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 : Knowledge and Data Representation and Management

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, HdR]
Bernard Senach [CR1 Inria]

External Collaborator

Marc Csernel [Assistant Prof., Univ. Paris IX Dauphine]
Nicolas Faure [Post-doc, University Jean Moulin of Lyon3, until March 31]

PhD Student

Alice Marascu [Univ. Nice Sophia Antipolis (UNSA-STIC) until September 15]
Alzennyr Gomes Da Silva [Univ. Paris IX Dauphine, until September 30]
Elena Smirnova [Univ. Nice Sophia Antipolis (UNSA-STIC), January 1–May 7 and from September 1]
Chongsheng Zhang [Univ. Nice Sophia Antipolis (UNSA-STIC), ANR Midas]
Julie Marlier [Univ. Nice Sophia Antipolis, Intermed Contract, Cemagref Montpellier, until March 1st]

Post-Doctoral Fellow

Celine Fiot [University of Montpellier, until February, 8th]

Visiting Scientist

Malika Charrad [PhD, INS & ENSI Tunisia & CNAM Paris, April 7–27]
Mounia Naamary [PhD, Univ. Casablanca & OFPPT, Morocco, November 13 – December 13]
El Moukhtar Zemmouri [PhD, ENSAM & Univ. Casablanca, Morocco, November 13 – December 13]
Franciso De Carvalho [Prof., Federal Univ. of Pernambuco, Brazil, March 28–April 3 and August. 28–Sept. 12]

Administrative Assistant

Stephanie Aubin [TR Inria, part-time until april]
Stephanie Chaix [Temporary staff, part-time since june]
Anais Cassino [AI Inria, part-time]

Other

Cristina Isai [University of Timisoara, Roumania, until September 30]

2. Overall Objectives

2.1. Overall Objectives

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 contribute to user-driven open innovation as a way to foster innovation, 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 involve the users in the design process and to empower them, so that they can become co-designers. This is a new way to anticipate the usage and its analysis and also to consider 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 Scientific Foundations), 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 researches are organised to support the disruptive process of continuous innovation.

In this continuous process: design is never ended and relies on very short test-adapt-test cycles users are co-designers: they can contribute to design before/after market launch as ideas providers, as participants in test beds or field experimentations or even as solution providers when they are given the convenient tools.

To support this process, a large collection of tools and methods are needed and numerous efforts have already been engaged at european level to provide infrastructures for experimentations (for instance the Future Internet Research & Experimentation (FIRE) initiative launched in summer 2008), tools for creativity or sharing (Laboranova, CoSpaces, etc.).

In this context, our team focuses its effort on the technical and methodological environment needed to extract meaning from the huge amount of data issued from large and distributed information system.

Our researches are organised in three research topics:

1. **Topic 1** - *Mining complex data and IS data* : mainly temporal data, semantic Web mining (ontologies and Web mining) and semantics checking of an evolving IS. Most effort is put into two problems related to mining temporal data: a) analysing the evolution of user behaviours and b) summarizing and mining data streams.
2. **Topic 2** - *Mining Web 2.0 and collective usage data*: mining social networks, community detection, finding expert, collaborative filtering based recommender systems for information retrieval, social networks based recommender systems, personalization, etc.
3. **Topic 3** - the *Focus methodological and technical experimentation platform*: methods and tools based on a multidisciplinary approach (SHS and TIC) for the design and the evaluation of innovative services and for user-driven open innovation..

3. Application Domains

3.1. Panorama - Sustainable Development for Smart Cities

The project addresses any applicative field

a) requiring usage/data storage, preprocessing and analysis

- for designing, evaluating and improving huge evolving hypermedia information systems (mainly Web-based ISs), for which end-users are of primary concern,
- for a better understanding of service/product use with data mining techniques and knowledge management (for example in Transportation systems, Tourism, Security and Anomaly Detection, Internet of Objects, etc.),
- for social network analysis (for example in Web 2.0 applications, Business Intelligence, Sustainable Development, etc.)

and b) requiring user-driven innovation methods.

Even if our know how, methods and algorithms have a cross domain applicability, our team chooses to focus on **Sustainable Development for Smart Cities** and with a particular stress put on three main domains:

- Transportation systems (cf. section 3.2),
- Tourism (cf. section 3.3),
- e-participation in Environment and Health (cf. section 3.4).

3.2. Transportation Systems

Major recent 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 traveller's information systems for trip planning, the ability to provide real time recommendations for changing transportation means according to traffic information, and the quality of embedded services in vehicles to provide enhanced navigation aids with contextualised and personalised information.

