Description
Download KEEL Software Tool
KEEL-dataset Data set repository
KEEL Included Algorithms
Publications
Links
Workshops
Identification Access

KEEL (Knowledge Extraction based on Evolutionary Learning) is an open source (GPLv software tool which empowers the user to assess the behavior of evolutionary learning a Computing based techniques for different kinds of DM problems: regression, classification, clu pattern mining and so on. KEEL is being developed under the Spanish National Projects T 04036-C05, TIN2005-08386-C05 and TIN2008-06681-C06 with the collaboration of the six f Spanish Research Groups:

SCI^{2}S (Spanish National Proyects TIC2002-04036-C05-01, TIN2005-08386-C05-01 and TIN2008-06681-C06-01)

Ayrna (Spanish National Proyects TIC2002-04036-C05-02, TIN2005-08386-C05-02 and TIN2008-06681-C06-03)

GRSI (Spanish National Proyects TIC2002-04036-C05-03, TIN2005-08386-C05-04 and TIN2008-06681-C06-05)

Intelligent Systems and Data Mining (Spanish National Proyects TIC2002-04036-C05-04, TIN2005-08386-C05-03 and TIN2008-06681-C06-02)

Metrology and Models (Spanish National Proyects TIC2002-04036-C05-05, TIN2005-08386-C05-05 and TIN2008-06681-C06-04)

Inte Syste Data (Sp Nationa TIN200 C0

If you want to refer to KEEL in a publication, please cite us using the following references:

KEEL description papers:

J. Alcalá-Fdez, L. Sánchez, S. García, M.J. del Jesus, S. Ventura, J.M. Garrell, J. Otero C. Romero, J. Bacardit, V.M. Rivas, J.C. Fernández, F. Herrera. KEEL: A Software Tool t Assess Evolutionary Algorithms to Data Mining Problems. Soft Computing 13:3 (2009 307-318, doi: 10.1007/s00500-008-0323-y.

J. Alcalá-Fdez, A. Fernandez, J. Luengo, J. Derrac, S. García, L. Sánchez, F. Herrera KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms an Experimental Analysis Framework. Journal of Multiple-Valued Logic and Soft Computin 17:2-3 (2011) 255-287.

