INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team AxIS

User-Centered Design, Improvement and Analysis of Information Systems

Sophia Antipolis - Méditerranée, Paris - Rocquencourt

THEME COG

Table of contents

1. Team

Research Scientist

Brigitte Trousse [CR1 Inria, Team Leader]
Yves Lechevallier [DR2 Inria, Vice-Team Leader]
Thierry Despeyroux [CR1 Inria, part-time 70%]
Florent Masseglia [CR1 Inria]
Fabrice Rossi [CR1 Inria, on secondment until Sept. 1st, HdR]
Bernard Senach [CR1 Inria]
Anne-Marie Vercoustre [DR2 Inria, part-time 75% until Oct. 19th]
Marie-Aude Aufaure [Assistant Prof., Supélec Gif-sur-Yvette until Sept.1st, HdR]
Marc Csernel [Assistant Prof., Univ. Paris IX Dauphine]

PhD Student

Julie Marlier [Univ. Nice Sophia Antipolis, Intermed Contract, Inria-Cemagref Montpellier, 50% from Nov. 17]
Alice Marascu [Univ. Nice Sophia Antipolis (UNSA-STIC) from June 6.]
Alzennyr Gomes Da Silva [Univ. Paris IX Dauphine]
Elena Smirnova [Univ. Nice Sophia Antipolis (UNSA-STIC), from Sept. 1st]
Chongsheng Zhang [Univ. Nice Sophia Antipolis (UNSA-STIC), from Sept. 15]

Post-Doctoral Fellow

Abdouroihamane Anli [Eiffel project, until Oct. 1st]
Zeina Jrad [Eiffel project, until Sept. 1st]
Nicolas Faure [CALFAT-VIlle project, from May 20 until Dec. 31]
Celine Fiot [University of Montpellier, from Sept. 10]
Wei Wang [ARC Sésur, Rennes from January to July, Sophia-Antipolis from Aug. to Dec. 2008]

Visiting Scientist

Malika Charrad [PhD, ENSI, Tunisi e, April 7–27]
Franciso De Carvalho [Prof.,Federal Univ. of Pernambuco, Brazil, March 28–April 3 and Aug. 28–Sept. 12]

Administrative Assistant

Stéphanie Aubin [TR Inria, part-time]
Anais Cassino [AI Inria, part-time]
Sophie Honnorat [AI Inria, part-time]

Other

Julie Marlier [ACET, from April 21]
Cristina Isai [University of Timisoara, Roumania, from Oct. 1st]
Bashar Saleh [Univ. Nice Sophia-Antipolis, until June 20]
Goverdhan Singh [Indian Inst.of Technology Guwahati, India, from May 11 - Aug. 1st]
Gaurav Panthari [LNM Inst. of Information Technology Rupa ki Nangal, India, from May 11 - Aug. 1st]
Nischal Verma [Indian Institute of Technology Guwahati, India, located in Nimes]
Dipankar Das [Indian Institute oF Technology Guwahati, India, located in Nimes]
Yoan Pitarch [Univ. Montpellier II, France, located in Montpellier]
Hassan Saneifar [Univ. Montpellier II, France, located in Montpellier]
Kelly Patricia Da Silva [Federal Univ. of Pernambuco, Sept.4–26]
Rania Soussi [ENSI University Tunis, Contract STIC Inria-Tunisie, June 2nd - July 30]
Ramzi Mohamed Haddad [ENSI-University Tunis,Tunisia, March 1st–June 30 until Dec; 15]
Vladimir Naumovski [Faculties of Management and Information Technologies-Faculty of Information Technology, Skopje, Macedonia, July 11 - Sept. 12]
Krir Myriam [ENSI-University Tunis, Contract STIC Inria-Tunisie, June 2nd - July 30]
Ben Mustapha Nesrine [ENSI-University Tunis, Contract STIC Inria-Tunisie, Nov. 17 - Dec.17]

2. Overall Objectives

2.1. Overall Objectives

Keywords: *KDD*, *Security*, *Semantic Web*, *Web mining*, *World Wide Web*, *anomaly detection*, *data mining*, *document mining*, *evaluation*, *information retrieval*, *information system*, *intrusion detection*, *knowledge discovery*, *knowledge management*, *ontology management*, *personnalisation data stream mining*, *recommender system*, *semantic Data mining*, *transportation system*, *usage mining*, *user-centered design*.

AxIS is carrying out research in the area of Information and Knowledge Systems (ISs) with a special interest in evolving large ISs such as Web based-information Systems. Our ultimate goal is to improve the overall quality of ISs, to support designers during the design process and to ensure ease of use to end users. We are convinced that to reach this goal, according to the constant evolution of actual and future ISs, it is necessary to anticipate the usage and its analysis and also the maintenance very early in the design process.

To achieve such a research, we have set up in July 2003 a multidisciplinary team that involves people from different computer sciences domains (Artificial Intelligence, Data Mining & Analysis, Software Engineering, Document Management from 2004) and at the end of 2005 from Ergonomics (Human Sciences), all of them focusing on information systems. Our goal is of course to improve efficiency of machine learning and data mining methods but also to improve the quality of results. The originality of AxIS project-team is to adopt a cognitive and inter-disciplinary approach for the whole KDD process and for each step (preprocessing, data mining, interpretation).

To address this challenge, relying on our scientific foundations (see our 2007 activity report,section 3 or http://ralyx.inria.fr/2007/Raweb/axis/), we had a first 4 years steps dedicated to the design of methodological and technical building blocks for IS mining (usage, content and structure). Our next steps started this year identified three applicative objectives:

- Analysing Collective Usage: the need to understand the usage of users or a group of users becomes more and more important in the context of organisations (academic or not) or enterprises for example. Building new KDD tools, able to address the real complexity of mining, that is mining from various sources which become larger and larger seems crucial to us. So is the capacity to adopt pluridisciplinary approaches.
- Offering new kinds of support tools for Information/knowledge retrieval and management: to face the increasing amount of available information/knowledge, tools for retrieving relevant documents have been built. In addition to theses tools, we believe that sophisticated tools are needed for supporting all kinds of users or groups of users in discovering new information/knowledge and more generally in managing their own information/knowledge. These new tools will have to be very flexible KDD based management tools a) for the end-users, they will support the annotation of their selected documents, exhibit and synthesize new information from large documents and selected data collections (clusters of documents for example) or b) for others types of actors such as webmasters or any responsible, they will help them in evaluating and maintaining the IS or in analysing the usage of the IS or a group of users, etc.
- Supporting the evolution of the future ISs (or a collection of XML documents) and of their use: future information systems will evolve faster and faster and will become larger and larger. To support the validation/evaluation of evolving information systems (in terms of content and structure), we believe in the design of powerful maintenance tools based on formal methods from software engineering. Such an evolution might also be due to the evolution of the usage by the end-users. Detecting such behavioural changes will be very useful in order to offer information retrieval support tools able to take into account small usage practice changes. They will be adaptable to the user profile/session and adaptive (i.e with some machine learning capabilities).

Our researches are organised in three topics according to the genericity level of results:

1. Mining complex data and IS data (mainly mining data streams and evolutive data)
2. Mining social networks, support tools for Information retrieval (recommender systems), personalization
3. SHS-STIC approach in evaluating the usage of Web-based information systems and applications.

2.2. Highlights

- AxIS organised with great success the main francophone conference in "Extraction et gestion des Connaissnaces' where there were more than 240 participants and prestigious invited speakers. It was the first conference organised in the new building Kahn at Inria Sophia Antipolis (january 2008). B. Trousse was chair of the organisation of EGC-09 [59], [17].
- Two new ANR contracts (Techlog INTERMED and MDCA MIDAS) started this year.
- Researches on social networks analysis made by F. Rossi in collaboration with N. Villa and Truong Quoc-Dinh [16], [25] were summarized in Le FIGARO.fr[1].
- AxIS was active this year inside the European Network Of Living Labs (ENoLL) [20] in relation with the restructuration of the Living Lab "ICT Usage Lab" of Sophia Antipolis according to the Regional Council and the Paca clusters (SCS, Pegase). Let us remind that our living lab was the first french living lab labelled in 2006 by the ENoLL network.

3. Application Domains

3.1. Panorama

Keywords: *Aeronautics*, *Anomaly Detection*, *Engineering*, *Environment Security*, *Health*, *Security*, *Telecommunications*, *Tourism*, *Transportation systems*, *adaptive service*, *e-business*, *e-democraty adaptive interface*, *information retrieval*, *personalization*, *web design*, *web usage mining*.

The project addresses any applicative field:

- on design, evaluation and improvement of huge hypermedia information systems, for which end-users are of primary concern (cf. section 3.2).
- where a better understanding of service use with data mining techniques and knowledge management could be useful: for example Transportation systems (cf. section 3.3), Tourism (cf. section 3.4), Security and Anomaly Detection (cf. section 3.5).
- where social network analysis could be useful: for example Sustainable development and E-democraty (cf. section 3.6) .

3.2. Evolving Hypermedia Information Systems

Keywords: *Multimedia*, *Telecommunications*, *consistency verification*, *design of Information Systems*, *evaluation of Information Systems*, *ontology*, *personalization*.

[1] See the article (may 28) http://www.lefigaro.fr/sciences/2008/05/24/01008-20080524ARTFIG00514-la-vie-des-paysans-du-moyen-ge-revelee-par-les-maths.php

We currently focus on web-based information systems (Internet, intranet), or parts of such ISs, offering one of the following characteristics:

1. presence or wanted integration of assistance in collaborative information search and personalization (ranking, filtering, addition of links, etc.);
2. a web-based IS containing information about the activities of a group of people, for instance when they are exchanging verbal or textual information.
   - a web-based IS containing information about the activities of a group of people, for example an institute (INRIA), a company, a scientific community, an European network on the Internet or intranet, etc.
   - a web-based IS indexing a wide range of productions (documents, products) from the Web or a company, according to a thematic criteria, eg. search engines (Yahoo, Voila), Internet guides for specific targets (FT Educadoc) or portals (scientific communities).
3. implicit user satisfaction (i.e. the interpretation of the user satisfaction according to the designer point of view) or explicit one, as it is the case for example for business sites, e-learning sites, and also for search engines.

The EPIA RNTL project (2007) is an example of such an evolving hypermedia information system. The Intermed project is an example of the living lab approach.

In summary, our fields of interest are the following:

- semantic specification and checking of an information system,
- usage analysis of an information system (Internet, intranet),
- document mining (XML documents, texts, Web pages),
- re-designing an information system based on usage analysis,
- re-designing an information system based on web mining (usage, content and structure),
- ontology construction and evolution,
- updating an ontology based on web mining (usage, content and structure),
- adaptive recommender systems for supporting information retrieval, collaborative search on the Internet,
- and in general personalization features of an Information System or a service, such as user profiling, personalied interfaces.

Ultimately, it should be noted that other fields (Life Science, Health, Transportation, etc.) could be subject to study since they provide an experimental framework for the validation of our research work in KDD, and in the reuse of experiences managing temporal data; this type of approach may be relevant in applications that are not well solved with methods in Automatics (e.g. nutrition of plants under greenhouses, control in robotics).

3.3. Transportation Systems

Several years ago we acquired experience in the design and evaluation of control rooms for transportation systems (previous work mainly with railway systems and partners such as RATP, SNCF, RTM, etc.). Presently, major evolutions in Intelligent Transportation Systems (ITS) are linked to rapid changes in communication technologies, such as ubiquitous computing, semantic web, contextual design. A strong emphasis is now put on mobility improvements. These improvements concern both the quality of traveler's information systems for trip planning and the quality of embedded services in vehicles to provide enhanced navigation aids with contextualised and personalised information.

The MobiVIP project (2004-2007) has been an opportunity to collaborate with local institutions (Communauté d'Agglomération de Sophia Antipolis - CASA) and companies (VU Log) and apply AxIS' know-how in data and web mining to the field of transportation systems (cf. section 5.13). Our cooperation about car-sharing has also been pursued in 2008 with CASA related to the site Otto&co in the context of the action COLOR Cuscov (cf. section 7.1).

3.4. Tourism

Local tourism authorities have developed Web sites in order to promote tourism and to offer services to citizens. Unfortunately the way information is organised does not necessarily meet Internet users' expectations. Mechanisms are necessary to enhance their understanding of visited sites. Tourism is a highly competitive domain. Thus if only for economical reasons, the quality and the diversity of tourism packages have to be improved, for example by highlighting cultural heritages.

AxIS is involved in the RNTL Eiffel project (cf. section 6.1.1) whose goal is to provide users with an intelligent and multilingual semantic search engine dedicated to the tourism domain. This should allow tourism operators and local territories to highlight their resources; customers could then use a specialised research tool to organize their trip on the basis of contextualised, specialised, organised and filtered information.

Other researches have been carried out using log files from the city of Metz. This city was chosen because their Web site is in constant development and has been awarded several times, notably in 2003, 2004 and 2005 in the context of the Internet City label.

The objective was to extract information about tourists' behaviours from this site's log files and to identify possible benefits in designing or updating a tourism ontology [55].

AxIS is also interested in providing users with transportation information while looking for tourism information such as cultural information, leisure etc.

3.5. Security and Anomaly Detection

Since 2007, we investigated research fields related to security and anomaly detection by means of data mining. Security of information systems is getting more and more difficult because of the streaming aspect of their trace. Data streams have two major characteristics: 1) they are the vital signals of the considered system and their analysis is of great interest and 2) their production rate is so high that actual technology is not able to process them in a satisfying way.

We are currently exploring two aspects of detection in data streams:

- Anomaly detection in data streams. This work aims at discovering records that are unusual compared to the majority of data. It can be used for monitoring devices in sensitive applications, such as nuclear energy production for instance.
- Intrusion detection. In this work, the first step (anomaly detection) is used for network and information system security.

