

# IN2GESOFT: Innovation and Integration of Methods for the Development and Quantitative Management of Software Projects TIN2004-06689-C03

José Javier Dolado \*

Universidad del País Vasco - Euskal Herriko Unibertsitatea

Isabel Ramos †

Universidad de Sevilla

Javier Tuya ‡

Universidad de Oviedo

## Abstract

This coordinated project intends to introduce new methods in *software engineering project management*, integrating different quantitative and qualitative technologies in the management processes. The underlying goal to all three subprojects participants is the generation of information adapted for the efficient performance in the directing of the project. The topics that are investigated are related to the capture of decisions in dynamical environments and complex systems, software testing and the analysis of the management strategies for the process assessment of the software in its different phases of the production.

The project sets up a methodological, conceptual framework and supporting tools that facilitate the decision making in the software project management. This allows us to evaluate the risk and uncertainty associated to different alternatives of management before leading them to action. Thus, it is necessary to define a taxonomy of software models so that they reflect the current reality of the projects. Since the *software testing* is one of the most critical and costly processes directed to guarantee the quality and reliability of the software, we undertake the research on the automation of the process of software testing by means of the development of new technologies test case generation, mainly based in metaheuristic and model checking techniques in the domains of database and internet applications. The software system developed will allow the integration of these technologies, and the management information needed, from the first phases of the cycle of life in the construction of a software product up to the last ones such as regression tests and maintenance.

The set of technologies that we investigate include the use of *statistical analysis* and of *experimental design* for obtaining *metrics* in the phase of analysis, the application of the

---

\*Email: [javier.dolado@ehu.es](mailto:javier.dolado@ehu.es)

†Email: [isabel.ramos@lsi.us.es](mailto:isabel.ramos@lsi.us.es)

‡Email: [tuya@uniovi.es](mailto:tuya@uniovi.es)

bayesian nets to the decision processes, the application of the standards of *process evaluation* and quality models, the utilization of metaheuristics algorithms and technologies of prediction to optimize resources, the technologies of visualization to construct control dashboards, hybrid models for the simulation of processes and others.

**Keywords:** Software engineering, project management, cost estimation, metrics, experimentation, project simulation, information visualization, software testing, process assessment, metaheuristics algorithms.

## 1 Project Overview

The IN2GESOFT project is devoted to the study of the different techniques that can be used by project leaders and developers to deal with the management decision problems in Software Engineering. The IN2GESOFT project deals with real problems that appear in software firms. The viewpoint chosen is to focus on the different processes. The basic assumption is that many of the problems in cost, quality and deadlines of software projects can be dealt with in the activities of managing products, processes and resources. The introduction of new methods in those phases of management is of paramount importance for improving the overall performance of the project.

The goal of the project is to explore new methods and technologies in the field of software management and apply them to the assessment and control of the different products, resources, quality and processes which conform the software project. The joint interest of the three IN2GESOFT subgroups is to improve the methods for innovating the methods that software project managers use for getting the project done reliably, on time and within budget.

The project is composed of three subprojects that research into different aspects of the questions about managing software projects. The subproject IN2QUANT is directed towards the quantitative and experimental issues in software engineering. The subproject IN2MED studies different methods for the definition of software management models and the subproject IN2TEST researches into many of the quantitative and experimental questions of the testing and verification processes. The project began on 13-Dec-2004 and will end on 12-Dec-2007. Next is the brief description of each subproject. Figure 1 plots the milestones of the project with the main tasks for each subproject.

### 1.1 Subproject 1: IN2QUANT. *Applying Empirical Techniques and Quantitative Tools to the Effective Management of Software Projects*

Among the topics addressed by this subproject are:

- *Statistical and experimental methods in software engineering.* We use new statistical methods in experimental software engineering in order to reduce the amount of information that is used in decision making. Specifically we use *principal component analysis*, factorial analysis for selecting variables in software experiments. [Tasks *iq3.1*, *iq3.2*, *iq3.3*] Other area of interest is the use of different data mining techniques for improving the estimations performed over software project data bases. The data bases are split using