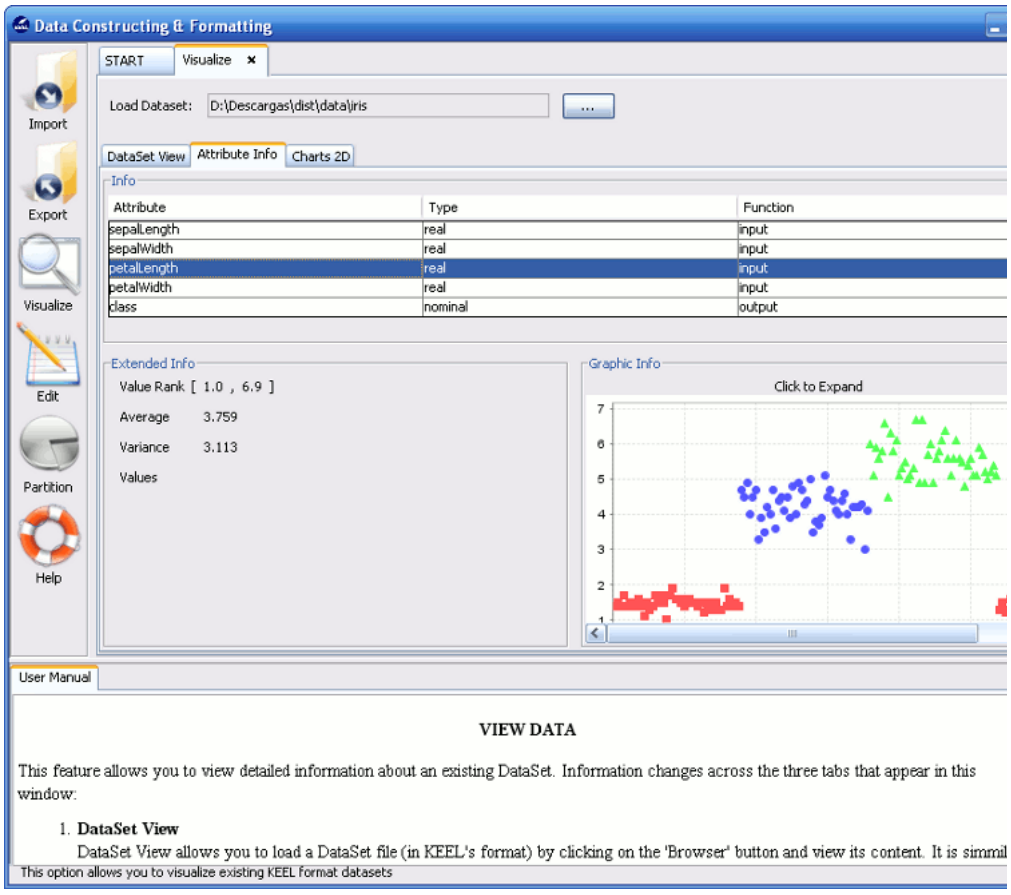
Contents

KEEL description

KEEL is a software tool to assess EAs for DM problems including regression, classification, clu pattern mining and so on. The version of KEEL presently available consists of the following blocks:

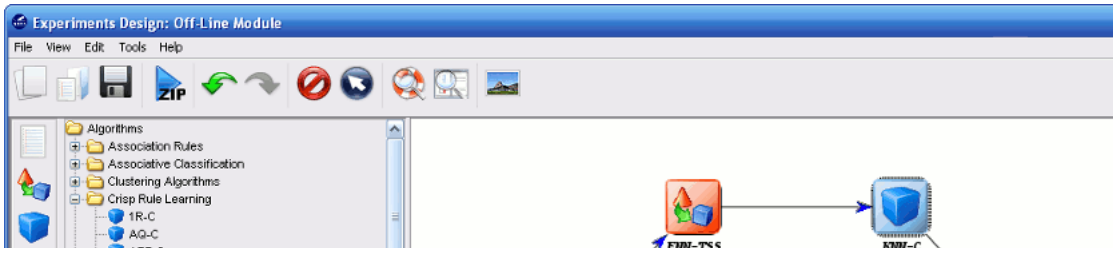
- *Data Management*

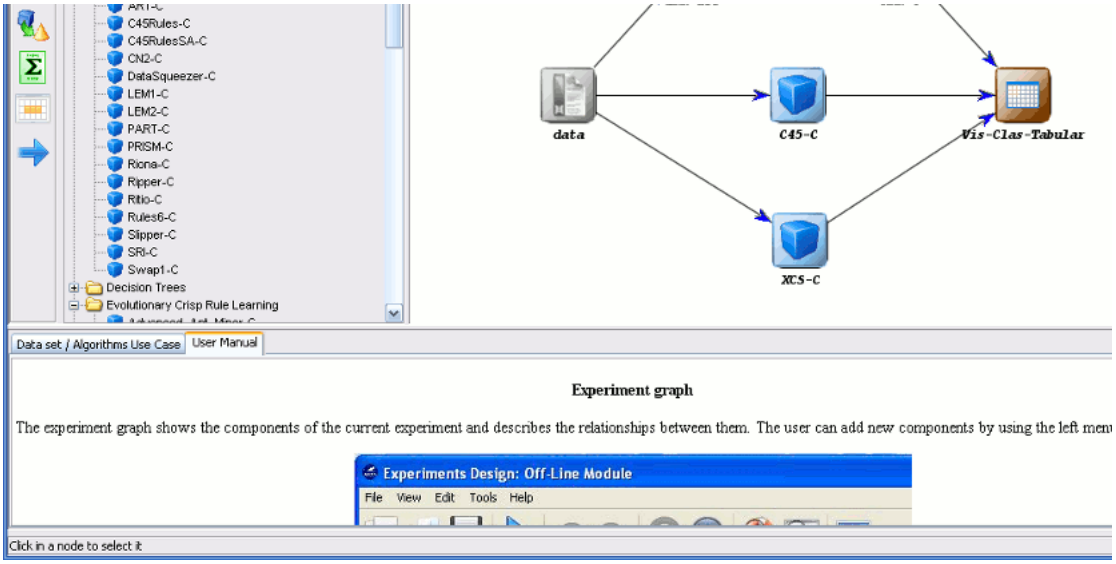
  This part is composed by a set of tools that can be used to build new data, export an data in other formats to KEEL format, data edition and visualization, apply transfor and partitioning to data, etc...