Let us cite various projects where AxIS was/is involved to:

- **Mining Mobility Data** - PREDIT (2004-2007): The MobiVIP project has been an opportunity to collaborate with local Institutions (*Communauté d'Agglomération de Sophia Antipolis - CASA*) and SMEs (VU Log) and apply AxIS know-how in mining spatial and temporal data issued from vehicules equipped with GPS and from the reservation server and in clustering trajectories (with semantic distances). Even if we didn't apply our know how in mining data streams in this project, we think it will be crucial in the future with more and more equipped vehicules with GPS.
- **Traveller's Information Systems** - evaluation of two Web sites:
  - the Envibus web site provides information about a bus network its evaluation was done by coupling ergonomic analysis and usage mining
  - the Otto&co web site support car-sharing our cooperation about car-sharing done in 2008 with an evaluation of the Otto&co site in the context of the action COLOR Cuscov is still lasting.
- **Advanced Transportation Systems - Multimodality**: PREDIT (2010-2012): the TIC TAC project (cf. section 7.1.6) aims to optimize travel time by providing in an area with weak transportation services, a just in time on demand shuttle based on real time information.

AxIS participated also in two proposals to the FUI9 call for proposals related to multimodality in mobility:

- The NaVTI project (resp: Continental Automotive - Rambouillet) labelled by the Moveo and System@tic clusters
- The e-transports project (resp. Moviken - Paris) labeled by the Advancity cluster: "Mise au point d'un système global de distribution des transports pour le développement du tourisme durable et expérimentation sur le territoire des alpes maritimes".

3.3. Tourism

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

Again to illustrate our role in such a domain, let us cite some past and current projects where AxIS is involved related mainly to **Semantic Web Mining** [1] and **information retrieval**.

- AxIS was involved in the **RNTL Eiffel project** (cf. section 7.2.1), whose goal is to provide users with an intelligent and multilingual semantic search engine dedicated to the tourism domain. This allows 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.In this context we worked on improving usage data mining by exploiting ontologies.
- Other researches have been carried out using log files from the city of Metz. This city was chosen because its 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 log files and to identify possible benefits in designing or **updating a tourism ontology** [88].
- **Providing Tourism Information linked to Transportation information** AxIS has already studied recommender systems in order to provide users with personalised transportation information while looking for tourism information such as cultural information, leisure etc (cf. our recommender Be-TRIP (2006) and [72].
- We submitted this year a Pacalabs (call 2) proposal called HOTEL-REF-PACA in order **to better refer the web sites of hotels** from the region of TOURVAL in PACA (mainly *Vésubie* territory), with an approach based on a better understanding of usage from the internauts.

3.4. e-participation in Environment and Health

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. So, sustainable development is closely associated with citizen participation. 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) and numerous different topics are addressed in the field.

Below are some topics where AxIS was or is involved in:

- **Preprocessing and analysing collective usage data and social networks** from group discussions related to a document: In the Intermed project related to water management, we started to study new services based on synthesizing discussions or on analysing disucssions (cf. section 7.2.2) and we pursued this year by elaborating a preprocessing tool based on disucssions data.
- **Methods and tools for open innovation**: we submitted a proposal for supporting the design of innovative services by citizens from public data in collaboration with Fing (Marseille) and Ademe (Sophia Antipolis) for the Inria Color call for actions.

Main AxIS topics relevant for these domains are: social network analysis, personalization and information retrieval, recommender systems, expert search, design and evaluation of methods and tools for open innovation, usage mining, mining data streams.

[1] By Semantic Web Mining, we mean the mutual benefits between two communities *Semantic Web* and *Web Mining*. In our case, we exploit a) ontologies and semantic data for improving usage analysis, personnalised services, the quality of results of search engines and for checking the content of an IS and also b) we exploit usage data for updating ontologies.

4. Software

4.1. Introduction

AxIS has developed several softwares or toolboxes:

- CLF for generating efficient parsers,
- AxIS Log preprocessing tool (AxISLogMIner, AWLH) for web usage mining and Methods for sequential pattern extraction,
- Clustering and Classification toolbox,
- SODAS 2 software,
- CBR*Tools for reusing experiences (requiring the management of historical data),
- Broadway*Tools for designing adaptive Web-based recommendation systems and collaborative information retrieval support.

We participated also in the SODAS 2, result of a past European project and in BibAdmin for the management of a collection of publications.

We have pursued our study of a new platform called FOCUS (CPER Télius) by testing, redesigning part of these software (cf. section 5.14) and by developing new tools before integrating them. Indeed this year new algorithms and tools have been designed and implemented related to mine data streams from A. Marascu's thesis and to analyse the evolution of user behaviours from A. Da Silva's thesis [18] (cf. section 5.2). Let us notea first moke-up for preprocessing data from debates (cf. section 5.13).

4.2. CLF - Computer Language Factory

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 [62], [79], [58] 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 [66].

4.3. AxIS Web Log Preprocessing and Methods for Sequential Pattern Extraction

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

AxISLogMiner issued from D. Tanasa's thesis is a software application that implements

- our preprocessing methodology [16] for Web Usage Mining (WUM)
- and methods for sequential pattern extraction with low support (*Cluster & Divide* and *Divide & Discover*[14]): See Chapter 3 of Tanasa's thesis [87] for more details.

In 2008 in the context of the Eiffel project, we isolated and redesigned the core of AxISlogMiner preprocessing tool (we called it AWLH) composed of a set of tools for pre-processing web log files. AWLH 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.