For both aspects, the major issue is to lower the threshold of false alarms. AxIS is involved in two projects related to data streams: MIDAS (an ANR project, section 6.1.3) and SéSur (an ARC of INRIA, section 7.2.1). AxIS is also managing MUTAN (Color project of INRIA, section 7.1.2) a project dedicated to intrusion detection. The MIDAS project allows us to work on real data from Orange (usage of a mobile portal), EDF (fluctuation of energy consumption for each client) and also GPS information from a large of vehicules (real time road traffic information). In projects SéSur and MUTAN, we have applied our methods to client behaviors on Web sites of INRIA Sophia Antipolis and IRISA.

3.6. Sustainable Development and E-democracy

Following the Rio Conference (1992) and the Agenda for the 21st Century, local territories are now directly concerned with the set up of actions for a sustainable development. In this frame, ICT tools have been supposed to be very efficient to re-engage people in the democratic process and to make decision-making more transparent, inclusive and accessible.

The emerging research field of E-democracy (so called Digital Democracy or eParticipation), concerned with the use of communications technologies such as the internet to enhance the democratic processes is now a very active field. Though still in its infancy, a lot of literature is already available (see for instance : http://itc.napier.ac.uk/ITC/publications.asp or http://www.demo-net.org/ for a global view of work in Europe).

Intermed (cf. section 6.1.2) is a sustainable development project concerned with water management. IN this project AxIS' usage analysis tools will help to understand: how the e-democray platform developed within the project is really used, which are the barriers to overcome to increase citizen participation in public decision making, how to reinforce the link between government and the governed population.,

4. Software

4.1. Introduction

http://www-sop.inria.fr/axis/software.html.

AxIS has developed several software or languages: 1) CLF for generating efficient parsers 2) AxISLogMIner for web usage mining (preprocessing and sequential pattern extraction), 3) a clustering toolbox used in our researches in clustering user visits, 4) CBR*Tools for Knowledge management and Broadway*Tools for designing adaptive Web-based recommendation systems. We participated also in the SODAS 2, result of a past European project and in two other software (Ralyx for the exploitation of INRIA activity reports and BibAdmin for the management of a collection of publications). Part of these software will be integrated and proposed next year via the new platform called FOCUS (CPER Télius) (cf. section 5.13.

4.2. CLF -Computer Language Factory

Keywords: *consistency verification*, *natural semantics*, *parser*, *validation*.

Participant: Thierry Despeyroux [correspondant].

CLF is a toolbox designed to ease the development of efficient parsers in Prolog. It currently contains a couple of tools. The first one uses flex to perform lexical analysis and the second is an extension of Prolog DCGs [77], [90], [73] to perform syntactical analysis. It allows right recursion, take advantage of hash-coding of prolog clauses by modern prolog compilers and keep an automatic link to the source code to ease the development of tools as compilers with accurate error messages.

This toolbox has been used to produce a parser for XML. It has also been used to produce the specification formalism SeXML. The generated parsers have been intensively used in our team to parse and analyze XML files, mainly related to our research applied to the Inria annual activity reports.

A complete documentation is available in [82].

4.3. Web Log Preprocessing and Sequential Pattern Extraction

Keywords: *http logs*, *pre-processing*, *web usage mining*.

Participants: Brigitte Trousse [co-correspondant], Yves Lechevallier [co-correspondant], Anli Abdouroihamane, Celine Fiot, Cristina Isai.

AxISLogMiner is a software application that implements our preprocessing methodology [101] for Web Usage Mining (WUM) and our work on sequential pattern extraction with low support.

We used Java to implement our application as this gives several benefits both in terms of added functionality and in terms of implementation simplicity. The application uses Perl modules for the operations carried on the log file such as: log files join, log cleaning, robot requests filtering and session/visit/episode identification. To store the preprocessed log file, in our relational model we used JDBC with Java. The result of this preprocessing is then used in data mining tool to extract, for instance, sequential patterns consisting in sequences of Web pages frequently requested by users. We endowed this software with the ability of recording the keywords employed by users in search engines to find the browsed pages.

This year in the context of Eiffel project and based on AxISlogMiner preprocessing tool, we isolated and redesigned the core (called AWLH) composed of a set of tools for pre-processing web log files. It can extract and structure log files from one or several Web servers, using different input format. The web log files are cleaned as usually before to be used by the datamining tool, as they contains many noisy entries (for example, robots bring a lot of noise in the analysis of user behaviour then it is important in this case to identify robot requests). The data are stored within a database whose model has been improved. The features of the current version of the AWHL are:

- Processing of several log files from several servers (different formats);
- Support of several input formats (CLF, ECLF, IIS, custom, ...);
- Incremental pre-processing;
- Java API to help integration of AWLH in external application.

We also developed a tool based on an open source project called "OpenSympony ClickStream". Using OpenSympony ClickStream we recorded the click actions made by a user in real time. During the capture process we create a table that is used by the AWLH tool to fulfill the tables required for the preprocessing and processing phases of the WUM process.

4.4. Clustering and classification Toolbox

Participants: Marc Csernel, Alzennyr da Silva, Francesco de Carvalho, Yves Lechevallier [co-correspondant], Fabrice Rossi, Brigitte Trousse [co-correspondant].

We developed and maintained a collection of clustering and classification software, written in C++ and/or Java:

- a Java library (Somlib) that provides efficient implementations of several SOM variants [79], [78], [98], [97], [104], especially those that can handle dissimilarity data (available on Inria's Gforge server http://gforge.inria.fr/projects/somlib/, developed by AxIS Rocquencourt and Brieuc Conan-Guez from Université de Metz.
- a functional Multi-Layer Perceptron library, called FNET, that implements in C++ supervised classification of functional data [93], [96], [95], [94] (developed by AxIS Rocquencourt).
- two partitioning clustering methods on the dissimilarity tables issued from a collaboration between AxIS Rocquencourt team and Recife University, Brazil: CDis and CCClust [81]. Both are written in C++ and use the "Symbolic Object Language" (SOL) developed for SODAS.
- two improved and standalone versions of SODAS modules, SCluster and DIVCLUS-T [75] (AxIS Rocquencourt).
- a Java implementation of the 2-3 AHC (developed by AxIS Sophia Antipolis). The software is available as a Java applet which runs the hierarchies visualization toolbox called HCT for Hierarchical Clustering Toolbox (see [76]).

We developed a Web interface for the following methods: SCluster, Div, Cdis, CCClust.The interface is developed in C++ and runs on our Apache internal Web server.

4.5. CBR*Tools

Participant: Brigitte Trousse [correspondant].

CBR*Tools [87] is an object-oriented framework [88], [85] for Case-Based Reasoning which is specified with the UMT notation (Rational Rose) and written in Java. It offers a set of abstract classes to model the main concepts necessary to develop applications integrating case-based reasoning techniques: case, case base, index, measurements of similarity, reasoning control. It also offers a set of concrete classes which implements many traditional methods (closest neighbors indexing, Kd-tree indexing, neuronal approach based indexing, standards similarities measurements). CBR*Tools currently contains more than 240 classes divided in two main categories: the core package for basic functionality and the time package for the specific management of the behavioral situations. The programming of a new application is done by specialization of existing classes, objects aggregation or by using the parameters of the existing classes.

CBR*Tools addresses application fields where the re-use of cases indexed by behavioral situations is required. The CBR*Tools framework was evaluated via the design and the implementation of five applications (Broadway-Web, Educaid, BeCKB, Broadway-Predict, e-behaviour and Be-TRIP).

CBR*Tools is concerned by two current contracts: EPIA and MobiVIP.

CBR*Tools will be soon available for research, teaching and academic purpose under the INRIA Cecill license via the FOCUS platform. The user manual can be downloaded at the URL: http://www-sop.inria.fr/axis/cbrtools/manual/.

4.6. Broadway*Tools

Participant: Brigitte Trousse [correspondant].

Broadway*Tools is a toolbox supporting the creation of adaptive recommendation systems on the Web or in a Internet/intranet information system. The toolbox offers different servers, including a server that computes recommendations based on the observation of the user sessions and on the re-use of user groups' former sessions. A recommender system created with Broadway*tools observes navigations of various users and gather the evaluations and annotations of those users to draw up a list of relevant recommendations (Web documents, keywords, etc).

Different recommender systems have been developed:

- for supporting Web browsing with Broadway-Web,
- for supporting browsing inside a Web-based information system with educaid (France Telecom Lannion - Inria contract), e-behaviour (Color Action, use of the mouse and eye-tracking events) and Be-TRIP (information retrieval and mobility, only specified),
- for supporting query formulation with Be-CBKB (XRCE-Inria contract), etc.

Broadway*Tools has been used in the EPIA contract (2002-2007) and Mobivip project (2004-2007).

4.7. SODAS 2 Software

Participants: Yves Lechevallier [correspondant], Marc Csernel.

The SODAS 2 Software [89] is the result of the European project "ASSO" (Analysis System of Symbolic Official data), that started in January 2001 for 36 months. It supports the analysis of multidimensional complex data (numerical and non numerical) coming from databases mainly in staistical offices and administration using Symbolic Data Analysis [72].

SODAS 2 is an improved version of the SODAS software developed in the previous SODAS project, following users' requests. This new software is more operational and attractive. It proposes innovative methods and demonstrates that the underlying techniques meet the needs of statistical offices.

SODAS allows for the analysis of summarised data, called Symbolic Data. This software is now in the registration process at APP. The latest executive version (version 2.50) of the SODAS 2 software, with its user manual can be downloaded at http://www.info.fundp.ac.be/asso/sodaslink.htm

The main contributions of AxIS to SODAS [49], [52], [57], [56] are:

1. a Symbolic Object Library (SOL [49]) that provides foundation tools, such as data loading and saving, selection, etc .
2. a divisive hierarchical clustering method on complex data tables called DIV [57]
3. a partitioning clustering method on complex data tables called SCLUST [57]
4. a supervised classification tree for symbolic data, called TREE [57]
5. a tool for extracting symbolic objects from databases [52], called DB2SO, jointly developed with EDF

Those contributions have been registered at APP.

4.8. Ralyx

Participant: Anne-Marie Vercoustre [correspondant].

In the context of her involvement with the IST department at Inria, Anne-Marie Vercoustre has been leading the Ralyx project. The goal of the Ralyx project is to publish and exploit dynamically the annual INRIA activity reports. Ralyx is based on the Xyleme system, a native XML database. Thanks to Xyleme, pages and links are no longer static but computed on the fly into different, possibly transversal views. Ralyx is operational since February 2007. The approach works very well but may be limited by the quality of the initial data. This brings us back to one of AxIS' objectives to control and increase the quality of document-based information systems. To access the fine granularity information embedded into the text, more advanced techniques based on natural language processing have been developed.

4.9. BibAdmin

Participant: Brigitte Trousse [correspondant].

"BibAdmin" developed by S. Chelcea (ex-PhD student). BibAdmin is a publication management tool corresponding to a collection of PHP/MySQL scripts for bibliographic (Bibtex) management over the Web. Publications are stored in a MySQL database and can be added/edited/modified via a Web interface. It is specially designed for research teams to easily manage their publications or references and to make their results more visible. Users can build different private/public bibliographies which can be then used to compile LaTeX documents. BibAdmin is made available from the end of 2005 under the GNU GPL license on INRIA's GForge server. A new version has been made in 2008 offering a better robustness of the code. BibAdmin is used by AxIS for its Web server.

5. New Results

5.1. Introduction

This year all obtained original results are related to our previous researches related to mining complex data and IS data which is our first research topic: analysis of evolving usage data (clustering, sequential pattern extraction), mining data streams, document processing and evolutive aspects of ontologies.

Let us note that some works made in 2007 were published in 2008:

- The work on "Mining Solid Itemsets" described in the activity report of 2007 in Section 6.3.7, has been published in 2008 by an international conference [32] and a national conference [47].
- The work on "Mining Generalised Web Data for Discovering Usage Patterns", described in the activity report of 2006 in Section 6.4.3, has been published in 2008 as a chapter of an international book [54].
- The work [16], [19] on Self Organizing Maps on Dissimilarity matrices and the work on feature selection [18] described in 2007 has been published in 2008.

Other work on functional data [46] and on kernel SOM [25], [30] has been published in 2008.

Let us note that after our first researches on mining social networks [16], [25], we pursued this topic inside the new project Intermed (public debates) (cf. section 6.1.2) and the starting thesis of E. Smirnova.

We have also pursued our research on an interdisciplinary SHS-STIC approach in usage mining (Sociology, Ergonomics, Computer Science) inside the CusCov project (cf. section 7.1) and the intermed project (cf. section 6.1.2.

5.2. Adaptive Distances in Clustering Methods

Keywords: *distances table*, *dynamic clustering algorithm*, *unsupervised clustering*.

Participants: Marc Csernel, F.A.T. de Carvalho, Yves Lechevallier.

The adaptive dynamic clustering algorithm [84] and [56] optimizes a criterion based on a fitting measure between clusters and their prototypes, but the distances used to compare clusters and their prototypes change at each iteration. These distances are not determined absolutely and can be different from one cluster to another. The advantage of these adaptive distances is that the clustering algorithm is able to recognize clusters of different shapes and sizes. The main difference between these algorithms lies in the representation step, which has two stages in the adaptive case. The first stage, where the partition and the distances are fixed and the prototypes are updated, is followed by a second one, where the partition and their corresponding prototypes are fixed and the distances are updated.

5.3. Divisive clustering with constraints

Keywords: *distances table*, *unsupervised clustering*.

Participant: Yves Lechevallier.

DIVCLUS-T is a divisive and monothetic hierarchical clustering method which proceeds by optimization of a polythetic criterion [74]. The bipartitional algorithm and the choice of the cluster to be split are based on the minimization of the within-cluster inertia. The complete enumeration of all possible bipartitions is avoided by using the same monothetic approach as Breiman et al. (1984) who proposed, and used, binary questions in a recursive partitional process, CART, in the context of discrimination and regression. In the context of clustering, there are no predictors and no response variable.

We propose an extension of DIVCLUS-T, called C-DIVCLUS-T which is able to take contiguity constraints into account. Because the new criterion defined to include these constraints is a distance-based criterion, C-DIVCLUS-T will be able to deal with complex data. In order to avoid the problem pointed out below concerning the definition of binary questions for complex data, we impose to the variables used in the the binary questions, to be classical. The variables used in the calculation of the distance-based criterion can however have complex descriptions.