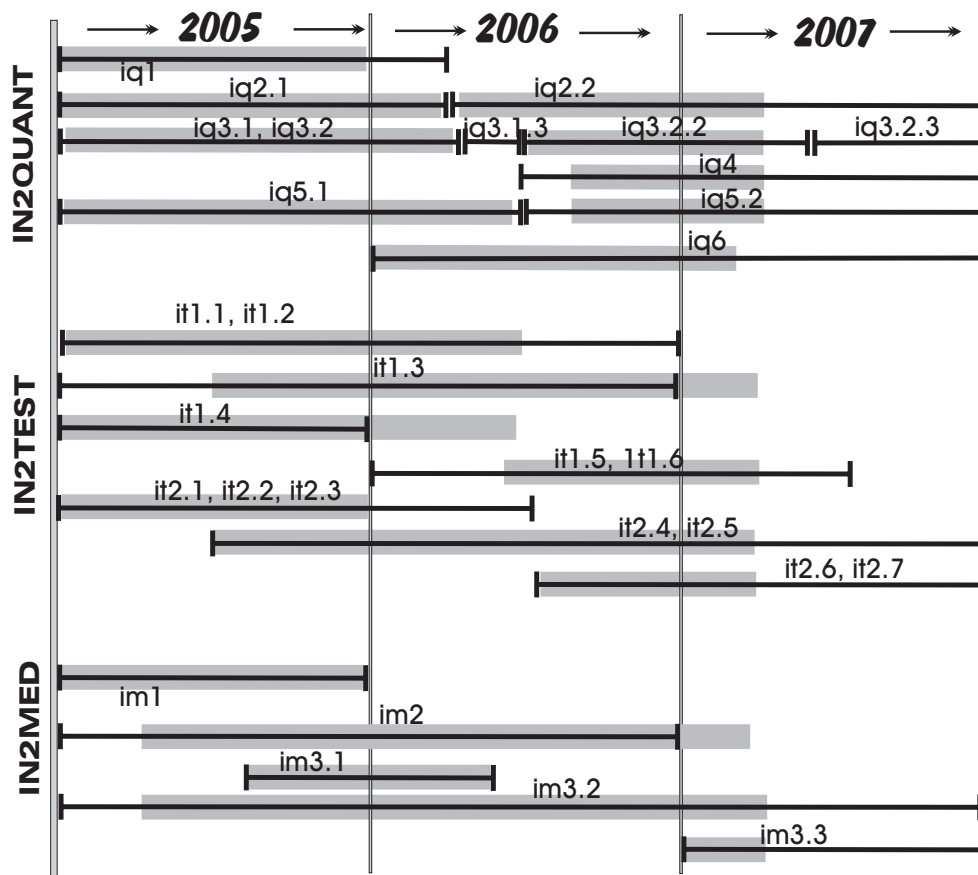


Figure 1: Milestones of the project tasks.

different algorithms in order to increase the predictions capabilities of the data sets [Task *iq3.1*].

Another line of research in this broad area is to characterise a software project data base. In this sense we try to extract the most representative instances of the data base so that the project manager can have a quick overview of the database.

- *Representation and visualization of software projects.* We have a twofold interest in this topic. On one hand there is interest in finding adequate representations for software project management and on the other hand we need to visualize many aspects of the software management process. In the first part of the topic we have worked different new representations for the Gantt charts [Task *iq2.1*]. In the second we are developing different metaphors that may visualize different aspects of the software management tasks [Task *iq2.2*].
- *Ontologies for software projects.* This is a key topic in this subproject. We are developing

ontologies for software engineering and software project management. In the case of the software engineering ontologies, the main question is how to define all terms and how to combine different specific sets of concepts in software engineering [Tasks *iq2.1*, *iq6*].

- *Dashboard for managing systems.* We build dashboards for the dynamic management of strategic decisions. We apply the system dynamics method for building simulators that may enhance the abilities of the project manager. Dashboards built with this method provide a flexible way of interacting with different potential actual situations [Tasks *iq2.2*, *iq6*].
- *Software process assessment.* We have studied different models of software process assessment, comparing different aspects of the Capability Maturity Model (CMM) and the standard SPICE (ISO/IEC TR-15504), Software Process Improvement and Capability dEtermination. The latter is used as the reference model for applying it in our research [Task *iq5*].
- *New cost estimation models and improvement of software project databases.* In this aspect we elaborate new parametric models combining regression models with fuzzy systems and other techniques [Task *iq3*].