- *Design of Experiments*

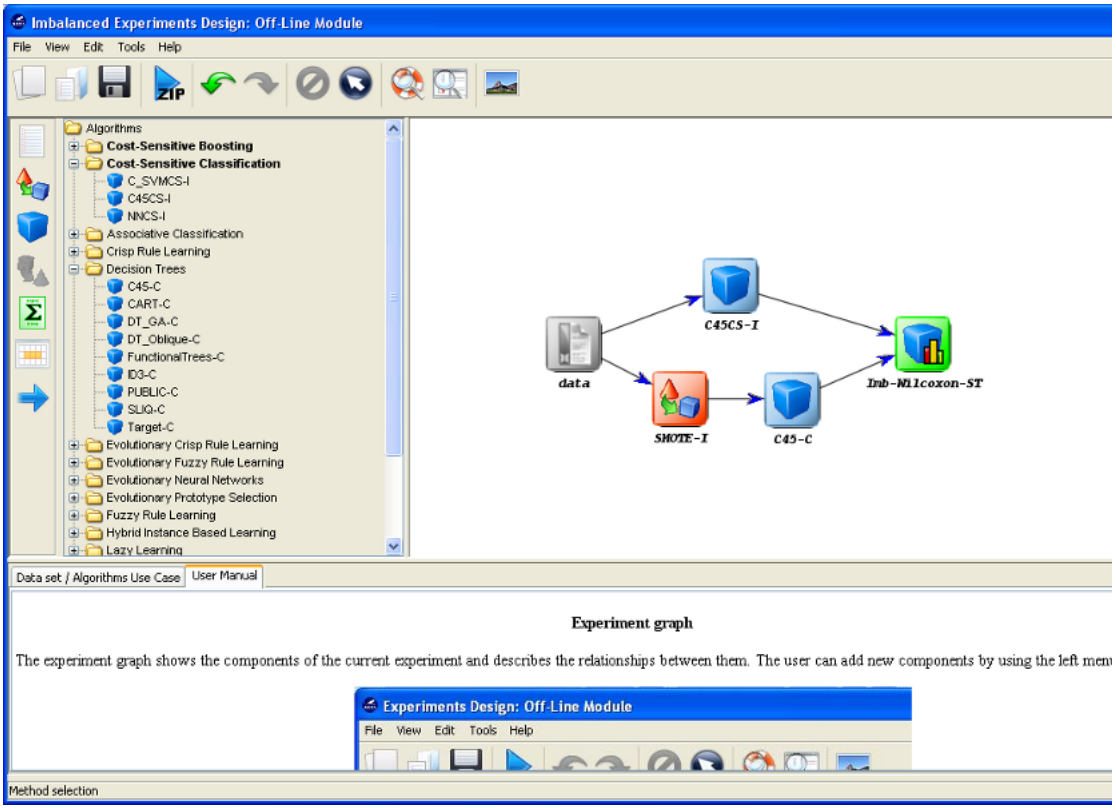
  The aim of this part is the design of the desired experimentation over the selected da It provides options for many choices: type of validation, type of learning (classi regression, unsupervised learning, subgroup discovery), etc...