The method [23] and [41] proposed has the specificity to be monothetic and its main advantage is then the simple and natural interpretation of the dendrogram and the clusters of the hierarchy. Of course these monothetic descriptions are also constraints which may deteriorate the quality of the divisions.

5.4. Analysis of evolving Web usage data

Keywords: *clustering*, *evolving data*, *web usage mining*.

Participants: Alzennyr Da Silva, Yves Lechevallier, F.A.T. de Carvalho.

In the analysis of Web usage traces, taking into account the time factor has become a necessity since the subjacent distribution of the data can evolve over time. A typical example concerns the Web surfer usage profiles. These models can be dependent on certain temporal factors (day of the week, period of promotions, etc). Our aim is to analyze the evolution of these profiles which can be related to the change in the number of cluster elements or to the displacement of clusters over time. In the article [43], we propose three strategies of clustering based on overlapping sliding windows. During the spring school [42], our work was presented and discussed with computer science students. Our newest results were published in two Springer books [51] [50].

5.5. Block clustering and Web Content Data Mining

Keywords: *Web Usage Mining*, *dynamic clustering algorithm*, *unsupervised clustering*.

Participants: Malika Charrad, Yves Lechevallier.

Our aim is to analyze textual data of a web site. Our approach [40] consists of three steps: Web pages classification, preprocessing of web pages content and block clustering. The first step consists in classifying web site pages into to major categories: auxiliary pages and content pages. In the second step, web pages content is preprocessed in order to select descriptors to represent each page in the web site. As a result, a matrix of web site pages and vectors of descriptors is constructed. In the last step, a simultaneous clustering is applied to rows and columns of this matrix to discover biclusters of pages and descriptors.

5.6. Gradual dependencies and temporal tendencies in sequential patterns

Keywords: *Gradual Rule*, *Sequence Database*, *Sequential Patterns*, *Trends*.

Participants: Celine Fiot, Florent Masseglia.

This work was done in collaboration with A. Laurent and M. Teisseire (LIRMM).

Temporal data can be handled in many ways for discovering specific knowledge. Sequential pattern mining is one of these relevant approaches when dealing with temporally annotated data. It allows discovering frequent sequences embedded in the records. In the access data of a commercial Web site, one may, for instance, discover that "5% of the users request the page register.php 3 times and then request the page help.html". However, symbolic or fuzzy sequential patterns, in their current form, do not allow extracting:

1. Gradual dependencies among objects in a sequential patterns.
2. Temporal tendencies that are typical of sequential data.

In [27], we have proposed GRASP, an algorithm intended to discover gradual trends in sequences. A gradual sequential pattern could be that *considering mail server breakdowns, the more the number of received e-mails is "high" and the more the average size of received e-mails is also "high" at time t, the higher the number of time delivery errors becomes later.* First, the database is converted into a membership degree database, such as for fuzzy sequential pattern mining, using predefined fuzzy sets automatically or from expert knowledge designed. Then, this membership degree database is converted into a time-related variation degree database. This dataset is the one mined for gradual sequential patterns.

This work allows discovering a new kind of knowledge, relying on the relationships between levels (or quantities) of objects in the corresponding patterns (*the more A is high, the more B is low*).However, they do not allow discovering patterns of co-evolution (*i.e.* relationships between fluctuations in these levels or quantities).

Therefore, in [28], [44] we have proposed TED and EVA, a couple of methods designed for evolution pattern mining. such patterns could be for instance that *An increasing number of requests to registration.php during a short period precedes an increasing number of requests to faq.html, after a very short period*. This knowledge would be explicit for the end-user (is the registration-form easy to fill-in?). However, modeling such temporal knowledge requires to handle a very large number of elements both in terms of attributes and records during the mining task. Searching for evolution patterns indeed requires to compare each record of a data sequence to the following ones which leads to a combinatorial space complexity. TED converts a numerical database into a trend database, describing evolution of numerical attribute values, according to time for several objects. These evolutions are represented as trend sequences. Then EVA searches for frequent evolution sequences in this trend sequence dataset.

5.7. Intrusion detection: mining common outliers

Keywords: *Intrusion*, *Outlier*.

Participants: Goverdhan Singh, Celine Fiot, Alice Marascu, Florent Masseglia.

In this work, done in collaboration with P. Poncelet (LIRMM), we have focused on Intrusion Detection Systems (IDS). An IDS aims at monitoring the events occurring in a network and analyzing them for trace of possible incidents, which are violations of computer security policies. Continuous analyses are performed in order to detect and stop possible incidents. There are multiple methods which aim at solving this problem, among them the unsupervised clustering. In this context, the unsupervised clustering allows grouping similar behaviours in clusters and, in a second step, finding among them the clusters corresponding to possible incidents. The idea is that the isolated behaviours are considered as possible incidents. The drawback of this method is that normal atypical behaviours may be considered as suspect. In this context, we have proposed the COD (Common Outlier Detection) method. The idea behind this method consist in a possible attack repetition in many systems. Our algorithm performs successive clustering steps for each site. At each step we check the potentially matching outliers between both sites. The clustering algorithm is agglomerative and depends on the maximum dissimilarity (MD) that has to be respected between two objects. Let us consider that n, the desired number of alarms, is set to 1 and the usage patterns are distributed as illustrated in figure 1. Let us also consider that for these sites the cluster labelled D at step 1 is the only one that corresponds to an intrusion attempt. For step one, MD is initialised with a very low value, so the clusters will be as tight and small as possible. Then we check correspondences between outliers of S_1 and S_2. Let us consider the clustering results on S_1 and S_2 at step one in figure 1. There are four matching of outliers between both sites (A, B, C and D). That would lead to 4 alarms (among which only one is true) which is more than desired by the user and . We thus have to increase the clustering tolerance (*i.e.* increase MD) so that bigger clusters can be built. After a few steps, we will find the clusters of step n in figure 1. The only common outlier is A, which corresponds to the intrusion attempt. Furthermore, this will trigger one alarm, as desired by the user, and there is no need to continue increasing MD until step m.

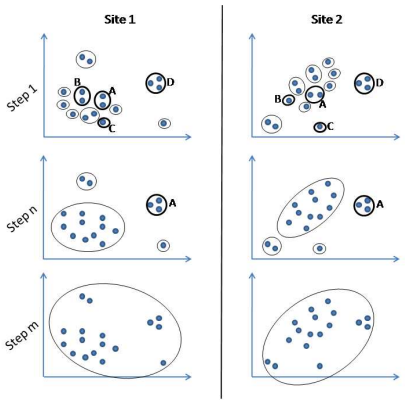

Figure 1. Detection of common outliers in the usage patterns of two Web sites

The network security services of INRIA and our own investigations allow us to confirm the intrusion attempts that have been discovered by our method (including “code injection”, “easter eggs” and “password”).

5.8. Detecting outliers in data streams: a self-adjusting method

Keywords: *Data streams*, *Multi-resolution*, *Outlier*, *Self-adjusting*, *Wavelets*.

Participants: Alice Marascu, Florent Masseglia, Yves Lechevallier.

Outlyingness is a subjective concept relying on the level of isolation of a (set of) record(s). Clustering-based outlier detection is a field that aims at clustering data and detecting outliers depending on their characteristics (small, tight and/or dense clusters might be considered as outliers). In order to separate the common behaviours from the outliers, the existing methods require a parameter such as a percent of small cluster to be considered as outliers or the top -n outliers. However, using a parameter is not always possible in a data stream environment. Starting from this idea, we propose a parameterless outlier detection method. We propose WOD (Wavelet-based Outlier Detection), a parameterless method intending to automatically extract outliers from a dataset. In contrast to previous work, our goal is to find the best division of a distribution and to automatically separate values into two sets corresponding to clusters on the one hand and outliers on the other hand. The tail of the distribution will be found thanks to a wavelet technique and will not depend on a user threshold. Our method will fit any distribution that depends on any characteristic such as distances between objects, objects density or clusters size. The key idea of WOD is to use a wavelet transform to cut down such a distribution. With a prior knowledge on the number of plateaux (we want two plateaux, the first one standing for small groups, or outliers, and the second one standing for big groups, or clusters) we can cut the distribution in a very effective manner. The advantages of WOD are i) to automatically adjust when the distribution shape changes and ii) to give a relevant and accurate detection of outliers with very natural results. Our experiments, performed on real data, confirm this separation feature of WOD compared to well-known outlier detection principles such as the top-k outliers or the percentage filter.

Two papers (PAKDD, SAC) and one poster (EGC) have been accepted in conferences in 2009.

5.9. Online and Adaptive Intrusion Detection in Unlabelled Audit Data Streams

Keywords: *adaptive*, *affinity propagation*, *data streams*, *intrusion detection*.

Participants: Wei Wang, Florent Masseglia.

Current anomaly based IDSs (Intrusion Detection System) have some difficulties for practical use. First, a large amount of precisely labelled data is very difficult to obtain in practical net-work environments. In contrast, many existing anomaly detection approaches need precisely labelled data to train the detection model. Second, data for intrusion detection is typically steaming and the detection models should be frequently updated with new incoming labelled data. However, many existing anomaly detection methods involve off-line learning. Third, many current anomaly detection approaches assume that the data distribution is stationary and the model is static accordingly. In practice, however, data involved in current network environments evolves continuously.

Our adaptive anomaly intrusion detection method addresses these issues through an online and unsupervised clustering algorithm in data streams, under the assumption that normal data is very large while abnormal data is rare in practical detection environments. Our method adaptively detects attacks with following three steps:

1. Building the initial model with Affinity Propagation (AP) [86] and its extension in streaming environments [106].
2. Identifying outliers and updating the model in the streaming environments.
3. Rebuilding the model and identifying attacks. An attack is identified if an outlier is detected again after rebuilding the model.

Online and adaptive intrusion detection is a difficult task because no a priori knowledge (e.g., data distribution as well as labelled information) can be provided to the learning method. Our method can detect intrusions with AP in an online and adaptive fashion through dynamical clustering of audit data streams. A very large real HTTP logs collected in Apache server of INRIA Sophia Antipolis as well as a subset of KDD 1999 bench mark data are used to evaluate our method. Experimental results show that our method is promising in terms of effectiveness and efficiency.

A poster on this work has been accepted in EGC09.

5.10. Validation and evolution of ISs and of ontologies

Keywords: *XML*, *consistency verification*, *natural semantics*, *ontologies*, *validation*.

Participant: Thierry Despeyroux.

To support the design, development and maintenance of complex but coherent sites we involve Software Engineering and Artificial Intelligence techniques as there is strong similarities between structured documents (such as Web sites) and programs.

Following previous experiments [83] we develop a new methodology for XML documents verification [7]. Nowadays, an IS does not only contains documents but it also make references to ontologies, in particular in Semantic Web applications and ontologies engineering is now an important activity

We make a parallel between ontologies and types in programming languages, and we use a small example to show that an ontology can be seen as a type system. When an ontology evolves, studying the impact of this evolution on the semantic annotations that use this ontology can be viewed as a type-checking process. First result has been published in [26].

5.11. Entity Ranking

Keywords: *Entity Ranking*, *INEX*, *Topic difficulty*, *XML Document*, *categories*, *linkrank*, *named entities*.

Participants: Anne-Marie Vercoustre, Vladimir Naumovski.

The goal of *entity ranking* is to retrieve entities as answers to a query. The objective is no longer to tag the names of the entities in documents (in batch mode) but rather to return a list of the relevant entity names, and possibly a page or some description associated with each entity. We have developed a system for Entity Ranking in Wikipedia that addresses two specific tasks: a task where the category of the expected entity answers is provided; and a task where a few (two or three) examples of the expected entity answers are provided. In our approach, candidate pages are ranked by combining three different scores: a linkrank score, a category score, and the initial search engine similarity score. The architecture of our system provides a general framework for evaluating entity ranking which allows for replacing some modules by more advanced modules and evaluate alternatives or different combinations of the score functions [34], [31]. This year we focused on evaluating our approach on taking into account topic difficulty. We show that the knowledge of predicted classes of topic difficulty can be used to further improve the entity ranking performance. To predict the topic difficulty, we generate a classifier that uses features extracted from an INEX topic definition to classify the topic into an experimentally pre-determined class. This knowledge is then utilised to dynamically set the optimal values for the retrieval parameters of our entity ranking system. Our experiments suggest that topic difficulty prediction is a promising approach that could be exploited to improve the effectiveness of entity ranking [70]. The current system has been developed in the context of the INEX (Initiative for the Evaluation of XML Retrieval) track on Entity Ranking [37], [33].

5.12. Comparison of textual documents

Keywords: *document analysis*, *mapping*, *ontologies*.

Participants: Nicolas Faure, Brigitte Trousse.

This research is conducted inside the Calfat Project which benefits from a grant issued by DRAST (Direction de la Recherche et des Affaires Scientifiques et Techniques), component of the Ministère de l'Ecologie, de l'Energie, du Déloppement durable et de l'Aménagement du territoire (French Ministry of Environment), and stands in the general context of text classification and text mining.

This work aims at allowing domain experts to automatically compare textual documents using a simple tool. Those documents are proposals following generic request; the automatic allocation of each proposal to a subtheme of the request for proposal would help domain experts a lot in this otherwise costly task.

The chosen methodology is related to the keyword technique as introduced in (Scot, M., 1997. PC analysis of key words - and key key words. System 25 (2), Elsevier, pp.233-245).

This technique is here used in order to produce a list of characterizing words for each document. Each document is tagged and lemmatised, reduced to a number of nominal lemmatised phrase, then compared to a reference corpus using log-likelihood measure. The results of these comparisons, lexical specificities, are then compared one to another, in order to establish lexical similarities.

Documents are then compared according to the result of this last comparison.

This approach differs from the original technique mainly regarding its application to nominal lemmatised phrases and systematic use of log-likelihood instead of Chi.

A component-based prototype, written in PERL, was developed as a mean to intensively and extensively test the underlying technique. This prototype also benefits from a web-based (PHP) front-end, allowing its use through a network, and comes in two versions, including a simple, "blackbox" version.