## 1.2 Subproject 2: IN2MED. *Taxonomy of models for software processes measurement and assessment*

The main objectives of the IN2MED subproject are:

- *Taxonomy of software models and definition of the processes for data collection.* Our intent is to analyze the current typology of the mostly common software projects of our industrial environment. From this analysis and by using the processes applied within their development cycle, different categories of software projects will be identified [Task *im1*].
- *Proposal for a methodological and conceptual framework to help in the decision making process regarding the management and assessment of software processes.* The proposal includes the development of machine learning techniques, prediction techniques, dynamic simulation modelling and process modelling according to the international standards of maturity models, total quality and self-evaluation of software processes [Tasks *im2*, *im3.1*].
- *Conceptual framework and their support tools.* The methodological framework proposed should have, among others, the following features: web application architecture, flexible enough as to accept new modules to add extra functionality regarding control elements, new variables and estimation, management and visualization procedures [Tasks *im3.2*, *im3.3*].

## 1.3 Subproject 3: IN2TEST. *Integration of automated testing techniques in the software development & maintenance process*

The topics addressed by this subproject are mainly related to Test Automation:

- *Test case generation with metaheuristic techniques.* Metaheuristic algorithms are a way to find good enough solutions for complex optimization problems. For test case generation the approach consists on represent the cost function to be minimized in terms of the fulfilment of a given coverage criteria. Tabu Search and Scatter Search are used for develop test cases using structural coverage criteria [Tasks 1.1, 1.2, 2.1].
- *Test case generation with model checking.* Model checkers are able to verify properties in a model. If properties are not satisfied a counterexample is shown. When used for test case development the approach consists on specify the test requirements in terms of a property that is negated, and represent the program or specification under test as a model for the model checker. This approach is used for develop test cases for BPEL specifications and XPath queries [Task 1.3].
- *Adequacy criteria for database testing.* Although many adequacy criteria have been developed for many programming languages, few work has been conducted in the development of adequacy criteria for database queries. In this case specifically tailored criteria are developed for SQL (structural and mutation based criteria) and for XPath queries [Tasks 1.4, 1.5].
- *Integration of tools and experimentation.* Several tools to support the above approaches have been developed to support the process of test case generation and evaluation. Additionally, controlled experiments are performed to assess the fault detection ability of the techniques [Tasks 2.2 to 2.7].

## 2 Level of Success of the Project

Part of the results obtained are listed in the section References, and here we detail some of the results that we have obtained. The figure 2 plots the topics that the members of the subprojects are researching.

### 2.1 Topics Developed in the IN2QUANT subproject

- *Characterization of software project databases* In [15] we have proposed a new method for global understanding of the software project database. We have found that it is possible to have a smaller database with less instances of the database, and this helps the project manager to deal with the data interpretation tasks.
- *Improving the parametric methods in software cost estimation.* We have applied different new methods for the problem of software cost estimation, including bayesian nets, fuzzy variables and other algorithms. Clustering has been a tool for finding also new equations for estimation. All results obtained improvements in the estimation process [14][12][13].
- *Managing projects with system dynamics.* A system dynamics model with several feedback loops represents different hypothesis for managing systems [16].
- *Representation in software projects.* The Gantt chart is augmented with new information related to the participants in the project [17].