Now the current version of our Web log processing offers:

- 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.

In 2009 we developed a tool based on an open source project called "OpenSympony ClickStream" for recording 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 Web Usage Mining process.

More an extended version of AWLH has been developed for capturing and structuring data issued from annotating documents inside discussion forums.

4.4. Clustering and Classification Toolbox

Participants: Marc Csernel, Alzennyr da Silva, Francesco de Carvalho, Yves Lechevallier [co-correspondant], 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 [64], [63], [85], [84], [89], 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 [80], [83], [82], [81] (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 [65]. 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 [60] (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 [61]).

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. SODAS 2 Software

Participants: Yves Lechevallier [correspondant], Marc Csernel.

The SODAS 2 Software [78] is the result of the European project "ASSO" (Analysis System of Symbolic Official data) (2001-2004). It supports the analysis of multidimensional complex data (numerical and non numerical) coming from databases mainly in statistical offices and administration using Symbolic Data Analysis [56]. This software is registrated at APP. The latest executive version 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 were:

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

Those AxIs contributions have been registered at APP.

4.6. CBR*Tools

Participant: Brigitte Trousse [correspondant].

CBR*Tools is an object-oriented framework [74], [69] 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 past contracts: EPIA and MobiVIP.

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

4.7. 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 from Jaczybski'thesis for supporting Web browsing, for supporting browsing inside a Web-based information system or for supporting query formulation in the context of a meta search engine.

4.8. BibAdmin

Participant: Brigitte Trousse [correspondant].

BibAdmin developed by S. Chelcea (ex-PhD student) 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 since the end of 2005 under the GNU GPL license on INRIA's GForge server.

5. New Results

5.1. Introduction

Below are our results obtained this year according to our three research topics,

1. **Topic 1** - *Mining complex data and IS data* : analysing evolutive data (cf. section 5.2), summarising and mining data streams (cf. sections 5.3 and 5.4), content analysis and ontologies (cf. sections 5.5 and 5.6), clustering and Web Content (cf. section 5.7). Let us not also a more generic result in data mining related to adaptive distances in clustering (cf. section 5.8).
2. **Topic 2** - *Mining Web 2.0 data*: expert search (cf. sections 5.9 and 5.10).
3. **Topic 3** - the *Focus methodological and technical experimentation platform* (cf. section 5.13 of the 2008 activity report): a) development of new tools for evolving usage data (cf. section 5.11), for mining data streams (cf. section 5.12), for structuring data from debates (cf. section 5.13) and b) improvment of our tool for mapping textual documents (cf. section 5.14).

Let us note also that some works made in 2008 and already described in the 2008 activity report were published in 2009:

- The work on "Evolution patterns and gradual trends" described in the activity report of 2008 in Section 5.6, has been published in 2009 by an international journal [24].
- The work on "Intrusion detection and common outliers" described in the activity report of 2008 in Section 5.7, has been published in 2009 by an international conference [34] and a national conference [43], [42].
- The work on "Outlier detection in data streams" described in the activity report of 2008 in Section 5.8, has been published in 2009 by two international conferences [31], [32] and a national conference [41].
- The work on "Online and adaptive intrusion detection in unlabelled audit data streams" described in the activity report of 2008 in Section 5.9, has been published in 2009 by an international conference [35] and a national conference [44].

A previous work on "Efficient mining of sequential patterns with time constraints" has been published by an international journal [25].

5.2. Analysis of Evolving Web Usage Data

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

Nowadays, more and more organizations are becoming reliant on the Internet. The growing number of traces left behind user transactions (e.g. customer purchases, user sessions, etc.) automatically increases the importance of usage data analysis. Indeed, the way in which a web site is visited can change over time. By consequence, the usage models must be continuously updated in order to reflect the current behaviour of the visitors. Such a task remains difficult when the temporal dimension is ignored or simply introduced into the data description as a numeric attribute.

It is precisely on this challenge that the present thesis of Ms A. Da Silva is focused [18], [46], [45], [48], [47]. In order to deal with the problem of acquisition of real usage data, we propose a methodology for the automatic generation of artificial usage data over which one can control the occurrence of changes and thus, analyse the efficiency of a change detection system. Guided by tracks born of some exploratory analyzes [45], we propose a tilted window approach for detecting and following-up changes on evolving usage data. In order to measure the level of changes, this approach applies two external evaluation indices based on the clustering extension. The approach also characterizes the changes undergone by the usage groups (e.g. appearance, disappearance, fusion and split) at each timestamp. The proposed approach is totally independent of the clustering method used and is able to manage different kinds of data in addition to the Web usage data. The effectiveness of the approach is evaluated on artificial data sets of different degrees of complexity and also on real data sets from different domains (academic, tourism and marketing).

Moreover, as a result of her thesis a new tool was developed: **ATWUEDA** (Axis Tool for Web Usage Evolving Data Analysis) [47] (cf. section 5.11).

Ms A. Da Silva defended her PhD in September 2009 at University of Paris IX Dauphine. All the results obtained from this work are described in her PhD thesis [18].