Further developments of this approach, includes the use of a terminological resource (ontology, thesaurus, semantic network) to allow classification and to better take into account lexical heterogenity, and various validation tests.

5.13. Towards the FOCUS platform for an inter-disicplinary Approach For Mining Collective Usage

Participants: Bernard Senach, Brigitte Trousse.

In 2006 AxIS has began to set up a new method for web site evaluation, articulating usage mining approach and human factors expertise (cf. our 2006 annual report). The first study during the MobiVIP Project [102] showed that combining Ergonomic and Web usage Mining Approaches was very fruitful and we want to go further in this direction. A rapid analysis of the state of the art as shown that the two INRIA research teams which have their focus on user interface design and evaluation (In-situ and Merlin) as well as other French academic laboratories (LIC/IIHM, LIG/Multicom, IRIT/I3C) and specialized laboratories in usage analysis (Laboratoires des Usages: Marsouin, LUCE, Lutin, Lucsi, LDU of Sophia Antipolis), do not presently use data mining technologies when evaluating web sites. Due to the rate at which usability evaluation methods have to be run after design changes, the international effort is mainly oriented toward automatization of the evaluation processes. Previous attempts have tried to compute web metrics (e.g. Rating Game, WebTango, etc.), to connect log files analysis and task interaction models (for instance, QUIP, KALDI) or to implement human factors expertise in knowledge bases (for instance: Sherlock, Ergoval, Synop, Ergo-conceptor). As full automatization is still often decepting, we believe much more in a cognitive coupling in which web site evaluation relies both on human ability and powerful technologies. The effort will be pursued with the FOCUS platform (see 7.1.6) and the started thesis of E. Smirnova on mining social networks..

6. Contracts and Grants with Industry

6.1. Grants with Industry

6.1.1. EIFFEL: RNTL Project (2006-2009)

Keywords: *Semantic Web*, *ontology*, *personnalisation*, *tourism*, *web usage mining*.

Participants: Abdouroihamane Anli, Marie-Aude Aufaure, Zeina Jrad, Yves Lechevallier [resp.], Bernard Senach, Brigitte Trousse.

The EIFFEL project related to Semantic Web and e-Tourism was labelled in 2006 by the RNTL program and started this year. Industrial partners are Mondeca and Antidot (leader) and academic partners are LIRMM, University of Paris X (Nanterre) and INRIA.

The main goal of the Eiffel project is to provide users with an intelligent and multilingual semantic search engine dedicated to the tourism domain. This solution should allow tourism operators and local territories to highlight their resources; the end users will then use a specialised research tool allowing them to organize their trip on the basis of contextualized, specialised, organised and filtered information. Queries and results will be guided by user profiles extracted from usage analysis. These profiles will facilitate the access to distributed and highly heterogeneous data. In this project, AxIS is in charge of the sub-package entitled SP8 and will define new paradigms dedicated to knowledge searching and visualizing, and will extract and exploit users' models and profiles from web logs.

A second deliverable (june 2008) concerns the preparation of web log data for the pattern discovery task. Our approach consists on using WUM techniques from AxisLogMiner in order to extract information from log files of tourism websites.

As we mentioned previously, in the Eiffel project we acquire our data from many sources. However, data derived from server web logs are the most relevant. A log file is a plain text file arranged in a particular format usually containing the host name (user IP), the date, the requested resource (page), the status (success, failure, error, etc.), the referrer and the user agent (browser identification).

The two main objectives of web log processing are (i) structuring data and (ii) improving the quality of data. In fact, web log files are not structured to be directly usable by datamining tools and they contains many noisy entries (for example, robots bring a lot of noise in the analysis of user behaviour, then it is important in this case to identify robot requests).

6.1.2. INTERMED: ANR Techlog Project (2008-2010)

Participants: Nicolas Faure, Celine Fiot, Cristina Isai, Julie Marlier, Bernard Senach, Elena Smirnova, Brigitte Trousse [resp.].

The Intermed project in response to the ANR TechLog call for proposals has been accepted in 2007. The Intermed kick-off meeting previously planned in December 2007 took place in April 2008. Academic partners are Cemagref (G-EAU and TETIS) LIRMM, CEPEL and industrial partners are SCRIPTAL, SIRENA, Normind, PIKKO. The aim of the InterMed project is to design and implement a set of tools fitting the requirements of users in charge of territory planning.

The goal is to use appropriate technologies to establish a functional link between citizen and local authorities. The technologies we are looking for will be progressively adapted to deal with human factors and constraints of the "field". The proposed experimental approach will rely on several iterations and active participation of people involved in the discussions.

The project follows a participatory design methodology in which involvement of professionals and final users is continuous and several meetings have been conducted with them (Users requirements meetings and so-called Evocation meeting in which solution options are introduced and discussed according to previously identified users' needs).

A first experiment has been conducted with scholars and two experiments are currently settled up. In the first experiment, scholars had to engage themselves in a role play :

- in a first step, they discussed with different professionals to understand the issues of a specific action of territory planning
- then they had to debate about the actions in a role play, half of the groups with Intermed tools. The experiments aims at understanding how a new tool can suit people and what can be its effects on the debate (argument quality and quantity, improvement of decisions, speed of decision).

AxiS participated to several meeting on the fields (Narbonne, Thau, Camargue) and to others Intermed meetings.

The technical participation of the AxIS team within the Intermed project in 2008 is related to the major aspect of usage analysis within a context of collaborative and debating tools.

For this first year experiments, the Intermed Web platform was designed as a composition of several modules using different technologies and databases for storing on one server user profile and document information and on another user actions and interactions.

Therefore the usage analysis on the global platform necessitate to integrate all the data sources within an analysis database with two requirements : a) to process the data generated by the use of the application and b) to understand what final users, politics and other partners would like to know about the collaborative work participants.

Then, the first step of this work was to study the technologies designed for the debating interface and collaborative work. As we need to be able to use every kind of data generated by using the platform (click stream on the standard Web interface, actions calling to Java embedded application), a list of tools was suggested to be integrated to the different modules in order to help tracking user actions and interactions on the Web plateform.

The second step was the design of a preprocessing database in order to unify and store all usage data coming for the different data sources, but also the preliminary data, regarding user profile (location, age, education,...) and debate information (what documents are available, who created them, what should be modified, etc.).The modeling of this database should also take into account the analysis that will be done: categorization and profiling of user and usage, but also social interaction and social network analysis, interaction in the debate.

The implementation of this model is currently in progress and should also integrate some dynamic parts that will be build depending on the debate or collaborative work to be analyzed. In fact each participation to a debate can be structured by some metadata specific to the section of the work. These metadata can be used to analyze more specifically the interactions in some specific context. Therefore we want to integrate them to the analysis database. Imported data are RDF annotations of which the schemata will be extended. Then the RDF schema will be used to dynamically create some part of the database.

We participated to two delivrables [64] and [62].

6.1.3. MIDAS: ANR MDCA Project (2008-2010)

Participants: Yves Lechevallier, Alice Marascu, Florent Masseglia [resp.], Brigitte Trousse, Chongsheng Zhang.

The MIDAS project “Mining Data Streams”, granted by ANR, started on January 2008 and will complete on December 2010. Partners are Ceregmia, EDF, France Telecom R&D, Lirmm, Telecom ParisTech and Inria.

The MIDAS project aims at studying, developing and demonstrating new methods for summarizing data streams. It tackles the following scientific challenges related to the construction of summaries:

- Summaries are built from infinite streams but must have a fixed or low increasing size;
- The construction of summaries must be incremental (done on the fly);
- The amount of CPU used to process each element of the streams must be compatible with the arrival rate of the elements;
- The summaries must cover the whole stream and enable to build summaries of any past part of the history of a stream.

In 2008, we participated in a delivrable written by all partners, on the related works [63]. AxIS was responsible of Chapters 2 and 3, and involved in Chapter 8.

6.2. Others actions with Industry

To identify future collaborations.We organised or participated to different industrial meetings with:

- SAP, Sophia Antipolis related to data mining and data streams (security and environnement problems) in relation with Marascu’s PhD thesis.
- Orange labs (Sophia Antipolis)

- France telecom (Paris) (Inria Rocquencourt, May and Paris Sept.)
- EDF Paris on usage mining and socila network analysis,
- Alcatel-Lucent (Paris) on recommender systems and social networks analysis.
- Accenture for a submission to the Pacalab call (cf. section 7.1.6).

We also had a lot of contacts and meetings this year with SMEs asking for advice and support in user interface design, personalization, recommendations and usage analysis: Perferencement, VU Log, TurnToWine, and from the Incubation Marseille Belle de Mai, GlobeKid, AccésCités and OpenDL;

The Perferencement SME is an actor in web traffic conversion. AxIS was contacted to provide assistance on page design : finding out what can be the optimal location of commercial links and how to avoid bouncing (one visit pages). A common proposal will be submitted for a Pacalab financing in December.

We are involved in the elaboration of a project called Mitra (managed by Paris Tech) for Multimodality Information for sustainable TRAnsport in commuting situation.

7. Other Grants and Activities

7.1. Regional Initiatives

Due to the bi-localization of the team, we are involved with two regions: PACA and Ile-de-France.

7.1.1. Color CusCOV (Inria, UNSA, CASA)

Participants: Julie Marlier, Bernard Senach, Brigitte Trousse.

A previous successfull collaboration during the MobiVIP project (See Delivrable 5.3. [102]) with Agglomeration Community of Antibes Sophia Antipolis (CASA) gave opportunity for further work related to the usage analysis of a Web site dedicated to car-sharing;

CUsCov stands for "Confiance USages et COVoiturage". This project (2008-2009) starts with a Color Action aimed at analysing the role of trust in car-pooling.

Our partners are: LAMIC Laboratory (B. Conein) of University of Nice Sophia Antipolis (UNSA) and the transportation department of a local authority (Agglomeration Community of Antibes Sophia Antipolis) with the participation of M. Relieu (Paris Tech). The goal of the study is to better understand the barrier on car pooling, to judge how ICT can encourage this mobility solution and to find out which "social functions" could encourage to adopt car pooling. Considering that car pooling is a mixture of connecting technologies and of sociology (motivation and reluctance), ICT could contribute to ease the spreading of this mobility solution. One could expect that providing tools for social networking could support crew constitution and community emergence. More precisely, the project aims at understanding the role of trust in "crew" composition and to see if "social functionalities" such as detailed personal information or photography would help to set up a social network and contribute to car pooling spreading.

An internship Julie Marlier located at Inria Sophia Antipolis conducted a survey with users of car pooling combined with a log analysis and a web site analysis. She collected data through interviews and questionnaires and her results draw interesting conclusions, notably that :

- in a professional context, trust is not a main issue
- there are 3 main profiles : ecologist / pragmatic / social, each one with its own motivations
- most of the "social functions" proposed to the participants are rejected because "a car pooling site is not a meeting site" or because they "denatures the spirit of car-pooling"
- using car pooling as a transportation solution is not as easy as one could imagine : after a first contact , realizing the travels is not guarantee due to different timetable, lack of flexibility, specific constraints (scholars) . There are in fact many reasons for which car pooling becomes rapidly not feasible
- a lot has still to be done in car pooling web site to improve mapping between supply and demand and to support effective use of car pooling [69]

7.1.2. Color MUTAN (Inria, LGI2P Nimes, LIRMM)

Participants: Florent Masseglia, Gaurav Panthari, Goverdhan Singh, Dipankar Das, Nischal Verma.

MUTAN (2008, 3 internships) for "MUTualisation des ANomalies pour la détection d'intrusions" is a project on intrusion detection. The main goal of MUTAN is to lower the number of false alarms in intrusion detection based on anomalies. Actually, an anomaly is not always due to malicious behaviour and considering anomalies for intrusion detection is impossible to use because of the high rate of false positives. The main idea of MUTAN is that successful attacks will be repeated from one site to another and will be considered as anomalies. Therefore, if the same anomaly is detected on two (or more) different sites, then it is very likely to be a true intrusion. In 2008, partners of experimentation MUTAN have written 3 research reports [66], [65], [71].

3 papers from this Color have been accepted in conferences in 2009 (EGC, PAKDD).

7.1.3. FOCUS Platform (PACA CPER TELIUS)

In 2007, in a framework agreement between french government and PACA territory (CPER Telius), AxIS proposed the creation of an experimentation platform FOCUS (Finding Out Collective Usage). This proposal was accepted and in 2008 AxIS received funding for buying equipments useful for future experiments in the Region. The FOCUS platform is dedicated to researchers interested in usage analysis, could be web usage, co-conception, user centred pilots in real life (scale 1) or experiments in laboratories contexts. Support offered to computer science community and industrial concerned by usage assessment, prototype validation and benchmarking will help them to improve practice in analyzing usage data.

The FOCUS platform relies on knowledge bases, data sets, data acquisition devices, software toolbox and specific methods for mining usage data. Main components of FOCUS platform are :

Data knowledge level is composed of :

- Data sets and benchmarks. The benchmarks available with the platform are well controlled data set that will be used by researchers to validate their analysis techniques,
- Experience data base : previous experimental results and conclusions in analyzing collective usage, past analysis cases for supporting new ones,
- Knowledge base: EAR methods, best practice, assessment criteria, recommendations, collective activities models.

Tool level is composed of devices and software resources

- Data acquisition tools : audio and video devices, advances tape recorder to synchronize heterogeneous data, eye-trackers, sounds to catch web usage,
- KDD toolbox : clustering and data stream mining algorithms, XML document checking,
- Experience reuse and IR support : object oriented framework for case based reasoning.

Multi-disciplinary approach: Stakeholders in FOCUS platform are coming from different disciplines: computer and social sciences (cognitive ergonomics, economy, sociology) which provide a very interesting confrontation of points of view and enriches usage analysis, for instance through multidisciplinary assessment criteria.