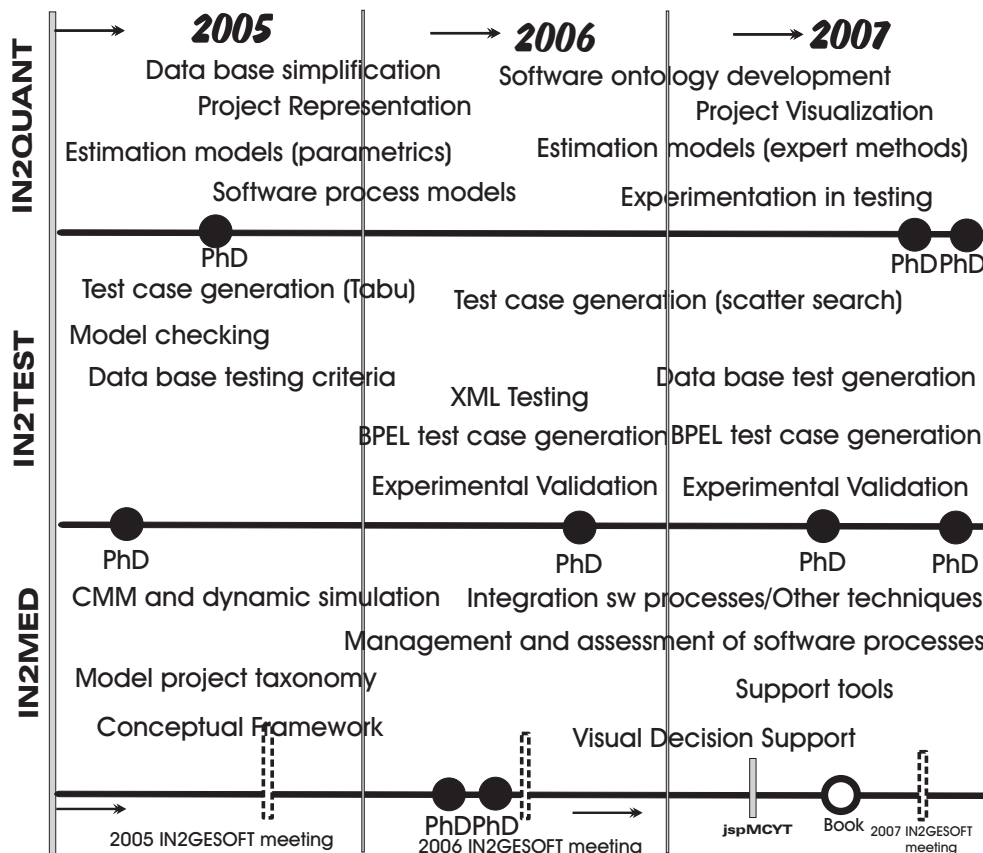


Figure 2: Topics developed and topics in development.

- *New topics in software management: visualization and coordination.* We are developing a new metaphora for the visualization of different activities in management. An initial approach has been proposed after reviewing the related literature. Also, we have presented an initial work dealing with the new problems in coordination in software projects.
- *Experimentation in software engineering.* Different experiments have been carried out and new analysis have been performed [10] [11].

## 2.2 Topics Developed in the IN2MED subproject

- *Simulation modelling for understanding and improving interactive systems usability.* The developed dynamic model helps visualize the behaviour of User-Centred Design (UCD) activities during the development life cycle of a software system. To model the UCD process the process model described in ISO standard 13407:1999 has been used as a reference [26] [27].

- *Integrating software process simulation models with other techniques.* The integration of continuous modelling with CMM, static and algorithmic models and the design of a metrics collection program have been collected in [22] [25]. In [23] a software project simulator based on System Dynamics is used to analyze the different sources of error in the initial estimates of parametric models.
- *Qualitative modelling and decision making in real environments.* The set of algorithms and tools previously developed have been improved and adapted to obtain qualitative models based on fuzzy rules from data coming from real environments [24].
- *Feature Selection.* We characterize several software engineering databases selecting attributes with the final aim that project managers can have a better global vision of the data they manage. The results show that in general, the smaller databases maintain the prediction capability with a lower number of attributes than the original databases [18] [15].
- *Data Streams.* Software Project Simulation has lead to massive data generation in a data streams manner. In this field, we have developed a decision support system (DDS) that is aimed to provide knowledge models from project management data. Experimental results show an excellent performance of our approach as both a general purpose method and a software project DDS, particularly [19] [21].
- *Visual Decision Support.* Visual data exploration techniques provide the user with graphic views or metaphors that represent potential patterns and data relationships. In this field, we have developed a visual exploration technique – named VETIS - that automatically extracts relevant attributes and displays their interesting measures in order to support two decision support tasks: classification and feature selection [20].