5.3. Mining Sequential Patterns in Data Streams and Managing their History

Participants: Alice Marascu, Florent Masseglia, Yves Lechevallier.

Mining data streams is an important challenge nowadays, due to i) their strong characteristics and ii) the growing number of systems that produce them. Analyzing and managing the history of these streams implies to use a degree of approximation. Marascu's thesis (partially founded by PACA Region) was devoted to such problems with numerous publications in 2009 [32], [41], [31], [42], [34] (cf. section 5.1).

More precisely, this year, Marascu's thesis focused on designing *A Fast Approximation Strategy for Summarizing a Set of Streaming Time Series*.

Summarizing a set of streaming time series is an important issue that reliably allows information to be monitored and stored in domains such as finance, frequent pattern extraction, networks, etc. To date, most of existing algorithms have focused on this problem by summarizing the time series separately. Moreover, the same amount of memory has been allocated to each time series. Yet, memory management is an important subject in the data stream field, but a framework allocating equal amount of memory to each sequence is not appropriate.

In fact, most of existing solutions to the problems raised by the data stream history management consider that old events are less interesting and recent events should be kept at fine granularity. The main idea of these methods is that, according to the human memory, we are generally more interested in recent events than we are in older ones. As a matter of fact, the human memory keeps a detailed report of the "salient" events but a careless record of the "unimportant" events (i.e. an accident versus a breakfast). Based on this principle, we have proposed a novel approach where the most important events are kept with high fidelity, while the less important events kept at a coarser resolution. For this purpose we have presented **GEAR** (Global-Error Aware Representation), a fast on-line approximation algorithm.

Solving this problem calls for an approximation technique that would fastly merge the lines representing a set of values. Marascu's Ph.D. thesis report [19] gives the description of GEAR and **MiTH**[2] (Middle THrough), a new model for merging two lines.

This work has been submitted and accepted as a long paper by an international conference (ACM SAC 2010) and by a national conference (EGC 2010).

Ms A. Marascu has defended her Ph.D thesis [19] (University of Nice Sophia Antipolis) in September 2009 at Inria.

5.4. Discovering Informative Feature Sets in Data Streams

Participants: Chonghseng Zhang, Florent Masseglia.

One particular case of data streams is that of categorical data, where each object is a set of features. Unfortunately, in the real world, objects do not arrive directly as a set of features. Let us consider, for instance, the usages of a Web site. Each user will be considered as an object and each page requested by that user will be considered as a feature. The users make their request one after the other. Which means that, at each step, the set of features for that client will evolve. Analyzing such kind of data streams, made of evolving objects, is an important topic today, associated to numerous challenges.

Chongsheng Zhang is studying these challenges in his Ph.D thesis. Part of this thesis is funded by MIDAS (Mining Data Streams), an ANR project (cf. section 7.2.3).

[2] REGLO (Regression attentive à l'Erreur GLObale) and AMi (Approximation par les MIlieux) in French publications

This year, Chongsheng has particularly studied feature extraction from such data streams. Feature selection is the task of selecting interesting or important features, and removing irrelevant or redundant ones. There is a lot of existing works on feature selection [75]. Depending on different applications and needs, we may have different interestingness measures to assess the weights of the features or feature sets [77]. Information based feature selection methods require us to compute the probabilities for all the features and all the possible feature subsets. As a result, it is extremely time consuming and exhaustive and the case is much worse for the streaming data because the probabilities for feature and feature sets are always changing. To deal with this problem, we introduce the heuristic algorithm *StreamHI*. It is based on a candidate generation and a pruning principle, the former will keep and monitor as many high likely candidates as possible while the latter will remove the redundant and hopeless candidates. Our contributions are i) a definition of the problem of online informative feature set selection from the data streams and ii) *StreamHI*, a heuristic method for mining the informative feature sets in real time. We ran *StreamHI* against naive methods through a series of experiments, which demonstrate its efficiency and effectiveness:

1. Execution times of *StreamHI* are one order of magnitude lower than those of a naive approach based on a method from the litterature designed for static data.
2. The itemset extracted by *StreamHI* is usually the same, and sometimes better (in terms of entropy), than the itemset extracted by the naive approach.

This work has been submitted and accepted as a long paper by a national conference (EGC 2010).

For 2010, Chongsheng will work on discovering different schemas from data streams where objects are sets of features. Today, in usage data streams, most objects are made of only one feature. Extracting relevant and useful knowledge from these data streams is a challenge since one single feature is not very informative. Our goal is to improve that knowledge despite the rare occurences of objects having more than one feature.

5.5. Validation and Evolution of ISs and of Ontologies

Participant: Thierry Despeyroux.

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

Following previous experiments [68] we developed a new methodology for XML documents verification [8]. Nowadays, an IS does not only contain documents but it also makes 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 [9].

These results have been selected as a chapter of the book "Emergent Web Intelligence: Advanced Semantic Technologies" that will appear in April 2010 [67].

5.6. Representation of Ontologies

Participant: Thierry Despeyroux.