7.1.4. PACA Clusters or "Pôles de compétitivité"

- Cluster SCS - Solutions Communicantes Sécurisées: AxIS (B. Trousse and B. Senach) was involved in the preparation of the project Clic&Go. The goal of this project is to provide tools for proximity e-commerce. New technologies are used not to sell products on-line but to support new relationships between customers and shopkeepers. A better information will be given to the clients (availability of products with well-defined attributes - size, colors, etc.) and merchants will be able to develop loyalty through technologies. Usage mining will be used to improve user profiling and push suggestions according to similarity of buyers and profils. In Clic&Go project, academic partners are LSIS (UMR CNRS 6168), INRIA (AxIS) and industrial partners are STID, COMLINKS and AGEVIA. The

project labelled by “Pôle de competitivité SCS” in June last year did not received a grant from the region. After few meetings in 2008 the consortium suspended its activities.

- Pegase Cluster with the CEFH Platform (Centre d’Etudes des Facteurs Humains): Areva TA, leader of the CEFH project contacted us for a collaboration in the set up of an experimentation platform dedicated to human factors studies in risky domains (nuclear power plant, aeronautics industry, process control. After a first contact in April, B. Senach participated in the first meeting in Aix en Provence, a common proposal was submitted to the call "Plateforme d’innovation" (DGE and CDC) The CEFH platform project is supported by the PEGASE cluster.

7.1.5. Living Lab “ICT Usage Lab”

Participants: Bernard Senach, Brigitte Trousse.

AxIS is involved with PACA partners (Orange labs, CSTB, UNSA, ...) in the set up of the Living lab.”ICT Usage Lab” (Sophia-Antipolis) who was labelled during the first wave by the European Network of Living Labs (ENoLL) in 2006 and was the first French living lab. We also were involved in a submission (called INFRALLABS) with almost 15 other Living Labs to the P7 call “Research Infrastructure” [20]. A booklet and a flyer are under construction.

7.1.6. Other initiatives

PacaLabs Initiative: AxiS took part in the presentation of the Pacalab financing structure. We also organised a meeting with the FING [2] to envision future collaborations.

Two projects were submitted to Pacalab:

- PERFECT[3] with Accenture labs, ADEME and CASA: the goal of this project is to provide an information system aiding car drivers to reduce their gasoline consumption and to adopt driving behaviors contributing to sustainable development. Partners of the project are Accenture, ADEME, and CASA
- Hotel-Ref-PACA : the goal of the project is to identify from an experimental approach what can be the optimal location of commercial links and how to avoid bouncing (one visit pages). Partner are : Perferencement company and a local territory.

AxIS (B. Trousse and B. Senach) are involved in the preparation of a new project (AccésCité) aiming at providing guidance for thematic urban travelling. The device will be used, in particular by disabled people and usage mining tools will allow to improve information content delivered to users.

7.2. National Initiatives

AxIS is involved in one ARC from INRIA and several national working groups.

7.2.1. ARC SéSur

Participants: Wei Wang, Florent Masseglia, Yves Lechevallier, Brigitte Trousse, Alice Marascu.

SéSur (2007-2008) for “Sécurité et Surveillance dans les flots de données” is an ARC which involves Dream (IRISA), LGI2P (Ecole des Mines d’Alès) and LIRMM (Montpellier). The goal of SéSur is to propose solutions for the security, monitoring and diagnosis of data streams. Data streams have two major characteristics: 1) they are the vital signals of the considered system and their analysis is of great interest and 2) their production rate is so high that actual technology is not able to process them in a satisfying way. We are mainly interested in monitoring the systems that produce data streams. In 2008, partners of SéSur have published research papers in national and international conferences [91], [99] and 4 research reports [92], [100], [80] and [35], [67]. Three papers (EGC, PAKDD, SAC) and one poster (EGC) from this ARC have been accepted in conferences in 2009.

[2] FING:
[3] PERFECT: PERFormance ECologique pour Trajets automobiles

7.2.2. EGC Association

AxIS members participated actively this year to the Working Group "Fouille de données complexes" created by D.A Zighed in June 2003 in the context of the EGC association:

- B. Trousse co-chaired with F. Guillet the 8th francophon conference in "Extraction et Gestion des Connaissances" [59] and was responsible of the conference organisation at Inria Sophia Antipolis. More than 240 persons attended the conference (january 2008). [59], [17]
- M.-A; Aufaure, Y. Lechevallier, F. Masseglia, B. Trousse (23 January, 2007) are member of the workshop of "Fouille de données complexes" at EGC08.

7.2.3. GDR-I3

AxIS is concerned by three working groups of the GDR-PRC~I3 National Research Group "Information - Interaction - Intelligence" of CNRS: working Group 3.4 (GT) on Data Mining, GRACQ, Working Group 3.7 on "Sécurité des Systèmes d'Information".

7.2.4. Other Collaborations

- Telecom Paris: we collaborated with Georges Hébrail (ENST) via the MIDAS project and the BiLab laboratory [63].
- Telecom Bretagne: W. Wang opursuued his collaboration with Sylvain Gombault [36].
- LIRMM: F. Masseglia and C. Fiot [27], [28], [44] and also in the context of the MIDAS and Intermed projects: F. masséglia, B. Trousse, B. Senach and C. Fiot, J. Marlier
- CEMAGREF Montpellier: via the ANR intermed coordonnated by Nils Ferrand. A thesis (Julie Marlier) is started in November, co-supervised by B. Conein (UNSA), N. Ferrand (CEMAGREF), B. Senach and B. Trousse..
- Loria (S. Lamirel) and INIST (Claire Francois) in the context of the Quaero project;
- Paris Descartes: Marc Csernel collaborated regularly with Francois Patte on aspects concerning the Sanskrit.
- CNAM and Loria (Cortex Team): contacts have been established with research teams in human and social sciences.
- University of Bordeaux 1 and 2 (MAP laboratory): Y. Lechevallier collaborated with M. Chavent [41], [23]
- University of Toulouse Le Mirail (GRIMM-SMASH team): F. Rossi works with N. Villa Self Organizing Map for dissimilarity matrices (cf Section 4.4).
- University of Metz (LITA EA3097): F. Rossi works with Brieuc Conan-Guez on the Self Organizing Map for dissimilarity matrices (see Section 4.4) [53], [19]

7.3. European Initiatives

7.3.1. ENoLL: European Network of Living Labs

AxIS is very active in the ENoLL network and participated in the founding meetings.

As a member of ENoLL, AxIS participated with other partners to the INFRALLABS proposal for FP7 call.

B. Senach and B. Trousse made a Contribution to Living Labs Research in ECOSPACE Newsletter No 5 - Special Issue dedicated to Living Labs [20]

7.3.2. Other Collaborations

- Germany, Clausthal University of Technology, Department of Informatics (Prof. Barbara Hammer & Alexander Hasenfuss) and IPK Gatersleben, Pattern Recognition Group (Dr. Marc Strickert): F. Rossi [30]
- Germany: AxIS participated to the project "Core Technology Cluster" of the AII program "QUAERO" in the multimedia domain (ontology construction, personalization)
- Italy, University of Napoli II (Prof. R. Verde) [56],
- Belgium, Université Catholique de Louvain, DICE Laboratory (Prof. Michel Verleysen, Prof. Vincent Wertz, Dr. Damien Francois & Catherine Krier): F. Rossi [18], [45]

7.4. International Initiatives

7.4.1. Brazil

We continue our collaboration on clustering and web usage mining with F.A.T. De Carvalho from Federal University of Pernambuco (Recife) and his team.

- A scientific project submitted by Francisco De Carvalho and Yves Lechevallier has been accepted by FACEPE and INRIA. The project started from 04/2006 and ends on 03/2008. Researches and students are concerned by this project from AxIS and CIn-UFPE side. It aims at developing methods of clustering analysis and web usage mining tools.
- Francisco de Carvalho visited AxIS project. During their stays, in collaboration with Yves Lechevallier, they participated to the design of dynamic clustering models based on adaptative distances and they finalized the elaboration of dynamic clustering models based on adaptative distances suitable to symbolic interval data. A complete paper has been submitted and accepted to the "Pattern Recognition" journal.
- In collaboration with Yves Lechevallier, Alzennyr Da Silva and Fabrice Rossi, Francisco de Carvalho has participated to the elaboration of an approach concerning the construction of summaries via clustering methods of data which evolve overtime. An application of this approach has been done on data from web usage which evolves on the time.
- Marc Csernel and Yves Lechevallier visited CIn-UFPE during the first week of mars. Marc Csernel worked also with Kelly Silva on her Master Thesis on the adaptation of Fuzzy clustering on symbolic data, specially in presence of rules as background knowledge.
- During "L'année de la France au Brésil" the workshop on data mining has been labelled and is expected to happen in Recife during the first week of May 2009.

7.4.2. Canada

Y. Lechevallier pursued his collaboration with A. Ciampi (Univ of McGill, Montréal).

7.4.3. China

Marie-Aude Aufaure collaborates with Yanwu Yang, Institute of Automation, Chinese Academy of Science, Beijing, on user modelling for the semantic web [105] and Yves Lechevallier collaborates with Hueiwen Wang, BUAA, Beijing on clustering methods.

7.4.4. Morocco

The project (duration of 4 years) with Casablanca University and ENSAM (Meknes) about the WRUM proposal (Web Redesign based on Usage Mining) has been accepted in may 2008 and received a financing from Morocco Telecom. In this project, AxiS will have to use multi-sites log analysis to redesign Web sites. The project hasn't started this year.

7.4.5. Romania

We maintained our contacts with the Computer Science department of the West University of Timisoara (Prof Viorel Negru), in particular via the SYNASC conference every year. B. Trousse discussed with V. Negru about future collaborations at ICT conference (Lyon) in november 2008.

7.4.6. Tunisia

Marie-Aude Aufaure and Yves Lechevallier are involved in co-supervision of 4 masters in the context of the STIC Tunisie and/or 2 thesis (Riadi Lab, ENSI Tunis) and participated as invited speakers [22], [24] in the workshop Ecol'IA08 "Apprentissage et fouille de données : de la théorie à la pratique" at Hammamet (Tunisie) during march 2008. These masters and thesis subjects are about web mining (usage, content and structure, using different methods) and ontology construction from heterogeneous sources [48], [68], [39], [29], [15].

8. Dissemination

8.1. Promotion of the Scientific Community

8.1.1. Journals and Books

AxIS is involved in the management and the edition of 3 journals and 2 books:

- B. Trousse was co-editor of the second special issue of the RNTI journal on "Fouille de donnée complexes": B. Trousse
- B. Trousse is a member of the RSTI scientific committee related to the "ISI, L'OBJET, RIA, TSI" journals (Hermes publisher).
- The MODULAD electronic journal, http://www.modulad.fr/): Y. Lechevallier is one of the four editors. F. Rossi is a member of the editorial board and S. Aubin is the webmaster of the web site.
- F. Masseglia is co-editor of a special issue of "Transactions on Multimedia" [21]. based mainly on selected papers of the successful workshop organised in conjunction with the 2006 ACM KDD conference.

AxIS members belongs to editorial boards of four international journals, six national journals (or some of their special issues):

- the Co-Design Journal (Editor: S. Scrivener, Coventry University, UK - Publisher: Swets & Zeitlinger): B. Trousse
- the RSTI RIA journal ("Revue d'Intelligence Artificielle") (Hermes publisher; editor-in-chief: M. Pomerol): B. Trousse.
- RIA Special Issue "Intelligence Artificielle et Web Intelligence" (Publisher, Hermes-Lavoisier) Editions, editors-in-chief: Y. Demazeau and L. Vercouter): B. Trousse
- the Journal of Symbolic Data Analysis (JSDA) (Editor: E. Diday, electronic journal: Y. Lechevallier, F. Rossi
- Neurocomputing: F. Rossi (http://www.elsevier.com/locate/issn/09252312)
- European Journal of GIS and Spatial Analysis ("Revue Internationale de Géomatique") http://geo.e-revues.com/: M.-A. Aufaure

AxIS members were reviewers for 12 international journals:

- ACM Computing Surveys: B. Trousse
- Behaviour & Information technology (BIT) journal: B. Trousse
- TKDE Transactions on Knowledge and Data Engineering: F. Masseglia, B. Trousse (http://www.computer.org/tkde/)
- DKE International Journal on Data Knowledge and Engineering: C. Fiot, F. Masseglia
- Journal of Intelligent Information Systems (JIIS): C. Fiot
- Pattern Recognition (PR): C. Fiot
- Econometric Reviews: F. Rossi
- IEEE Transactions on Neural Networks: F. Rossi
- IEEE Transactions on Pattern Analysis and Machine Intelligence: F. Rossi
- Computational Statistics and Data Analysis: F. Rossi
- Neural Processing Letters: F. Rossi
- Neurocomputing: F. Rossi

8.1.2. National Conferences/Workshops

AxIS members as co-chairs organised or co-organised three new EGC workshops: http://www-sop.inria.fr/axis/egc08/ateliers-cours.html)

- Modélisation utilisateur et personnalisation d'interfaces Web [60]: Z. Jrad, A. Abdouroihalane, M.-A Aufaure
- Analyse des usages et Mobilité urbaine [61]: B. Senach, B. Trousse,
- Mesures de similarité sémantique [58]: M.-A. Aufaure (co-organiser with O. Boussaid (ERIC, Lyon) and P. Kuntz (LINA, Nantes)

Several AxIS members were involved in national conferences/worhshops as members of Program Committee.

- EGC 2008: Sophia-Antipolis, France (Jan.): M.-A. Aufaure, Y. Lechevallier, F. Masseglia, B. Trousse
- Ateliers EGC 2007: Sophia-Antipolis, France, 29 January 2008
  - Fouille de Données Complexes: M.-A. Aufaure, B. Trousse, F. Masseglia, Y. Lechevallier
  - Modélisation de Connaissances: B. Trousse
  - Analyse des usages et Mobilité urbaine: Y. Lechevallier
  - Modélisation utilisateur et personnalisation d'interfaces Web: Y. Lechevallier, B. Trousse
  - Fouille de données temporelles: F. Masseglia.
- BDA 2008: Guilherand-Granges, France (Oct. 21-24): F. Masseglia
- Atelier RàPC 2008, Nancy, march 31 - april 2nd: B. Trousse

8.1.3. International Conferences/Workshops

B. Trousse was chair of the track "Cognitive Modelling and Interaction" at the ECAI 2008, the 18th biennal European Conference on Artificial (Patras, Greece, july 21 - 25.)