- *Design of Imbalanced Experiments*

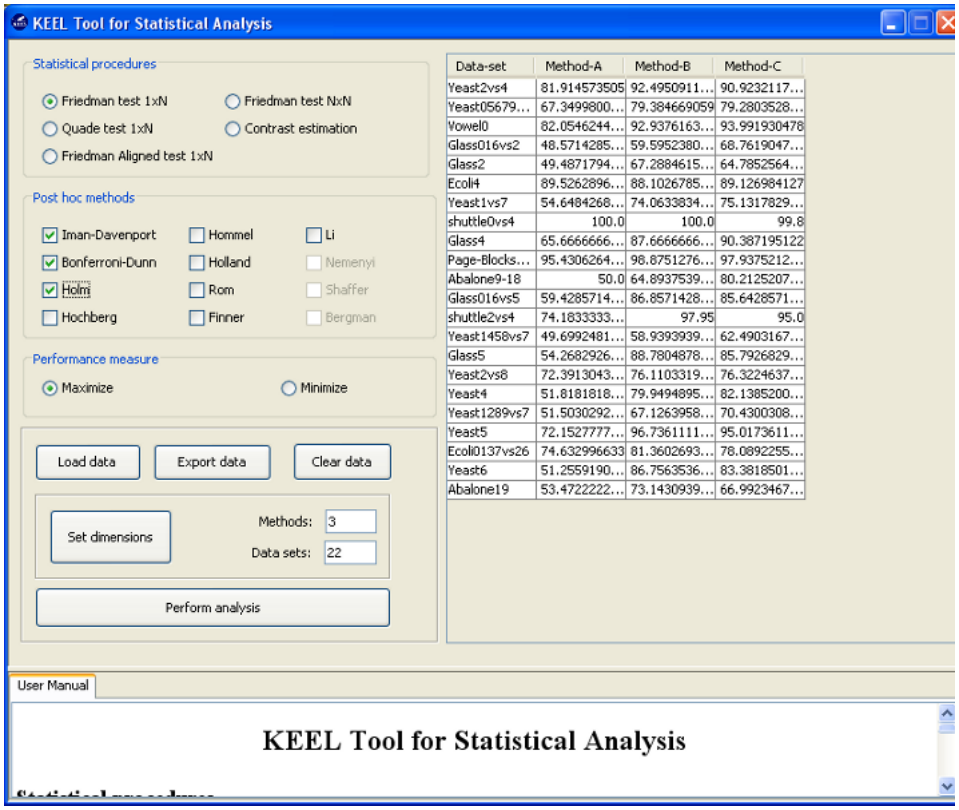
  The aim of this part is the design of the desired experimentation over the imbalanced data sets. These experiments are created for 5cfv datasets and include algorithms for imbalanced data and general classification algorithms.



- *Statistical Tests*

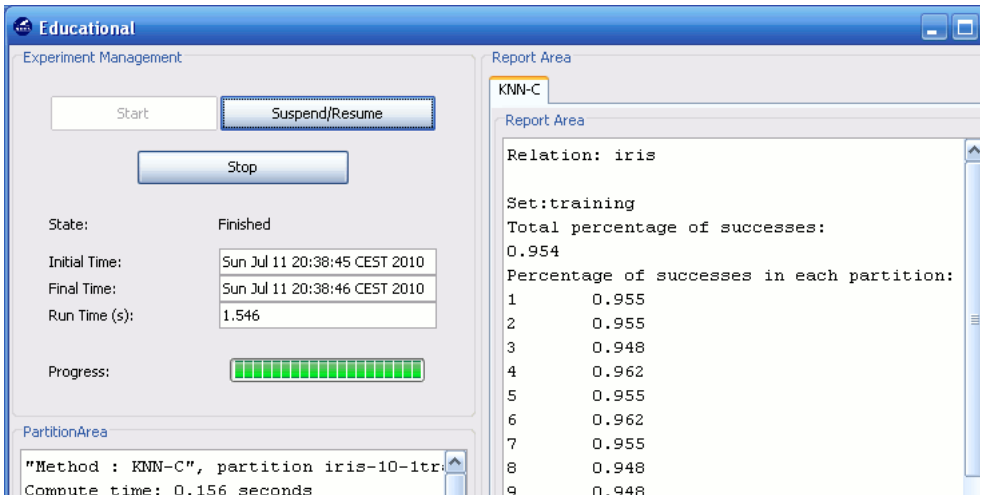
  KEEL is one of the fewest Data Mining software tools that provides to the resea complete set of statistical procedures for pairwise and multiple comparisons. Inside t environment, several parametric and nonparametric procedures have been codec