As an exploratory research in collaboration with M. Arnoux (University of Rennes), we compared ontology engineering and software engineering. We have investigated analogies between ontologies on one hand, types, objects and data bases on the other one, taking into account the notion of evolution of an ontology. We represent a unique ontology using different paradigms, and observe that the distance between these different concepts is small. We deduced from this constatation that ontologies and more specifically ontology description languages can take advantage of being fertilizated with some other computer science domains and inherit important characteristics as modularity, for example [26].

5.7. Block clustering and Web Content Data Mining

Participants: Malika Charrad, Yves Lechevallier.

Simultaneous clustering, usually designated by biclustering, co-clustering or block clustering, is an important technique in two way data analysis. The goal of simultaneous clustering is to find sub-matrices, which are subgroups of rows and subgroups of columns that exhibit a high correlation. Our aim is to analyze textual data of a web site. Our approach [28] 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.

One of the major problems of simultaneous clustering algorithms, similarly to the simple clustering algorithms, is the need of specifying the optimal number of clusters. This problem has been subject of wide research. Numerous strategies have been proposed for finding the right number of clusters. However, these strategies can only be applied with one way clustering algorithms and there is a lack of approaches to find the best number of clusters in block clustering algorithms. Our goal [28] is to extend the use of these indices to block clustering algorithms. We propose in [36] to use The Laplace operator and the differential operator in two dimensions to detect the good number of clusters.

5.8. Adaptive Distances in Clustering Methods

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

The adaptive dynamic clustering algorithm 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.

We extended these models in order to cluster objects described by interval-valued variables [23] and [22]. Adaptive distances were used to compare a pair of vectors of intervals simultaneously taking into account both the lower and the upper boundaries of the intervals assumed by the interval-valued variables. Moreover, various tools for the interpretation of the partition and cluster of interval-valued data were also presented.

Aditionally, we proposed a clustering algorithm for relational data (dissimilarity data) which is also able to cluster in a polynomial time objects described by set-valued or interval-valued variables constrained by rules [21]. The constraints given by the rules were taking into account through a suitable dissimilarity function.

5.9. User Oriented Expert Finding

Participants: Elena Smirnova, Brigitte Trousse.

The goal of expert search is to find a set of people who are knowledgeable in a given topic. Experts can be demanded for the variety of purposes: problem solving, question answering, providing more detailed information on a topic, to name a few. Models for expert finding proposed so far were solely focused on returning the most knowledgeable persons as experts on a given topic [57] [86]. In this work we argue that in user oriented setting the notion of the best expert depends on the user and his needs. We propose a rational user model to measure the system performance without involving a real user and show how given this model the resulted set of experts is changing relatively to the list returned by classical expert finding system. Our model is based on two factors that influence user satisfaction - time spent on the task and knowledge gained after [70]. Next steps in this work include experiments with real users to verify the positive impact of newly introduced user parameters and to uncover other factors that influence on user's choice.

5.10. Latent Variable Models for Expert Finding

Participants: Elena Smirnova, Brigitte Trousse.

Latent variable models are widely used in many areas of data analysis: speech recognition, image restoration, finance forecasting. In document analysis one can highlight two methods uncovering hidden structure of the documents. The PLSA method proposed by T. Hoffman [73] and topic models introduced by D. Blei et al. [59] acquired popularity and have proved to be very useful statistical description of document corpus. Latent variable models can be directly applied to expert finding task because it implies existence of "hidden" experts who create "observed" documents. Therefore the generative model of a document includes modeling hidden author's contribution to a document based on observed set of authors and modelling hidden expert profiles based on observed words in a document. In previous work [86] inference in this kind of models was done using EM algorithm. Here we note the empirical nature of this point estimation and in contrast we take a fuller Bayesian approach when prior knowledge is used for parameter estimation. In this case variational Bayesian methods are appropriate for relatively fast and reliable inference.

5.11. Tools for Analysing Evolving Web Usage Data

Participants: Alzennyr Da Silva, Yves Lechevallier.

As a result of Da Silva's thesis a new tool was developed: **ATWUEDA**[3] [47] for Web Usage Evolving Data Analysis. It is available at INRIA's gforce website: http://gforge.inria.fr/projects/atwueda/. A. Da Silva presented part of her work in a working research group at CNAM-Paris [48].

This tool was developed in Java and uses the JRI library in order to allow the application of R functions in the Java environment. R is a programming language and software environment for statistical computing (http://www.r-project.org/). The ATWUEDA tools is able to read data from a cross table in a MySQL data base, split the data according to the user specifications (in logical or temporal windows) and then apply the approach proposed in the Da Silva's thesis (cf. section 5.2) in order to detect changes in dynamic environment. The proposed approach characterizes the changes undergone by the usage groups (e.g. appearance, disappearance, fusion and split) at each timestamp. Graphics are generated for each window analyzed exhibiting statistics that characterizes changing points over time.

5.12. Softwares for Mining Sequential Patterns in Data Streams

Participants: Alice Marascu, Florent Masseglia, Yves Lechevallier.