M.-A. Aufaure was co-chair of the workshop "Semantic Web meets geospatial applications - AGILE 2008, Girona, Spain, May 5th, 2008.

Several AxIS permanents were involved in international conferences/worhshops as members of Program Committee.:

- CSCWD 2008: Xi'An, China (April 16-18, 2008): B. Trousse
- ECCBR 2008, Trier, germany, 1-4 september: B. Trousse
- SYNASC 2008, Timisoara, september, 2008: B. Trousse
- IEEE ICDM 2008: Pisa, Italy, December 15 thru 19, 2008: F. Masseglia.
- PIKM 2009: Napa Valley, California, USA, October 26-30 2008, held in conjunction with CIKM'08: F. Masseglia.
- Data Mining for Business Applications Workshop at KDD08, Las vegas, USA, August 24: B. Trousse
- DocEng 2008: A.-M. Vercoustre
- ESAIR'08 workshop (part of ECIR08): A.-M. Vercoustre
- ACM SIGIR 2008 Workshop on Focused Retrieval: QA, XML-IR and Passage Retrieval: A.-M. Vercoustre
- IASTED EuroIMSA 2008: Innsbrck, Austria (March 17-19,2008): A.-M. Vercoustre
- WDSA 2008: Beijing, China (October 27-29): Y. Lechevallier
- 3rd workshop on Hybrid Artificial Intelligence Systems HAIS 2008: F. Rossi
- ICANN 2008: F. Rossi

AxIS members were also additonnal reviewer for

- IEEE ICDM 2008: R. Rossi, C. Fiot
- IEEE/ACM International Conference on Soft Computing as Transdisciplinary Science and Technology: C. Fiot
- IPMU'08 (International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems): C. Fiot

8.1.4. Organization of Conferences /Workshops

- AxIS co-organised EGC 2008, the Tenth conference on "Extraction et Gestion des Connaissances" at Sophia Antipolis. B. Trousse was chair of EGC08 organisation committee.

  http://www-sop.inria.fr/axis/egc08/
- A.-M. Vercoustre was member of the advisory board of the track on Entity Ranking for Inex 2008 (Dec. 12-18, Dagstuhl).
- B. Trousse is Publicity Chair of KDD-09 the 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, which will be located in Paris , June 28th July 1st 2009.
- B. Senach is member of the local organization committee of the EPIQUE 2009 Colloquium which will be held in Sophia Antipolis in november.

8.1.5. Activities of General Interest

- T. Despeyroux is involved (30 %) as president of AGOS (Inria Works Council), a permanent member of the "Commission technique paritaire (CTP)" and a member of the Inria Board of Directors (Conseil d'Administration) as a scientific staff representative.
- B. Senach is member of an internal "think tank" working group (CUMIR) which has to envision the future users needs within INRIA to plan the required evolution of technological resources. During this year he was leader of an Inria Sophia Antipolis working group : "Formation et Communication" which has to analyze the information flow between employees (researchers and administration) and the technical support department (SEMIR) and to set up a formalized agreement between the different stakeholders.

- B. Trousse is the corresponding person with Servane Crave (Orange labs) of the Living lab “ICT usage Lab” inside the ENoLL network. The ICT Usage Lab was the first french lab labelled by the ENoLL network which is based on the past work of the “Laboratoire des Usages des NTIC” (2002-2006) of Sophia Antipolis. This year in collaboration with Orange labs, CSTB, UNSA and Paris Tech we redefined the main objectives of our lab.
- B. Trousse is member of the technical committe on Computer Supported Work in Design in the context of the IEEE SMCS (Systems, Man & Cybernetics Society (http://www.ieeesmc.org/)
- A.-M. Vercoustre is involved (25%) in the Department for Scientific Information and Communication (DISC), working on Inria policy and tools for scientific publications, in particular the development of the Open Archive HAL, in cooperation with CNRS. She is a member of the COST (Comité scientifique et technique du Comité stratégique) for the extension of HAL to become the French National Open Archive. As part of her DISC involvement, A.-M. Vercoustre is also leading the Ralyx project for exploiting the INRIA Activity Report (cf. section 4.8).

8.2. Formation

8.2.1. University Teaching

AxIS is an ”associated team” for the STIC Doctoral school at the University of Nice-Sophia Antipolis (UNSA) and AxIS team members are teaching in various universities:

- “Master PMLT” (resp. Mr Kounalis) at UNSA Sophia Antipolis: Tutorial (12h) on *Data Mining and Web Mining*: C. Fiot, F. Masseglia, B. Trousse (resp.).
- “Master IFI” (resp. Bruno Martin) at Polytech’Sophia: Lecture (6h) on *Data Mining: Sequences, Streams and Security*: F. Masseglia.
- “Application web et Web services” (resp. Mr Quafafou) at ESIL Luminy: Tutorial (18h) on "Web Design: ergonomics and usage analysis". B. Senach, (resp.).
- Licence professionnelle franco-italienne: Statistiques et Traitement Informatique de Données (STID) (resp. J. Lemaire) at UNSA, Menton: Supervision of a student project (60h by students, 8 students, 30h supervised) on *Mining HTTP Logs From Inria’s Web Sites*: B. Senach, B. Trousse (co-resp.).
- Master 2 Recherche Systèmes Intelligents (resp: S. Pinson) of the University Paris IX-Dauphine: Tutorial (12h) on *“Du data mining au knowledge mining”*: Y. Lechevallier.
- Master 2 Pro Mathématiques appliqués et sciences économiques (resp: P. Cazes) of the University Paris IX-Dauphine: Tutorial (15h) on *“Méthodes de classification”*: Y. Lechevallier.
- Master 2 Pro Ingéierie de la Statistique (resp: G. Saporta) of CNAM (12h) on *Méthodes neuronales*: Y. Lechevallier.
- ENSAE (“Ecole Nationale de la Statistique et de l’Administration Economique”): Tutorial (18h) on *“Data Mining”*: Y. Lechevallier.
- Master 2 recherche Informatique, Paris XI: Tutorial (3h) on *Ontology construction*: M.-A. Aufaure.
- Master 1 Pro Ingénierie Mathématique pour les Sciences du Vivant (resp: B. Le Roux et M. Kratz) of University of Paris V, introduction to artificial neural networks (15h): F. Rossi.

F. Masseglia and B. Trousse are reviewers each year for Master thesis in KDD (resp: D. Zighed) at the University of Lyon 2 (France).

8.2.2. H.D.R and Ph.D. Thesis

Ph.D. in progress:

1. **A. Marascu**, (start: October 2005), Extraction de Motifs Séquentiels dans les Data Streams, University of Nice-Sophia Antipolis (director: Yves Lechevallier, with the participation of F. Masseglia).
2. **A. Da Silva**, (start: October 2005), Modélisation de données agrégées ou complexes par l'approche symbolique, application au Web Usage Mining, University of Paris IX Dauphine (directors: Edwin Diday and Yves Lechevallier).
3. **E. Smirnova** (start: september 2008), "Mining social networks", (directors: B. Trousse, M. Rueher)
4. **C. Zhang** (start: october 2008), "Mining data streams: clustering and pattern extraction", University of Nice-Sophia Antipolis, (director: F. Masseglia)

Others Ph-D in progress supervised by some AxIS members::

Y. Lechevallier is co-superviser with G. Saporta for M. Charrad's thesis (start: end of 2005), CNAM. and University La Manouba (Tunisia).

B. Trousse co-supervises with A. marzark (UnIv. Of Casablanca) of the thesis of **H. Behja**, (start: 2003), "Gestion de points de vues multiples dans l'analyse d'un observatoire sur le Web",

B. Senach and B. Trousse with N. Ferrand (Cemagref Montpellier) are member of the thesis committee of **Julie Marlier** supervised by B. Conein (UNSA).

F. Rossi is a member of the thesis committee of **C. Krier** (start: October 2005) on "Analyse de données de grande dimension en particulier en spectométrie, Catholic University of Louvain, Belgium (director: Michel Verleysen).

AxIS researchers were members of H.D.R or Ph.D. committees in 2007:

- **E Boutin**, H.D.R., Recherche d'information sur internet au prisme de le théorie des facettes, University of Nice-Toulon Var (TIC), Toulon, november 13 : B. Trousse
- **M. Damez-Fontaine 2008**, PH.D, Apprentissage artificiel pour l'apprentissage humain: de la récolte de traces à la modélisation utilisateur, LIP6, Paris, september 18: B. Trousse;
- **S. Castagnos**, PH.D, Modélisation de comportements et apprentissage stochastique non supervisé de stratégies d'interactions sociales au sein de systèmes temps réel de recherche et d'accès à l'information, october, University of Nancy: B. Trousse
- **S. Marcellin**, Ph.D, Arbres de décision en situation d'asymétrie, September, University of Lyon II: Y. Lechevallier
- **N. Paul**, Ph.D, Une nouvelle approche pour l'estimation des mélanges de lois. Application à la radio cognitive, May, CNAM: Y. Lechevallier
- **S. Marret**, Ph.D, Identification de cibles sur critères cinématiques, July, University of Bordeaux I: Y. Lechevallier
- **L. Karoui**, Ph.D, Extraction contextuelle d'ontologie par fouille de données, November, University of Paris-Sud: Y. Lechevallier

8.2.3. Internships

We welcomed 13 students this year:

fou at Sophia Antipolis:

- **C. Isai** (supervisor B. Trousse), Master 2, University of Timisoara, Roumania, software components of the FOCUS platform (oct 2008-Feb 2009)
- **J. Marlier** (supervisors B. Senach , B. Trousse) post master Research, ENSAM Paris. "Analyse des usages d'un site de covoiturage : confiance, constitution des équipages et développement dun réseau social".
- **G. Singh** (supervisor F. Masseglia), Master 1, Indian Institute of Technology Guwahati, India, "Collaborative Mining of Outliers for Intrusion Detection" [66]
- **G. Panthari** (supervisor F. Masseglia), Master 1, LNM Institute of Information Technology Rupa ki Nangal, India, "Mining Sequential Patterns at Optimal Time Granularity"[65]

five at Rocquencourt :

- **R. Soussi** [15], [48], **K. Myriam** and **R. Mohamed** [68], ENSI, Contract STIC Inria-Tunisie
- **V. Naumovski** and **Ben M. Nesrine** [38], [39]

four students at Nimes (P. Poncelet) in the context of MUTAN Color and at Montpellier (LIRMM M. Teisseire) in the context of the ARC SéSUR:):

1. **N. Verma** (supervisors P. Poncelet and F. Masseglia, location Nimes), Master 1, Indian Institute of Technology Guwahati, India, "Intrusion Detection: how to preserve privacy in a collaborative environment ?" [103].
2. **D. Das** (supervisors P. Poncelet and F. Masseglia, ARC Sésur, location Nimes), Master 1, Indian Institute of Technology Guwahati, India, "Data steams: Some improvements on titled time windows" [80].
3. **Y. Pitarch** and **H.Saneifar** ARC SéSUR, located in Montpellier [92], [100].

8.3. Participation to Workshops, Conferences, Seminars, Invitations

Furthermore we attended the following conferences or workshops:

- Conference ICE 2008 (Lisbon), June 23-25: B. Trousse, B. Senach
- Conference ICT 2008 (Lyon), November 25-27: B. Trousse, B. Senach
- World usability day 2008, Sophia Antipolis, November 13, B. Trousse
- Seminars Jean-Pierre Fénelon on Data Analysis: Y. Lechevallier
- Bilab Seminars: Y. Lechevallier, A. Da Silva
- INEX'08 Workshop (Initiative for the Evaluation of XML Retrieval), Schloss Dagstuhl, Germany, December 17-19: A.-M. Vercoustre

Y. Lechevallier and M.-A. Aufaure made each an invited talk at Ecol'IA 2008 (March 20-22, Tunisia) organised by Khaled Ghedira (LI3 / ENSI, Tunis)

F. Rossi made an invited talk on *Introduction à l'apprentissage statistique* at the first *Réseau africain de statistique mathématique et ses applications (RASMA)* and also on his work with N. Villa and Quoc-Ding Truong on mining a medieval social network [25].

F. Rossi presented his work at various workshops: a) seminars SAMOS, University of aris 1 (march), b) seminar *probabilité et statistiques*, University of Lyon 1 (february) and c) colloquium Informatica, University of Groningen (Nederlands, february).

B. Senach made an invited talk at the World usability day 2008 at Sophia Antipolis [69].

9. Bibliography

Major publications by the team in recent years

[1] E. GUICHARD (editor). *Mesures de l'internet*, ouvrage collectif suite au Colloque Mesures de l'internet, Nice, France, 12-14 Mai, 2003, Les Canadiens en Europe, 2004.

[2] P. BERTRAND, M. F. JANOWITZ. *The k-weak hierarchical representations: an extension of the indexed closed weak hierarchies*, in "Discrete Applied Mathemathics", vol. 127, n^{o} 2, April 2003, p. 199–220.

[3] M. CHAVENT, F. DE CARVALHO, Y. LECHEVALLIER, R. VERDE. *New clustering methods for interval data*, in "Computational Statistics", vol. 21, n^{o} 23, 2006, p. 211-230.

[4] S. CHELCEA, P. BERTRAND, B. TROUSSE. *Un Nouvel Algorithme de Classification Ascendante 2-3 Hiérarchique*, in "Actes de 14ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle (RFIA 2004), Centre de Congrès Pierre BAUDIS, Toulouse, France", vol. 3, 28-30 Janvier 2004, p. 1471-1480, http://www.laas.fr/rfia2004/actes/ARTICLES/388.pdf.

[5] B. CONAN-GUEZ, F. ROSSI, A. EL GOLLI. *Fast Algorithm and Implementation of Dissimilarity Self-Organizing Maps*, in "Neural Networks", vol. 19, n^{o} 6-7, August 2006, p. 855–863, http://dx.doi.org/10.1016/j.neunet.2006.05.002.