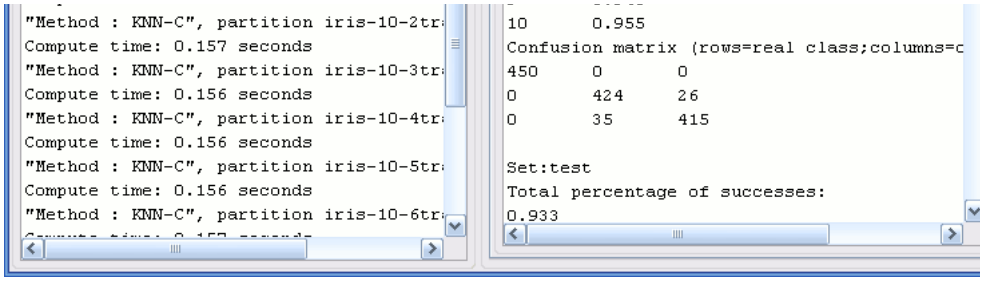
should help to contrast the results obtained in any experiment performed with the s
tool.



- *Educational Experiments*

  With a similar structure to the *Design of Experimets* part, allows us to design an exp
  which can be step-by-step debugged in order to use this as a guideline to show the
  process of a certain model by using the platform with educational objectives.





Taking into account each one of the function blocks, KEEL can be useful by different types which expect to find determined features in a Data Mining (DM) software.

In the following, we describe the user profiles who it is designed for, its main features different ways of working integrated in the software tool.

User Profiles

KEEL is an integration of an environment with a defined architecture and a developr knowledge extraction as expandable modules. It is mainly intended for two categories o researchers and students. Either group has a different set of needs:

- *KEEL as a research tool*

  The most common use of this tool for a researcher will be the automated exec experiments, and the statistical analysis of their results. Routinely, an experimenta includes a mix of evolutionary algorithms, statistical and AI-related techniques. Spe was taken to make possible that a researcher can use KEEL to assess the relevanc own procedures. Since the actual standards in machine learning require heavy compu work, the research tool is not designed to offer a real-time view of the progres algorithms, it is designed to rather generate a script and be batch-executed in a cl computers. The tool allows the researcher to apply the same sequence of pre-pro experiments and analysis to large batteries of problems and focus his attentior summary of the results.

- *KEEL as an educational tool*

  The needs of a student are quite different to those of a researcher. Generally speak objective is no longer that of making statistically sound comparisons between alg There is no need of repeating each experiment a large number of times. If the tool used in class, the execution time must be short and a real-time view of the evolutio algorithms is needed, since the student will use this information to learn how to ad parameters of the algorithms. In this sense, the educational tool is a simplified versio research tool, where only the most relevant algorithms are available. The execution in real time. The user has a visual feedback of the progress of the algorithms, and ca the final results from the same interface used to design the experimentation.

Both types of user require an availability of a set of features in order to be interested in usin Then, this is when we describe the main features of the KEEL software tool.

Main Features

KEEL is a software tool developed to ensemble and use different DM models. We would like to that this is the first software toolkit of this type containing a library of evolutionary algorithms with open source code in Java. The main features of KEEL are:

- Evolutionary Algorithms (*EAs) are presented in predicting models, pre-processing (evol feature and training set selection) and post-processing (evolutionary tuning of fuzzy rules).*

- *It includes data pre-processing algorithms proposed in specialized literature: data transfo discretization, training set selection, feature selection, imputation methods for missing val noisy data filtering methods.*