As a result of Marascu's thesis [19], a collection of softwares have been developed for knowledge discovery (sequential patterns and clusters) and security in data streams.

- 3 **clustering algorithms for mining sequential patterns** (Java) in data streams have been developped by A. Marascu during her thesis [19]. The softwares take batches of data in the format "Client-Date-Item" and provide clusters of sequences and their centroids in the form of an approximate sequential pattern calculated with an alignment technique.
  - SMDS compares the sequences to each others with a complexity of $O(n2)$.
  - SCDS is an improvement of SMDS, where the complexity is enhanced from $O(n2)$ to $O(n.m)$ with n the number of navigations and m the number of clusters.
  - ICDS is a modification of SCDS. The principle is to keep the clusters' centroids from one batch to another.
- **GEAR** is an implementation (Java) of the history management strategy proposed in Marascu's thesis [19]. It takes a set of time series and provides a memory representation of these series based on a new principle, where salient events are important (in contrast to the recent events of decaying models).
- **WOD** is an implementation (Java) of our outlier detection method proposed in [32]. Given a set of clusters and their sizes (a list in the format "clusterId-size") WOD gives a natural separation between small ang big clusters thanks to a wavelet decomposition of the list.

[3] ATWEDA: Axis Tool for Web Usage Evolving Data Analysis

More on intrusion detection, let us cite **COD** an implementation of our intrusion detection method proposed in [34]. It takes data files in the format "Client-Date-Item" obtained after having preprocessed the usages of two web sites. It gives their common outliers by means of WOD.

5.13. Tool for Preprocessing Data from Debates based on Annotations

Participants: Céline Fiot, Cristina Isai, Brigitte Trousse.

For the purpose of participative democraty, we studied the annotation process related to a document in the context of a debate. We proposed a model for structuring data issued from annotating documents during debates by users and implemented it via our log preprocessing core AWLH (cf. section 4.3). Experiments on large datasets and various contexts of discussions are needed to validate our model and the integration of this work in the Focus platform.

5.14. Tool for Mapping Textual Documents

Participants: Nicolas Faure, Cristina Isai, Brigitte Trousse [resp.].

In 2008, the DRAST (Direction de la Recherche et des Affaires Scientifiques et Techniques), component of the French Ministry of Environment[4], issued a grant for the design and implementation of a tool which automatically compares textual documents submitted in the context of proposals following generic requests.

The work undertaken, included in what was named Calfat-Ville Project, resulted in the end of 2008 in a crude prototype and a methodology which met DRAST's objectives.

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).

During 2009, the tool was improved by fixing minor bugs and most notably expanding the number of syntactic patterns used to extract syntagms. Also, it was intensively tested on various documents, including literary documents. The results of these tests show a necessity to evolve towards semantics to improve the methodology, by including some taxonomical resources (ontology, thesaurus, lexical network,...) in the process.

Work currently in progress includes the re-use of the extraction process in order to automatically produce such a taxonomy from the documents and obtain fully satisfying results.

6. Contracts and Grants with Industry

6.1. France Telecom - Orange Labs

Participants: Yves Lechevallier, Alice Marascu, Florent Masseglia, Bernard Senach, Brigitte Trousse.

Orange labs CRE (2009-2010) Contract 905201

Orange Labs and Inria have signed a CRE (*Contrat de Rercherche Externalisée*) for 6 months (2009-2010) under the scientific responsability of F. Masseglia, where the goals are:

- To transfer a technology from AxIS to Orange Labs.
- To add a new step devoted to specific needs of Orange.
- To perform experiments on Orange Data.

The AxIS technology involved in this project is the SMDS [76] method proposed in the Ph.D thesis of A. Marascu. SMDS is an algorithm intending to **extract sequential patterns from data streams**. The data stream is processed by batches that accumulate the data for the same time period. Afterwards, each batch is processed in order to i) obtain clusters of data and ii) extract approximate sequential patterns that will summarize the clusters. Each such sequential pattern will be considered as a center of the cluster and will evolve from one batch to another.

[4] Ministère de l'Ecologie, de l'Energie, du Développement Durable et de l'Aménagement du territoire

The new step added by request of Orange is a second clustering process, applied to the approximate sequential patterns.Eventually, Orange Labs provided us with 3 months data coming from the mobile portal usage. These log files contain the request of users on the portal, made from their mobile devices.

Strategic Partnership between France Telecom - Inria

B. Trousse, with her colleague M. Boullé (Orange Labs) was in charge of animating one of the four research topics selected by Orange Labs for the strategic partnership between France Telecom - Orange Labs and Inria. Two workshops (Paris march 13, Rocqunecourt may 15) were organised by the Inria direction. Two meeting reports were made and three project proposals on this topic have been proposed by the interested EPIs and Orange labs colleagues: two with 5 EPIs (AxIS, DOLPHIN, DREAM, SEQUEL, TAO) and another with 2 EPIs (TAO and DOLPHIN).

B. Senach and B. Trousse are exploring with Orange Labs (Sophia Antipolis) different topics related to Smart Cities for future collaborations in the context of the living lab *ICT Usage Lab* (cf. section 7.1.1).

