underway for some years, with several millions of Dollars of funding per year.

Papers

- JR. Rabuñal, J. Dorado, A. Pazos, J. Pereira y D. Rivero. "New Approach to the Extraction of Rules from ANN and their Generalization Capacity through GP". **Neural Computation**. MIT Press. 1483-1524. Vol. 16, Número 7. Julio 2004
- AF. Castro, J. Pereira, JM. Vázquez, JA. Pernas, A. Pazos, J. Teijeiro. Remote follow-up system for patients with a hip-prosthesis. Journal: **Telemedicine Journal and e-Health**. Ed. Mary Ann Liebert, Inc. publisher, N 78. Vol. 10, Sup 1-2004. Mayo 2004
- Francisco J. Novoa, Antonio F. Castro, Javier Pereira, Alejandro Pazos. Development of a DICOM server for the reception and storage of medical images in digital format. **WSEAS Transactions on Computer**. 2134-2139. Vol. 3, Issue 6. Diciembre 2004
- M. Míguez, J. Dorado, A. Pazos, G. Bueno, J. Crespo, J. Díaz, J. V. Manjón, J. Pereira, N. Pedreira, J. R. Rabuñal. Almacenamiento, procesado y análisis de imagen en un sistema de Informática Biomédica. Journal: **I + S Informática y Salud**. Ed. MIC. 22,24. Vol. II. Abril 2004

Books and chapters

- J. Dorado, N. Pedreira, JR. Rabuñal, A. Pazos. "**Bioinformática**". Ed. Fundación Alfredo Brañas. Santiago de Compostela. 2004
- J. Pereira, S. Pellit, J. Teijeiro y A. Pazos. "**Medicina y Nuevas Tecnologías: e-Medicina**" in **E-World**. Ed. Fundación Alfredo Brañas. Pp 91,98. Santiago. 2005
- J. Dorado, M. Míguez, AF. Castro, JM. Vázquez, J. Pereira, A. Pazos, J. Teijeiro. **El tratamiento de imágenes en la investigación médica en Informática Biomédica** in **Informática Biomédica**. Ed. INBIOMED. Pp 185-194. Madrid. 2004
- J. Dorado, M. Míguez, AF. Castro, JM. Vázquez, J. Pereira, A. Pazos, J. Teijeiro. **Almacenamiento de imágenes médicas en Informática Biomédica** in **Informática Biomédica**. Ed. INBIOMED. Pp 195-208 Madrid. 2004

Congress

- J. Pereira, JM. Vazquez, FJ. Novoa, N. Vazquez, JM. Vazquez, J. Teijeiro, A. Pazos. Experiences in the Development of a Secure DICOM PACS for a Hemodynamics Unit. **CinC2005** Computers in Cardiology. 25 – 28 sept. 2005. Lyon. France

- FJ. Novoa, AF. Castro, J. Pereira, A. Pazos. Development of a DICOM server for the reception and storage of medical images in digital format in 4th WSEAS International Conference on Applied Informatics and communications (**AIC'04**). 17-19 dec 2004. Tenerife
- FJ. Novoa, J. Pereira, JM. Vázquez Naya, AF. Castro, A. Pazos, J. Teijeiro. "Network Analysis and Design: a Customized Methodology for a DICOM PACS" in **MEDINFO 2004**. 11th World Congress on Medical Informatics. 7-11 sep 2004. San Francisco (USA)
- JM. Vázquez, AF. Castro, J. Pereira, FJ Novoa, A. Pazos, J. Teijeiro. "Application of the DICOM standard to Galicia's Hospital environment" in The 8th World Multi-conference on systemics, cybernetics and informatics. **CISCI2004**. 18-21 jul 2004. Orlando. Florida (USA)
- JM. Vázquez, AF. Castro, J. Pereira, JL. Pérez, J. Teijeiro, A. Pazos. "The impact of design on the implementation of medical imaging information systems" in **CARS 2004**. 18th Int. Congress and Exhibition "Computer Assisted Radiology and Surgery". 23-26 jun 2004. Chicago. USA
- AF. Castro, J. Pereira, JM. Vázquez Naya, JA. Pernas, A. Pazos, J. Teijeiro. "Remote follow-up system for patients with a hip-prosthesis" in **ATA 2004**. Ninth Annual Meeting & Exposition of the American Telemedicine Association. 2-5 may 2004. Tampa. Florida (USA)
- J. R. Rabuñal, J. Dorado, A. Pazos, M. Gestal, D. Rivero, N. Pedreira. "Search the Optimal Architecture, Reduce the Training Set and Make the Process by a Distribute Genetic Algorithm" in IASTED International Conference. **Feb 2004**. Innsbruck (Austria)

Software

- Pazos, J. Teijeiro, J. Pereira, J. Dorado, A. Castro, JM. Vázquez, JL Pérez, S. Pellit. Servidor multithread para el almacenamiento e intercambio de ficheros en un entorno hospitalario©. 15 mar 2004

Acknowledgements

This work was supported in part by the Spanish Ministry of Education and Culture (Ref TIC2003-07593), the IMBIOMED network (Ref G03/160) financed by the Carlos III Health Institute, and grants from the General Directorate of Research of the Xunta de Galicia (Ref. PGIDIT03-PXIC10504PN PGIDIT04-PXIC10503PN, PGIDIT04-PXIC10504PN)

4 References

- [1] J. Pereira. A new perspective on developing an information system for medical images. A tool to support making clinical decisions. [Doctoral Thesis]. A Coruña: Universidade da Coruña; 2004
- [2] Francisco J. Novoa, Antonio F. Castro, Javier Pereira, Alejandro Pazos. Development of a DICOM server for the reception and storage of medical images in digital format. WSEAS Transactions on Computer. 2134-2139. Vol. 3, Issue 6. Diciembre 2004
- [3] J. Pereira, JM. Vazquez, FJ. Novoa, N. Vazquez, JM. Vazquez, J. Teijeiro, A. Pazos. Experiences in the Development of a Secure DICOM PACS for a Hemodynamics Unit. CinC2005 Computers in Cardiology. 25 – 28 sept. 2005. Lyon. France
- [4] A. Castro. Study and validation of advanced segmentation algorithms in biomedical images. Integration in a computer system. [Doctoral Thesis]. A Coruña: Universidade da Coruña; 2004

TIC2003-07593

[5] S. Fernandez. Development of a technique for evaluating coronary stenosis with hemodynamic images. Integration in an Information System for medical images. [Graduate Thesis]. A Coruña: Universidade da Coruña; 2004.