[6] A. DA SILVA, Y. LECHEVALLIER, F. DE CARVALHO, B. TROUSSE. *Mining Web Usage Data for Discovering Navigation Clusters*, in "11th IEEE Symposium on Computers and Communications (ISCC'06), Pula-Cagliari, Italy", IEEE Computer Society, 26-29 June 2006, p. 910-915, http://doi.ieeecomputersociety.org/10.1109/ISCC.2006.102.

[7] T. DESPEYROUX. *Practical Semantic Analysis of Web Sites and Documents*, in "The 13th World Wide Web Conference, WWW2004, New York City, USA", 17-22 May 2004, http://www-sop.inria.fr/axis/papers/04www/despeyroux-www2004.pdf.

[8] A. EL GOLLI, B. CONAN-GUEZ, F. ROSSI, D. TANASA, B. TROUSSE, Y. LECHEVALLIER. *Une application des cartes topologiques auto-organisatrices à l'analyse des fichiers Logs*, in "Actes des onzièmes journées de la Société Francophone de Classification, Bordeaux, France", Septembre 2004, p. 181–184.

[9] G. HÉBRAIL, Y. LECHEVALLIER. *Data mining et analyse des données*, in "Analyse des données", Hermes, June 2003, p. 340-360.

[10] A. MARASCU, F. MASSEGLIA. *Mining Sequential Patterns from Data Streams: a Centroid Approach*, in "Journal of Intelligent Information Systems (JIIS).", vol. 27, n^{o} 3, November 2006, p. 291-307.

[11] F. MASSEGLIA, P. PONCELET, M. TEISSEIRE, A. MARASCU. *Web Usage Mining: Extracting Unexpected Periods from Web Logs*, in "Data Mining and Knowledge Discovery (DMKD) Journal.", DOI 10.1007/s10618-007-0080-z, 2007.

[12] F. MASSEGLIA, D. TANASA, B. TROUSSE. *Web Usage Mining: Sequential Pattern Extraction with a Very Low Support*, in "Advanced Web Technologies and Applications: 6th Asia-Pacific Web Conference, APWeb 2004, Hangzhou, China. Proceedings", LNCS, vol. 3007, Springer-Verlag, 14-17 April 2004, p. 513–522.

[13] F. ROSSI, B. CONAN-GUEZ. *Functional Multi-Layer Perceptron: a Nonlinear Tool for Functional Data Analysis*, in "Neural Networks", vol. 18, n^{o} 1, January 2005, p. 45-60, http://hal.inria.fr/inria-00000599.

[14] D. TANASA, B. TROUSSE. *Advanced Data Preprocessing for Intersites Web Usage Mining*, in "IEEE Intelligent Systems", vol. 19, n^{o} 2, March-April 2004, p. 59–65.

Year Publications

Articles in International Peer-Reviewed Journal

[15] H. BAAZAOUI-ZGHAL, M.-A. AUFAURE, R. SOUSSI. *Towards an on-line Semantic Information Retrieval System based on Fuzzy Ontology*, in "Journal of Digital Information Management", vol. 6, n^{o} 5, October 2008, p. 375–385.

[16] R. BOULET, B. JOUVE, F. ROSSI, N. VILLA. *Batch kernel SOM and related Laplacian methods for social network analysis*, in "Neurocomputing", vol. 71, n^{o} 7–9, March 2008, p. 1257–1273.

[17] F. GUILLET, B. TROUSSE. *8th French-speaking conference on knowledge discovery and management (EGC2008): conference report*, in "SIGKDD Explorations", vol. 10, n^{o} 1, 2008, p. 53-54, http://doi.acm.org/10.1145/1412734.1412743.

[18] C. KRIER, F. ROSSI, D. FRANÇOIS, M. VERLEYSEN. *A data-driven functional projection approach for the selection of feature ranges in spectra with ICA or cluster analysis*, in "Chemometrics and Intelligent Laboratory Systems", vol. 91, n^{o} 1, March 2008, p. 43–53.

Articles in National Peer-Reviewed Journal

[19] B. CONAN-GUEZ, F. ROSSI. *Accélération des cartes auto-organisatrices sur tableau de dissimilarités par séparation et évaluation*, in "Revue des Nouvelles Technologies de l'Information", RNTI-C-2 Classification : points de vue croisés. Rédacteurs invités : Mohamed Nadif et Francois-Xavier Jollois, June 2008, p. 1–16.

Articles in Non Peer-Reviewed Journal

[20] M. PALLOT, B. TROUSSE, B. SENACH, B. RYETERS. *Living Lab Research*, in "ECOSPACE Newsletter No 5 - Special Issue dedicated to Living Labs", n^{o} 5, 2008, http://www.ami-communities.eu/wiki/ECOSPACE_Newsletter_No_5.

[21] Z. ZHANG, F. MASSEGLIA, R. JAIN, A. D. BIMBO. *Editorial: Introduction to the Special Issue on Multimedia Data Mining*, in "IEEE Transactions on Multimedia", vol. 10, n^{o} 2, 2008, p. 165-166.

Invited Conferences

[22] M.-A. AUFAURE. *Construction et évolution d'ontologies : application au web sémantique*, in "Ecol'IA 2008, Hammamet - Tunisia", 20-22 March 2008.

[23] M. CHAVENT, Y. LECHEVALLIER, F. VERNIER, K. PETIT. *Monothetic Divisive Clustering with Geographical Constraints*, in "COMPSTAT2008", P. BRITO (editor), Physica-Verlag, 2008, p. 67-76.

[24] Y. LECHEVALLIER, B. TROUSSE, A. DA SILVA, A. ABDOUROIHALANE, F. ROSSI, D. TANASA. *Apprentissage et fouille de données : Application au Web Usage Mining*, in "Ecol'IA 2008, Hammamet - Tunisia", 20-22 March 2008.

[25] N. VILLA, F. ROSSI, Q.-D. TRUONG. *Mining a medieval social network by kernel SOM and related methods*, in "Proceedings of MASHS 2008 (Modèles et Apprentissages en Sciences Humaines et Sociales), Créteil, France", June 2008.

International Peer-Reviewed Conference/Proceedings

[26] T. DESPEYROUX. *Evolution of Ontologies and Types*, in "WWW/Internet 2008 Conference, Freiburg, Germany", October 2008.

[27] C. FIOT, F. MASSEGLIA, A. LAURENT, M. TEISSEIRE. *Gradual Trends in Fuzzy Sequential Patterns*, in "12th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU'08)", 2008.

[28] C. FIOT, F. MASSEGLIA, A. LAURENT, M. TEISSEIRE. *TED and EVA: Expressing Temporal Tendencies Among Quantitative Variables Using Fuzzy Sequential Patterns*, in "17th IEEE International Conference on Fuzzy Systems (FuzzIEEE'08)", 2008.

[29] M. HADJOUNI, H. BAAZAOUI-ZGHAL, M.-A. AUFAURE, C. CLARAMUNT, H. BEN GHEZALA. *Towards a personalized spatial web architecture*, in "1st workshop Semantic Web Meets GeoSpatial Application, collocated with the AGILE Conference", may 2008.

[30] A. HASENFUSS, B. HAMMER, F. ROSSI. *Patch Relational Neural Gas - Clustering of Huge Dissimilarity Datasets*, in "Proceedings of the Third IAPR Workshop on Artificial Neural Networks in Pattern Recognition (ANNPR 2008), Paris, France", L. PREVOST, S. MARINAI, F. SCHWENKER (editors), Lecture Notes in Artificial Intelligence, vol. 5064, Springer, July 2008, p. 1–12.

[31] J. PEHCEVSKI, A.-M. VERCOUSTRE, J. THOM. *Exploiting Locality of Wikipedia Links in Entity Ranking*, in "The 30th annual European Conference on Information Retrieval (ECIR), Glasgow, Scotland", vol. LNCS, Volume 4956/2008, Springer, 2008, http://hal.inria.fr/inria-00232794/en/.

[32] B. SALEH, F. MASSEGLIA. *Time Aware Mining of Itemsets*, in "15th IEEE International Symposium on Temporal Representation and Reasoning (TIME'08)", 2008, p. 93-100.

[33] A.-M. VERCOUSTRE, J. PEHCEVSKI, J. THOM. *Using Wikipedia Categories and Links in Entity Ranking*, in "Proceedings of the sixth International Workshop of the Initiative for the Evaluation of XML Retrieval (INEX 2007) Focused Access to XML Documents, Germany, Schloss Dagstuhl", N. FUHR, J. KAMPS, M. LALMAS, S. MALIK, A. TROTMAN (editors), vol. LNCS, Volume 4862/2008, 2008, p. pp. 321-335.

[34] A.-M. VERCOUSTRE, J. THOM, J. PEHCEVSKI. *Entity Ranking in Wikipedia*, in "the 23rd Annual ACM Symposium on Applied Computing, Brésil Fortaleza", March 2008.

[35] W. WANG, S. GOMBAULT, T. GUYET. *Towards fast detecting intrusions: using key attributes of network traffic*, in "Proceedings of The Third International Conference on Internet Monitoring and Protection (ICIMP'2008)", Bucharest, Romania, IEEE computer society press, july 2008.

[36] W. WANG, S. GOMBAULT. *Efficient Detection of DDoS Attacks with Important Attributes*, in "3rd International Conference on Risks and Security of Internet and Systems (CRiSIS 2008)", Tozeur,Tunisia, IEEE press, october 2008.

[37] A. DE VRIES, A.-M. VERCOUSTRE, J. THOM, N. CRASWELL, M. LALMAS. *Overview of the INEX 2007 Entity Ranking Track*, in "Proceedings of the sixth International Workshop of the Initiative for the Evaluation of XML Retrieval (INEX 2007), Focused Access to XML Documents, Germany, Schloss Dagstuhl", N. FUHR, J. KAMPS, M. LALMAS, S. MALIK, A. TROTMAN (editors), vol. LNCS, Volume 4862/2008, Springer, 2008, p. pp. 1-23, http://hal.inria.fr/inria-00192489/en/.

National Peer-Reviewed Conference/Proceedings

[38] N. BEN MUSTAPHA, H. BAAZAOUI-ZGHAL, M.-A. AUFAURE, H. BEN GHEZALA. *Une MetaOntologie pour l'enrichissement incrémental d'ontologie de domaine basée sur la fouille du contenu Web*, in "Ecol'IA 2008, Hammamet - Tunisia", 20-22 March 2008.

[39] N. BEN MUSTAPHA, R. SOUSSI, H. BAAZAOUI-ZGHAL, M.-A. AUFAURE. *Une metaontologie pour l'enrichissement d'ontologie de domaine dans un système de recherche d'information*, in "2èmes journées francophones sur les ontologies", march 2008.

[40] M. CHARRAD, Y. LECHEVALLIER, G. SAPORTA, M. BEN AHMED. *Web Content Data Mining : la classification croisée pour l'analyse textuelle d'un site Web*, in "EGC", F. GUILLET, B. TROUSSE (editors), Revue des Nouvelles Technologies de l'Information, vol. RNTI-E-11, Cépaduès-Éditions, 2008, p. 43-54.

[41] M. CHAVENT, Y. LECHEVALLIER, F. VERNIER, K. PETIT. *Méthode de classification avec contraintes géographiques*, in "First Joint Meeting SFC - CLADAG, Caserta-Italy", Edizione Scientifiche Italiane, Société Francophone de Classification, June 2008, p. 225-228.

[42] A. DA SILVA, Y. LECHEVALLIER. *Stratégies de classification non supervisée basées sur fenï¿½res superposées dans le Web Usage Mining*, in "EcolIA 2008, Hammamet - Tunisia", 20-22 March 2008.

[43] A. DA SILVA, Y. LECHEVALLIER. *Stratégies de classification non supervisée sur fenêtres superposées : application aux données dusage du Web*, in "EGC", F. GUILLET, B. TROUSSE (editors), Revue des Nouvelles Technologies de l'Information, vol. RNTI-E-11, Cépaduès-Éditions, 2008, p. 219-220.

[44] C. FIOT, F. MASSEGLIA, A. LAURENT, M. TEISSEIRE. *Des séquences aux tendances*, in "26ème Congrès Informatique des organisations et systèmes d'information et de décision (InforSID'08)", 2008.

[45] D. FRANÇOIS, C. KRIER, F. ROSSI, M. VERLEYSEN. *Estimation de redondance pour le clustering de variables spectrales*, in "Actes des 10èmes journées Européennes Agro-industrie et Méthodes statistiques (Agrostat 2008), Louvain-la-Neuve, Belgique", January 2008, p. 55–61.

[46] F. ROSSI, Y. LECHEVALLIER. *Constrained variable clustering for functional data representation*, in "Proceedings of the first joint meeting of the Société Francophone de Classification and the Classification and Data Analysis Group of the Italian Statistical Society (SFC-CLADAG 2008), Caserta, Italy", June 2008.

[47] B. SALEH, F. MASSEGLIA. *Extraction d'itemsets compacts*, in "Extraction et gestion des connaissances (EGC'2008), Actes des 8èmes journées Extraction et Gestion des Connaissances, Sophia-Antipolis, France, 29 janvier au 1er février 200, 2 Volumes", 2008, p. 409-414.

[48] R. SOUSSI, N. BEN MUSTAPHA, H. BAAZAOUI-ZGHAL, M.-A. AUFAURE. *Un système d'aide à la recherche d'information en ligne basé sur les ontologies*, in "Conférence francophone en Recherche d'Information et Applications (CORIA), Trégastel, 12-14 march", mars 2008.

Scientific Books (or Scientific Book chapters)

[49] M. CSERNEL, F. DE CARVALHO. *The normal symbolic form*, in "Symbolic Data Analysis and the SODAS Software", E. DIDAY, M. NOIRHOMME-FRAITURE (editors), Wiley, 2008, p. 93-108.

[50] A. DA SILVA, Y. LECHEVALLIER, F. ROSSI, DE CARVALHO. FRANCISCO. *Clustering Dynamic Web Usage Data*, in "Innovative Applications in Data Mining", 2008, p. 71-82, http://dx.doi.org/10.1007/978-3-540-88045-5_4.