### 2.3 Topics Developed in the IN2TEST subproject

- *Test case generation with metaheuristic techniques.* We have explored the feasibility of using techniques (alternative to the genetic algorithms) such as Tabu Search and Scatter Search. [3]. The most outstanding result is that the local search performed by Tabu Search improves the finding at the latter stages the most difficult test cases faster than the others. On the other hand, Scatter Search has been developed to perform the same task [4]. Performance of Scatter Search improves the finding of test cases at early stages of the generation. Both approaches are then complementary and its joint use is currently being explored.
- *Test case generation with model checking.* The use of model checking has been previously used for the verification of state-based systems [5]. The research has focused on the generation of conformance test suites for BPEL compositions of web services [1] [9], using the BPEL specification as the only information. In the same line of work, but applied to data repositories, automatic generation of test cases for XML queries specified in XPath have obtained promising results.
- *Adequacy criteria for database testing.* The first line of work is the research on coverage criteria for testing SQL queries based on a structural view of the query, extended with

the tailoring of criteria based on MCDC and equivalence classes [6]. The second approach to the evaluation of coverage of SQL queries is based on mutation analysis. Mutation operators have been developed [2] and integrated in a mutation tool [7]. The Third approach is driven towards the development of test cases for semi-structured data stored in XML documents, based on Category Partition for developing test cases for XPath queries [8].

- *Tool development.* The tools for generating test cases using Tabu Search and Scatter Search, first developed separately, have been refactored to allow the generation using a common framework. A tool named SQLTest has been developed to support training and experimentation in SQL test case development, and SQLMutation to support the generation of the SQL mutants developed [7]. Both tools are publicly available and the latter one is also accessible as a web service. The development of similar tools for testing of XML and BPEL is an ongoing work.
- *Experimental validation.* A series of four controlled experiments with students have been conducted to evaluate the fault detection capability of database test cases written using different criteria. The first analysis of data collected has shown that test cases developed for SQL using the criteria above mentioned are most effective than conventional approaches such as equivalence partitioning.

### 3 Outcomes of the project and main results

Table 1 summarizes the results, so far, of the project. The figures present the works and activities in which the members of the project have taken part in.

One of the most important results of the project, from the perspective of the heads of the subprojects, is the book **Técnicas Cuantitativas para la Gestión en la Ingeniería del Software**, that has 16 chapters, around 400 pages and has a participation of 35 authors. It will be published during 2007. The book is the expression of the many techniques in which the members of the project are experts at. The content of the book describes topics such as: *software process assessment, CMM, SPICE, process simulation, project representation, cost estimation by machine learning, system support for project management, principal component analysis, experimental software engineering, bayesian nets, software estimation by fuzzy methods, data mining for decision support, software test methods and software test management.*

#### 3.1 Publications in Journals and Books

The References show part of the articles that have been published by the members of the project. The journals targeted have been, among others: Journal of Systems and Software, Information and Software Technology, Computers and Operations Research, International Journal of Engineering Intelligent Systems for Electrical Engineering and Communications, SIGPLAN Notices, IEEE Latin America Transactions, Procesos y Métricas de las Tecnologías de la Información, etc.



## 3.2 Conferences

The conferences, workshops and seminars in which the results have been presented are, among others:

[2005] ADIS 2005, CAISE05, IWANN05, SACACM05, WIST05, WASVWS05, IDA05, PROSIM05, IWSM05, GECCO05, IMWA05, IADIS 2005, MIC 2005, JISBD 2005

[2006] ADIS 2006, JISBD 2006, ICEIS06, ICEISDC06, ICWIST06, WSTV06, KES06, SACACM06, ICCI06, CISTI06, DS06, SQM06, Mutation 2006, ICSEA 2006, TAIC-PART 2006, WS-MaTe 2006, ICKEDS 2006, STV 2006, RedBD 2006, PRIS 2006, JSWEB 2006, ONTOSE 2006.

## 3.3 Events organized

ADIS 2005, PRIS 2006, ADIS 2006, the Session *Ontologies for Software Engineering* (SEKE06), and the International Conference on Software Process and Product Measurement MENSURA 2006.

## 3.4 Research Grants

The subproject IN2QUANT has a researcher under the FPI-MEC grant associated to the project.