- *It has a statistical library to analyze algorithms' results. It comprises a set of statistical t analyzing the normality and heteroscedasticity of the results and performing parametric a parametric comparisons among the algorithms.*
- *Some algorithms have been developed by using a Java Class Library for Evolutionary Com (JCLEC)*
- *It provides an user-friendly interface, oriented to the analysis of algorithms.*
- *The software is aimed to create experimentations containing multiple data sets and alg connected among themselves to obtain a result expected. Experiments are independentl generated from the user interface for an off-line run in the same or other machines.*
- *KEEL also allows to create experiments in on-line mode, aiming an educational support in learn the operation of the algorithms included.*
- *It contains a Knowledge Extraction Algorithms Library, remarking the incorporation of evolutionary learning algorithms, together with classical learning approaches. Th employment lines are:*
    - *Different evolutionary rule learning models have been implemented*
    - *Fuzzy rule learning models with a good trade-off between accuracy a interpretability.*
    - *Evolution and pruning in neural networks, product unit neural networks, and rad base models.*
    - *Genetic Programming: Evolutionary algorithms that use tree representations f extracting knowledge.*
    - *Algorithms for extracting descriptive rules based on patterns subgroup discove have been integrated.*
    - *Data reduction (training set selection, feature selection and discretization). EAs f data reduction have been included.*

Members of the KEEL project

University of Granada (Subproject TIN2008-06681-C06-01)

The head of group is *Ph.D. Francisco Herrera Triguero* of *Department of Computer Science and Artificial In* University of Granada.

This group is the **coordinating node** in the development of KEEL.

Name	E-mail	Institution
Ph.D. Herrera Triguero, Francisco (IP)	herrera@decsai.ugr.es	University of Granada
Ph.D. Alcalá Fernández, Jesus	jalcala@decsai.ugr.es	University of Granada
Ph.D. Alcalá Fernández, Rafael	alcala@decsai.ugr.es	University of Granada
Ph.D. Bull, Larry	larry.bull@uwe.ac.uk	University of the west of England, Br
Ph.D. Casillas Barranquero, Jorge	casillas@decsai.ugr.es	University of Granada
Mr Derrac Rus, Joaquin	carrdelling@gmail.com	University of Granada
Ph.D. Fernández Hilario, Alberto	alberto@decsai.ugr.es	University of Jaén
Ph.D. García López, Salvador	salvagl@decsai.ugr.es	University of Jaén
Ms. López Morales, Victoria	vlopez@decsai.ugr.es	University of Granada
Mr. Luengo Martín, Julián	julianlm@decsai.ugr.es	University of Granada

Ph.D. Martínez López, Francisco José	fjmlopez@ugr.es	University of Granada
Ph.D. Pérez Márquez, Elena	elena@eis.uva.es	University of Valladolid
Mr. Sáez Muñoz, José Antonio	smja@decsai.ugr.es	University of Granada
Mr. Triguero Velázquez, Isaac	triguero@decsai.ugr.es	University of Granada
Ph.D. Villar Castro, Pedro	pvillarc@ugr.es	University of Granada

University of Jaen (Subproject TIN2008-06681-C06-02)

The head of group is *Ph.D. María José Del Jesus* of *Department of Computer Science* of the University of Jaen.

Name	E-mail	Institu
Ph.D. Del Jesus, María José (IP)	mjjesus@ujaen.es	University of Jaé
Mr. Aguilera García, José Joaquín	jjaguile@ujaen.es	University of Jaé
Mr. Berlanga Rivera, Francisco José	berlanga@ujaen.es	University of Jaé
Ph.D. Cano de Amo, Jose Ramon	jrcano@ujaen.es	University of Jaé
Ph.D. González García, Pedro	pglez@ujaen.es	University of Jaé
Ph.D. Pérez Godoy, Mº Dolores	lperez@ujaen.es	University of Jaé
Ph.D. Rivas Santos, Víctor Manuel	vrivas@ujaen.es	University of Jaé
Ph.D. Rivera Rivas, Antonio Jesús	arivera@ujaen.es	University of Jae

University of Cordoba (Subproject TIN2008-06681-C06-03)

The head of group is *Ph.D. Sebastián Ventura Soto* of *Department of Computer Science and Numerical* University of Cordoba.