[51] A. DA SILVA, Y. LECHEVALLIER, F. DE CARVALHO. *Comparing Clustering on Symbolic Data*, in "Intelligent Text Categorization and Clustering", 2008, p. 81-94, http://dx.doi.org/10.1007/978-3-540-85644-3_5.

[52] Y. LECHEVALLIER, A. EL GOLLI, G. HÉBRAIL. *Improved generation of symbolic objects from relational databases*, in "Symbolic Data Analysis and the SODAS Software", E. DIDAY, M. NOIRHOMME-FRAITURE (editors), Wiley, 2008, p. 45-60.

[53] F. ROSSI, B. CONAN-GUEZ. *Multi-Layer Perceptron and Symbolic Data*, in "Symbolic Data Analysis and the SODAS Software", E. DIDAY, M. NOIRHOMME-FRAITURE (editors), chap. 20, Wiley, January 2008, p. 373–391.

[54] D. TANASA, F. MASSEGLIA, B. TROUSSE. *Mining Generalized Web Data for Discovering Usage Patterns*, in "Encyclopedia of Data Warehousing and Mining, 2nd edition", J. WANG (editor), IGI global, 2008.

[55] B. TROUSSE, M.-A. AUFAURE, B. LE GRAND, Y. LECHEVALLIER, F. MASSEGLIA. *Web Usage Mining for Ontology Management*, in "Data Mining with Ontologies: Implementations, Findings and Frameworks.", N. HÈCTOR OSCAR, GONZALEZ CISARO. SANDRA ELISABETH G, X. DANIEL HUGO (editors), chap. 3, Information Science Reference, 2008, p. 37-64.

[56] A. DE BAENST-VANDENBROUCKE, Y. LECHEVALLIER. *Clustering methods in symbolic data analysis*, in "SODAS2 Software: Overview and methodology", E. DIDAY, M. NOIRHOMME-FRAITURE (editors), Wiley, 2008, p. 429-444.

[57] F. DE CARVALHO, Y. LECHEVALLIER, R. VERDE. *Clustering methods in symbolic data analysis*, in "Symbolic Data Analysis and the SODAS Software", E. DIDAY, M. NOIRHOMME-FRAITURE (editors), Wiley, 2008, p. 181-204.

Books or Proceedings Editing

[58] M.-A. AUFAURE, O. BOUSSAID, P. KUNTZ (editors). *Atelier Mesures de similarité sémantique. 8èmes journées Extraction et Gestion des Connaissances*, january 2008, 132 pages.

[59] F. GUILLET, B. TROUSSE (editors). *Extraction et gestion des connaissances (EGC'2008), Actes des 8èmes journées Extraction et Gestion des Connaissances, Sophia-Antipolis, France, 29 janvier au 1er février 2008, 2 Volumes*, Revue des Nouvelles Technologies de l'Information, vol. RNTI-E-11, Cépaduès-Éditions, 2008.

[60] Z. JRAD, A. ABDOUROIHALANE, M.-A. AUFAURE (editors). *Atelier Modélisation utilisateur et personnalisation d'interfaces Web. 8èmes journées Extraction et Gestion des Connaissances*, january 2008, 68 pages, http://www-sop.inria.fr/axis/egc08/atelier_7.pdf.

[61] B. SENACH, B. TROUSSE (editors). *Atelier Analyse des usages et Mobilité urbaine. 8èmes journées Extraction et Gestion des Connaissances*, 2008, 38 pages.

Research Reports

[62] CEMAGREF, PIKKO, LIRMM, INRIA, NORMIND, SCRIPTAL. *Vers une infrastructure appropriée pour la conception participative (entre terrain et internet) et la mise en uvre des chartes de territoire - Plan de management du projet phase 1*, Projet INTERMED 2008-2010, ANR Deliverable, n^{o} 2, 2008.

[63] GET-ENST, LIRMM, CEREGMIA, INRIA, LGI2P, EDF R&D, FT R&D. *Etat de l'art des approches de résumés des flots de données*, Projet MIDAS 2008-2010, ANR Deliverable, n^{o} Tache 1.1, 2008.

[64] PIKKO, CEMAGREF, LIRMM, INRIA, NORMIND, SCRIPTAL. *Vers une infrastructure appropriée pour la conception participative (entre terrain et internet) et la mise en oeuvre des chartes de territoire - Plan d'expérimentation phase 1* , Projet INTERMED 2008-2010, ANR Deliverable, n^{o} 1, 2008.

[65] G. PANTHARI. *Mining Sequential Patterns at Optimal Time Granularity*, Technical report, Juillet 2008.

[66] G. SINGH. *Collaborative Mining of Outliers for Intrusion Detection*, Technical report, Juillet 2008.

[67] W. WANG, T. GUYET, R. QUINIOU, M.-O. CORDIER, F. MASSEGLIA. *Online and adaptive anomaly Detection: detecting intrusions in unlabelled audit data streams*, ARC SéSur intern document, Technical report, 2008.

Other Publications

[68] O. HADDAD. *Intégration dans un système de personnalisation de contrainte géographique et évaluation de cette approche*, Mémoire de fin d'étude, ENSI, Université de la Manouba, Tunis, Masters thesis, 2008.

[69] B. SENACH, J. MARLIER. *Deux études de mobilité urbaine et Inspection ergonomique de site et analyse de log - Confiance et covoiturage*, Presentation at World Usability Day, 2008, http://www.use-age.org/wud/2008/Slides_INRIA.pdf.

[70] A.-M. VERCOUSTRE, J. PEHCEVSKI, V. NAUMOVSKI. *Topic Difficulty Prediction in Entity Ranking*, submitted for publication, 2008.

[71] N. VERMA. *Detection Intrusion: how to preserve privacy in a collaborative environment?*, Masters thesis, LGI2P, France, Juillet 2008.

References in notes

[72] H.-H. BOCK, E. DIDAY (editors). *Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data*, Springer Verlag, 2000.

[73] P. BLACKBURN, J. BOS, K. STRIEGNITZ. *Learn Prolog Now!*, Texts in Computing, vol. 7, College Publications, 2006.

[74] M. CHAVENT. *A monothetic clustering method*, in "Pattern Recognition Letters", vol. 19, n^{o} 11, September 1998, p. 989-996.

[75] M. CHAVENT, Y. LECHEVALLIER, O. BRIANT. *DIVCLUS-T: A monothetic divisive hierarchical clustering method*, in "Computational Statistics and Data Analysis", vol. 52, 2007, p. 687-701.

[76] S. CHELCEA. *Agglomerative 2-3 Hierarchical Classification: Theoretical and Applicative Study*, Ph. D. Thesis, University of Nice Sophia Antipolis, March 2007.

[77] W. F. CLOCKSIN, C. S. MELLISH. *Programming in Prolog*, 5th edition, Springer Verlag, 2003.

[78] B. CONAN-GUEZ, F. ROSSI. *Speeding Up the Dissimilarity Self-Organizing Maps by Branch and Bound*, in "Proceedings of 9th International Work-Conference on Artificial Neural Networks (IWANN 2007), San Sebastian (Spain)", F. SANDOVAL, A. PRIETO, J. CABESTANY, M. GRANA (editors), Lecture Notes in Computer Science, n^{o} 4507, June 2007, p. 203–210.

[79] B. CONAN-GUEZ, F. ROSSI, A. EL GOLLI. *Fast Algorithm and Implementation of Dissimilarity Self-Organizing Maps*, in "Neural Networks", vol. 19, n^{o} 6–7, August 2006, p. 855–863, http://hal.inria.fr/inria-00174196.

[80] D. DAS. *Data steams: Some improvements on titled time windows*, Masters thesis, LGI2P, France, Juillet 2008.

[81] F. DE CARVALHO, R. M. C. R. DE SOUZA, M. CHAVENT, Y. LECHEVALLIER. *Adaptive Hausdorff distances and dynamic clustering of symbolic interval data*, in "Pattern Recognition Letters", vol. 27, n^{o} 3, February 2006, p. 167–179.

[82] T. DESPEYROUX. *Developing efficient parsers in Prolog: the CLF manual (v1.0)*, Technical Report, n^{o} RT-0328, INRIA, 2006, http://hal.inria.fr/inria-00120518/en/.

[83] T. DESPEYROUX, B. TROUSSE. *De la sémantique des langages de programmation à la vérification sémantique des sites Web*, in "Journées scientifiques de l'action spécifique Web Sémantiques, CNRS, Paris", October 2002, http://www.lalic.paris4.sorbonne.fr/stic/octobre/programme0209.html.

[84] E. DIDAY, G. GOVAERT. *Classification automatique avec distance adaptatives*, in "R.I.A.R.O Informatique Computer Science", vol. 11, n^{o} 4, 1977, p. 329-349.

[85] M. E. FAYAD, D. C. SCHMIDT. *Object-Oriented Application Frameworks*, in "Communication of the ACM", vol. 40, n^{o} 10, 1997, p. 32-38.

[86] B. J. J. FREY, D. DUECK. *Clustering by Passing Messages Between Data Points.*, in "Science", January 2007.

[87] M. JACZYNSKI, B. TROUSSE. *Patrons de conception dans la modélisation d'une plate-forme po ur le raisonnement à partir de cas*, in "Revue l'Objet", Numéro Spécial sur les patterns orientés objets, D. Rieu et J-P. Giraudon (guest editors), vol. 5, n^{o} 2, 1999.

[88] R. E. JOHNSON, B. FOOTE. *Designing Reusable Classes*, in "Journal of Object-oriented programming", vol. 1, n^{o} 2, 1988, p. 22–35.

[89] M. NOIRHOMME-FRAITURE, ET AL.. *User manual for SODAS 2 Software*, version 1.0, FUNDP, Belgique, april 2004.

[90] R. A. O'KEEFE. *The craft of Prolog*, MIT Press, Cambridge, MA, USA, 1990.

[91] Y. PITARCH, A. LAURENT, M. PLANTEVIT, P. PONCELET. *Fenêtres sur cube*, in "24ièmes journees Bases de Données Avancées (BDA'08)", 2008.

[92] Y. PITARCH. *Fouille de flot de données multidimensionnelles*, Masters thesis, University of Montpellier 2, France, Juillet 2008.

[93] F. ROSSI, B. CONAN-GUEZ. *Functional Multi-Layer Perceptron: a Nonlinear Tool for Functional Data Analysis*, in "Neural Networks", vol. 18, n^{o} 1, January 2005, p. 45–60, http://hal.inria.fr/inria-00000599.

[94] F. ROSSI, B. CONAN-GUEZ. *Un modèle neuronal pour la régression et la discrimination sur données fonctionnelles*, in "Revue de Statistique Appliquée", vol. LIII, n^{o} 4, 2005, p. 5–30, http://hal.inria.fr/inria-00001190.

[95] F. ROSSI, B. CONAN-GUEZ. *Theoretical Properties of Projection Based Multilayer Perceptrons with Functional Inputs*, in "Neural Processing Letters", vol. 23, n^{o} 1, February 2006, p. 55–70, http://hal.inria.fr/inria-00001191.

[96] F. ROSSI, N. DELANNAY, B. CONAN-GUEZ, M. VERLEYSEN. *Representation of Functional Data in Neural Networks*, in "Neurocomputing", vol. 64, March 2005, p. 183–210, http://hal.inria.fr/inria-00000666.

[97] F. ROSSI, A. HASENFUSS, B. HAMMER. *Accelerating Relational Clustering Algorithms With Sparse Prototype Representation*, in "Proceedings of the 6th International Workshop on Self-Organizing Maps (WSOM 07), Bielefeld (Germany)", ISBN: 978-3-00-022473-7, September 2007, http://dx.doi.org/10.2390/biecoll-wsom2007-144.

[98] F. ROSSI. *Model collisions in the dissimilarity SOM*, in "Proceedings of XVth European Symposium on Artificial Neural Networks (ESANN 2007), Bruges (Belgium)", April 2007, p. 25–30, http://apiacoa.org/publications/2007/dsom-collision-esann.pdf.

[99] H. SANEIFAR, S. BRINGAY, A. LAURENT, M. TEISSEIRE. *S2MP: Similarity Measure for Sequential Patterns*, in "Seventh Australasian Data Mining Conference (AusDM 2008), Glenelg, Australia", J. F. RODDICK, J. LI, P. CHRISTEN, P. KENNEDY (editors), CRPIT, vol. 87, ACS, 2008.

[100] H. SANEIFAR. *Clustering de motifs séquentiels et ses application à la détection d'intrusion*, Masters thesis, University of Montpellier 2, France, Juillet 2008.

[101] D. TANASA, B. TROUSSE. *Data Preprocessing for WUM*, in "IEEE Potentials", vol. 23, n^{o} 3, August 2004, p. 22–25.

[102] B. TROUSSE, B. SENACH, C. MANGEAT, G. CLOUET. *Analyse de l'usage d'un site d'information voyageurs (rapport final)*, 89 pages, Délivrable Projet PREDIT MobiVIP, n^{o} D5.3, Inria, March 2007.

[103] N. VERMA. *Detection Intrusion: how to preserve privacy in a collaborative environment?*, Masters thesis, LGI2P, France, Juillet 2008.

[104] N. VILLA, F. ROSSI. *A comparison between dissimilarity SOM and kernel SOM for clustering the vertices of a graph*, in "Proceedings of the 6th International Workshop on Self-Organizing Maps (WSOM 07), Bielefeld (Germany)", ISBN: 978-3-00-022473-7, September 2007, http://dx.doi.org/10.2390/biecoll-wsom2007-139.

[105] Y. YANG, M.-A. AUFAURE, C. CLARAMUNT. *Towards a DL-Based semantic user model for web personalization*, in "The First International Workshop on Knowledge-based User Interface, KUI 2007, Athens, Greece", IEEE Computer Society Press, 19 June 2007.

[106] X. ZHANG, C. FURTLEHNER, M. SEBAG. *Data Streaming with Affinity Propagation*, in "ECMLPKDD (2)", 2008, p. 628-643.