6.2. Other Industrial Collaborations

Participants: Brigitte Trousse, Yves Lechevallier, Florent Masseglia, Bernard Senach.

To identify future collaborations. we pursued our industrial contacts and/or participated to different industrial meetings:

- SAP, Sophia Antipolis related to open innovation and the living lab approach,
- Alcatel Lucent related to recommender systems,
- EDF Paris on usage mining and social network analysis in the context of the Bilab (Business Intelligence Laboratory).

Following requests from the Perferencement SME (Nice) and also Senuceo (Sophia Antipolis) in the context of its incubation with the Paca-Est incubator, we are elaborating two proposals for a collaboration at the beginning of next year.

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. We were this year very active both in two regions,

- in Paca:a new PREDIT project TIC-TAC starting in 2010 (Sophia Antipolis), 3 Pacalabs proposals, living lab “ICT Usage Lab” and numerous contacts with innovation actors, CEFH, a lot of contacts with regional SMEs and territories
- and in Ile de France: a new Web 2.0 project SCAR starting in 2010 (Paris) and 2 FUI9 proposals.

.

7.1.1. Living Lab “ICT Usage Lab”

Participants: Brigitte Trousse, Bernard Senach.

AxIS has pursued with current founding partners (Orange Labs, INRIA, CSTB, UNSA) the setting 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. A booklet, a flyer and a web site are under construction. Several contacts have been set up to extend the current partnership (SAP, GFI, City Of Nice, SCS and pegase clusters, Fing,..) and to decide the legal entity.

Several invited talks related to the living lab "ICT Usage Lab" have been done:

- at the working days *Open Innovation & Living Labs* organised by the ICT Mission of PACA Regional Council.(6-7 April, Marseille): B. Trousse & R. Decorme of CSTB (with B. Senach and S. Crave of Orange Labs).
- at the administration committee of the Telecom valley association,, june 9: B. Trousse with B. Senach and S. Crave (Orange Labs).
- at the innovation commission of the Telecom Valley association: B. Trousse, B. Senach, V. Giraudon (Orange Labs) and R. Decorme (CSTB).
- at ICE conference, workshop COINS, Noordwyck, Nederlands, june 24: B. Trousse with B. Senach [55].

Locally several meetings have been conducted in order to coordinate activities of several research teams involved in SHS from UNSA and Telecom ParisTech related to usage analysis and open innovation in the context of the Campus STIC.

We collaborate with Mediterranée Technologies and the other Paca living lab *TPmed* to write down a position paper "La voix des LL en PACA"

At the national level, the ICT Usage Lab organised the second national meeting of french living labs which was hosted at CSTB in Sophia Antipolis the 14th of april. It was an opportunity to set up a cartography of the elevent current french living labs.

Several contacts have been established in order to answer to different calls for proposals:

- with other living labs in France and in Europe for the FP7-ICT call 5: three proposals (Fireball, UDIHouse, Elliot) were submitted (cf. section 7.3.1).
- with SMEs and academic partners for the Pacalabs call 2. Three proposals were submitted (cf. section 7.1.2),
- with the Inria EPI Lognet for the INTERREG IV ALCOTRA proposal called MyMed (resp. L. Liquori). This proposal has been accepted.
- with Fing and Ademe for the Inria Color call in order to initialize a collaboration related to the design by citizens of innovative services based on public data (such as environmental data): the proposal is called CDISOD [5].

The ICT Usage Lab will organize the 1st LL Summer school in Sophia Antipolis which will be held in August 2010. Inria will be in charge of the local organisation.

7.1.2. PacaLabs Initiative

Participants: Bernard Senach, Brigitte Trousse.

The Regional Council has launched the Pacalabs initiative for supporting experimentations of innovative services (from SMEs) or exploratory ideas (from academic world) with real users.

B. Trousse belonged to the expert committee of the Pacalabs Call 1 and to the strategic committee of the Pacalabs orientation of the Regional Council.

AxIS participated in three proposals to the Pacalabs call 2 (under review)

- **ECOFFICES** is an eco-challenge within an enterprise. Some offices are equipped with sensors and feed back concerning energy consumption is provided. The goal of the project is to provoke behavioural changes. In this project our team is in charge of the evaluation phase: usage data concerning actions on actuators will be registered and employee behaviour will be tracked.

[5] CDISOD: Citizen Driven Innovation Services based on Open data

- **TAC-VAL** is a project conducted with the SME Ivès which provides guidance to disabled people. In collaboration with the town council, we will establish a "usability diagnosis" of the urban space in a small town (Valbonne), finding out the problems that disabled encounter on their way to their affairs. In this project, our team will provide a methodological assistance to set up experiments
- **HOTEL-REF-PACA**: this project is conducted with Perferencement an SME specialized in web site referencing. This project aims at improving hinterland tourism. Experiments of different new referencing rules will be conducted with Web site visitors in order to study their effect on behavioural changes and touristic choices.