## 3.5 Ph.D. Dissertations

The following researchers, members of IN2GESOFT, have defended their corresponding Ph.D.:

- María José Suárez Cabal, Mejora de casos de prueba en aplicaciones con bases de datos utilizando medidas de cobertura de sentencias SQL (2006)
- María Eugenia Díaz Fernández, Generación automática de pruebas estructurales de software mediante Búsqueda Tabú (2005).
- Roberto Ruiz Sánchez, Heurísticas de Selección de Atributos para Datos de Gran Dimensión (European Thesis) 2006.
- Francisco Ferrer Troyano, Aprendizaje Incremental y Modelado Visual de Reglas de Decisión (2006).
- Antònia Mas Pichaco, Un Nuevo Modelo para la Implantación de un Sistema de Gestión de Calidad en Pymes de Desarrollo de Software basado en SPICE (ISO/IEC 15504) (2005).

Two members obtained during 2005 the special prize for PhDs at their corresponding universities

- Claudio de la Riva, Premio Extraordinario de Doctorado (2005).
- Mari Carmen Otero Vidal, Premio Extraordinario de Doctorado (2005).

### 3.6 Collaborations with groups and other projects

There have been several collaborations with researchers at other universities: Université du Québec at Montréal and University of Reading at UK.

In 2005 a Marie Curie EU International Incoming Fellowship was submitted and it was approved in 2006. The recipient is *Prof. Alain Abran* from the Université du Québec à Montréal who will visit the University of the Basque Country for the next two years. The members of the project have promoted the following activities: RePRIS (a network for software testing), and projects OPTISOFT and HEMPIC2.

Type of contribution	Number
International Journals and book chapters	32
Book chapters	7
National Journals	12
Ph.D. Dissertations	5
International Conferences	55
National Conferences	26
International Program Committees	9
National Program Committees	6
Events organized	5
PhD Students under FPI grants	1

Table 1: Current results of the project.

## References

- [1] José García-Fanjul, Claudio de la Riva and Javier Tuya, Testing BPEL compositions of web services using SPIN. *Journal of Systems and Software*, (Under revision, selected from WS-MaTe 2006).
- [2] Javier Tuya, M<sup>a</sup> José Suárez-Cabal, Claudio de la Riva, Mutating database queries. *Information and Software Technology*, (In press, July 2006)
- [3] Eugenia Díaz, Javier Tuya, Raquel Blanco, Javier Dolado, A tabu search algorithm for Structural Software Testing. *Computers and Operations Research* (focus issue on Search-Based Software Engineering), (Accepted).
- [4] Raquel Blanco, Javier Tuya, Eugenia Diaz, Adenso Diaz, A Scatter Search approach for the Automated Branch Coverage in Software Testing. *International Journal of Engineering Intelligent Systems*, (Accepted).
- [5] Claudio de la Riva, Javier Tuya, Automatic Generation of Assumptions for Modular Verification of Software Specifications. *Journal of Systems and Software*, 79 (9), pp. 1324-1340, September 2006.

- [6] Javier Tuya, M<sup>a</sup> José Suárez-Cabal, Claudio de la Riva, A practical guide to SQL white-box testing. *SIGPLAN Notices*, 41(4), pp. 36-41, April 2006.
- [7] Javier Tuya, M<sup>a</sup> Jose Suarez-Cabal and Claudio de la Riva, SQLMutation: a Tool to Generate Mutants of SQL Database Queries. Second Workshop on Mutation Analysis (Mutation2006). Raleigh, North Carolina, November 2006.
- [8] Claudio de la Riva, José García-Fanjul, Javier Tuya, A Partition-based Approach for XPath Testing. *International Conference on Software Engineering Advances*. Tahiti, October 2006.
- [9] José García-Fanjul, Claudio de la Riva, Javier Tuya, Generation of Conformance Test Suites for Compositions of Web Services Using Model Checking. *Testing: Academic & Industrial Conference - Practice and Research Techniques (TAIC-PART)*. Windsor, Agosto 2006.
- [10] M.C. Otero and J.J. Dolado, An Empirical Comparison of the Dynamic Modeling in OML and UML, *Journal of Systems and Software*, vol. 77, 91-102, 2005.
- [11] D Rodríguez, MA Sicilia, JJ Cuadrado, D Pfahl, e-Learning in Project Management using Simulation Models: A Case Study based on the Replication of an Experiment, *IEEE Transactions on Education*, 2006. IEEE Education Society.
- [12] Cuadrado-Gallego, J.J., Fernández, L., Sicilia, M.A.. Enhancing input value selection in parametric software cost estimation models through second level cost drivers. *Software Quality Journal*, 2006.
- [13] Rodríguez, D., Cuadrado-Gallego, J.J., Garre, M., Sicilia, M.A. Multiattribute clustering in software effort estimation parametric models. *Journal of Computer Science Technology*, 2006.
- [14] Cuadrado-Gallego, J.J., Sicilia, M.A., Rodríguez, D., Garre M. An empirical study of process-related attributes in segmented software cost-estimation relationships. *Journal of Systems and Software*, 2006, 3 (79), 351-361.
- [15] Dolado, J., Rodríguez, D., Riquelme, J., Ferrer, F., Cuadrado-Gallego, J.J. A two stage zone regression method for global characterisation of a project database. In *Advances in Machine Learning Application in Software Engineering*, Idea Group Publishing, 2007, 0-0. ISBN: 1-59140-941-1.
- [16] Morlán, I., Nieto, M., Davalillo, A., Dolado, J. A System Dynamic model to facilitate understanding and building stakeholder support for new management strategies in universities 9th Toulon-Verona Conference. *Quality in higher Education, Health Care and local Government*, Paisley. Escocia (Reino Unido), Septiembre, 2006
- [17] Fran J Ruiz-Bertol, Javier Dolado: Human resource representation in Gantt charts, *Software Quality Management XIV: Perspectives in Software Quality (SQM 2006)*, pp. 219-230. 2006.