Name	E-mail	Institution
Ph.D. Ventura Soto, Sebastián (IP)	sventura@uco.es	University of Córdoba
Mr. Fernández Caballero, Juan Carlos	fernandezcaballero@gmail.com	University of Córdoba
Mr. González Espejo, Pedro	pgonzalez@uco.es	University of Córdoba
Mr. Gutiérrez Peña, Pedro Antonio	pagutierrez@uco.es	University of Córdoba
Ph.D. Hervás Martínez, César	chervas@uco.es	University of Córdoba
Ph.D. Martínez Estudillo, Francisco	fjmestud@etea.com	INSA-ETEA
Ph.D. Martínez Estudillo, Alfonso Carlos	acme@etea.com	INSA-ETEA
Ph.D. Pechenizkiy, Mykola	m.pechenizkiy@tue.nl	Eindhoven University of Tecl
Ph.D. Romero Morales, Cristóbal	cromero@uco.es	University of Córdoba
Ph.D. Romero Salguero, José Raúl	jromero@uco.es	University of Cordoba
Ph.D. Zafra Gómez, Amelia	ameliazg@gmail.com	University of Córdoba

University of Oviedo (Subproject TIN2008-06681-C06-04)

The head of group is *Ph.D. Luciano Sánchez Ramos* of *Department of Computer Science* of the University of Ov

Name	E-mail	Instit
Ph.D. Sánchez Ramos, Luciano (IP)	luciano@ccia.uniovi.es	University of Ov
Ph.D. de la Cal, Enrique Antonio	delacal@uniovi.es	University of Ov
Mr. Junco Navascues, Luis	junco@lsi.uniovi.es	University of Ov
Ph.D. Otero Rodríguez, José	jotero@uniovi.es	University of Ov
Mr. Otero Rodríguez, Adolfo		University of Ov

Mr. Otero Rodríguez, Adolfo	otero@lsi.uniovi.es	University of Ov
Mrs. Suárez Fernández, M. de Rosario	mrsuarez@correo.uniovi.es	University of Ov
Ph.D. Villar Flecha, Jose Ramon	villarjose@uniovi.es	University of Ov

University Ramon Llull (Subproject TIN2008-06681-C06-05)

The head of group is *Ph.D. Ester Bernadó Mansilla* of *Department of Computer Science* of the University Ramor

Name	E-mail	Institutio
Ph.D. Bernadó Mansilla, Ester (IP)	esterb@salleurl.edu	University Ramon Llull
Ph.D. Bacardit Peñarroya, Jaume	jbacardit@salleURL.esu	University Ramon Llull
Mr. Camps Dausà, Joan	joanc@salleURL.edu	University Ramon Llull
Mr. Farguell Matesanz, Enric	efarguell@salle.url.edu	University Ramon Llull
Ph.D. Garrel Guiu, Josep M.	josempg@salleURL.edu	University Ramon Llull
Ph.D. Ho, Tin Kam	tkh@research.bell-labs.com	Bell Laboratories
Mrs. Macià Antolínez , Núria	nmacia@salle.url.edu	University Ramon Llull
Ph.D. Orriols Puig, Albert	aorriols@salleurl.edu	University Ramon Llull
Mr. Rios Boutin, Joaquim	jrios@ati.es	University Pompeu Fal
Mr. Teixidó Navarro, Francesc	fteixido@salle.url.edu	University Ramon Llull

University of Huelva (Subproject TIN2008-06681-C06-06)

The head of group is *Ph.D. Antonio Peregrin Rubio* of *Department of Computer Science* of the University of Huel

Name	E-mail	Ins
Ph.D. Peregrin Rubio, Antonio (IP)	peregrin@diesia.uhu.es	University o
Mr. López Gómez, Ignacio	nacho@dti.uhu.es	University o
Ph.D. Márquez Hernández, Francisco Alfredo	alfredo.marquez@diesia.uhu.es	University o
Ph.D. Moreno Velo, Francisco José	francisco.moreno@dti.uhu.es	University o
Mr. Rodríguez Román, Miguel Ángel	miguel.rodriguez@dti.uhu.es	University o