- [18] Giráldez Raúl, JS Aguilar-Ruiz, Riquelme JC. Knowledge-based Fast Evaluation for Evolutionary Learning. *IEEE Systems, Man and Cybernetics Part C*. Vol. 35, n° 2, pp. 254-261, 2005. Ed. IEEE Press ISSN 0018-9472.
- [19] Ferrer-Troyano Francisco, Aguilar-Ruiz JS, Riquelme JC. Incremental Rule Learning and Border Examples Selection from Numerical Data Streams. *Journal of Universal Computer Science*. Vol. 11, n° 8, pp. 1426-1439, 2005. Ed. Know-Center Springer ISSN 0948-695x.
- [20] Ferrer-Troyano Francisco, Aguilar-Ruiz JS, Riquelme JC. Connecting Segments for Visual Data Exploration and Interactive Mining of Decision Rules. *Journal of Universal Computer Science*. Vol. 11, n° 11, pp. 1835-1848, 2005. Ed. Know-Center Springer ISSN 0948-695x.
- [21] Giráldez, R. "Improving the performance of Evolutionary Algorithms for Decision Rules Learning", *AI Communications*, Vol. 18 (1), pp. 63-65, 2005.
- [22] Ruiz M., Ramos I., Toro M. Software Process Dynamics: Modeling, Simulation and Improvement. *New Trends in Software Process Modelling*. Vol. 18, pp. 21-56, World Scientific Publishing Company, 2006.
- [23] Mercedes Ruiz Carreira, JJ Cuadrado Gallego, D Rodríguez García, MA Sicilia. Using Simulation to determine the sensitivity of error sources. *Internacional Workshop on Software Measurement (IWSM'05)*, pp. 1-10, 2005, Montreal (Canada).
- [24] de Toro F.J, Aroba J, López J.M. Fuzzy-Genetic Methodology for web-based computer-aided diagnosis in medical applications. *IX International Conference on Computational Intelligence*. Dortmund (Alemania), 2006.
- [25] de la Villa M, Ruiz-Carreira M, Ramos I. Un estudio crítico comparativo de ISO 9001, CMMI e ISO 15504. *Conferência Ibérica de Sistemas e Tecnologias de Informação*. Ofir (Portugal). Junio 2006.
- [26] Nuria Hurtado Rodríguez, Mercedes Ruiz Carreira, Jesús Torres Valderrama. Modelado y simulación como herramienta para la mejora de la usabilidad. *Actas del VI Congreso de interacción persona-ordenador*, pp. 1-4, 2005, Granada.
- [27] Nuria Hurtado Rodríguez, Mercedes Ruiz Carreira, Jesús Torres Valderrama. Towards Interactive Systems Usability improvement through simulation modeling. *6th International Workshop on Software Process Simulation and Modeling (PROSIM'05)*, pp. 117-125, 2005, San Louis, MO (USA).