7.1.3. FOCUS Platform (PACA CPER TELIUS)

Participants: Brigitte Trousse, Bernard Senach, Yves Lechevallier.

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, such as 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. The FOCUS platform is composed of two levels:

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.

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.

This year we bought material: 10 labtops for experiments, 1 eye-tracker, This material has been used by two Inria research projets: Arobas team project used one of our labtop for an advanced robotic project and Edelweiss an another one for eye-tracking research.

7.1.4. CEFH

Participants: Bernard Senach, Brigitte Trousse.

The CEFH plateform (*Centre d'Etudes des Facteurs Humains*), supported by the PEGASE cluster, was submitted last year to the call "Plateforme d'innovation" (DGE and CDC). This experimentation platform dedicated to human factors studies in risky domains (nuclear power plant, aeronautics industry, process control was accepted in a first selection step. This year B. Senach and B. Trousse had regular meetings in Aix-en-Provence between march and june to finalize the proposal of the second selection step. The submission of the proposal was delayed according to the decision of the founding partners.

7.1.5. Collaborations with Territories in PACA

Participants: Bernard Senach, Brigitte Trousse.

Links with territories:

- CASA, City of Antibes: transportation projects (Mobivip 2007, Cuscov 2008, TIC TAC 2010)
- CASA, City of Valbonne: accessible mobility project (TAC-VAL) (cf. section 7.1.2)
- Urban Community of Nice Provence Côte d'Azur, City of Nice: Support of the FP7-ICT proposal called FIREBALL related to a specific coordination action on *Smart Cities* at the european level which is also to be linked to the Living Lab approach and Internet of Objects (cf. section 7.3.1).
- City of Nice, Plaine du var "Ecovallée" (November 11): meeting with O. Sassi (Development Director) and S. Lartigue (innovation and economics aspects).

We have also contact with M. Dioudonnat of the "Institute of Mediterranée" related to the **Medlab** project.

7.1.6. Collaborations with SMEs

Participants: Bernard Senach, Brigitte Trousse, Yves Lechevallier.

Two SMEs involved in **two accepted projects**:

- **Wozaik (Paris)** related to *Social networks and Recommender Systems*: the SCAR 2.0 project was one of the 44 accepted projects from 340 submissions to the Call for proposals *Web 2.0 - Web innovants*. It is one of the three accepted projects at the national level where Inria is partner.
- **VuLOG (Antibes)** related to Usage Analysis and Experimentation for the *PREDIT* call on Mobility in Regions, the TIC-TAC project (advanced transportation system - shuttle on demand and real time traveller information).

Four other SMEs involved in common proposals with AxIS:

- **Perferencement (Nice):** in 2008 we had a first contact with Perferencemnt an SME specialized in web site referencing. In 2009 we started a collaboration within a research project aiming to optimize the position of results provided by search engines such as Google. Our contributions will concern: usage mining and experimentation (cf. section 7.1.2).
- **Coexel (Marseille):** Axis participated to "speed dating" meeting organized by the SCS cluster at Inria and aiming at matching SMEs problems with Inria research projects' solutions. With Coexel, a SME specialized in knowledge management we proposed the "Smart Docs santé" Project for labeling it by the cluster SCS. AxIS will be concerned by *usage analysis, personnalization and recommender systems*.
- **Ivès** (Biot, Grenoble) and **AccésCités**: accessible mobility, partner in the Pacalab proposal called TAC-VAL (cf. section 7.1.2).
- **Osmose** (Roquefort-les-Pins): energy, partner in the Pacalab proposal called ECOFFICES (cf. section 7.1.2).

Via two proposals to FUI9, we are involved with others SMEs :

- several SMEs (Numtech, Vincipark, Cityway, etc.) involved in the NAVTI proposal (labelled by the Movéo and System@tic clusters) related to multimodality and mobility
- Moviken (including SLE Sophia Antipolis) in e-transports proposal (labelled by the Advancity cluster) related to a Dlobal System of Distribution of transport tickets for tourism and dedicated to the department of *Alpes Martimes*.

7.2. National Initiatives

7.2.1. RNTL Project (2006-2009): EIFFEL

Participants: 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 in 2007 with industrial partners (Mondeca and Antidot (leader)) and academic partners (LIRMM, University of Paris X (Nanterre) and INRIA). This project concerns **Tourism and Semantic Web**.

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 was in charge of the sub-package entitled SP8 and of the defination of new paradigms dedicated to knowledge searching and visualizing, of the extrcation and exploitation of users' models and profiles from web logs.

The Eiffel project was completed on october 31st.

7.2.2. ANR Techlog Project (2008-2009): INTERMED

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 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.

This project concerns **participative democraty** and involved four natural territories: *Etang de Thau*, *le parc de la Camargue*, *le Parc de la Narbonnaise* et *le Parc de la Bretagne*)

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.

We pursued our work in Intermed till march to finish our commitments concerning technical clusters and usability analysis [54]: usage analysis,comparison of two doucments for supporting a debate, data collecting software implemented as a Web service and the transformation of Sommières data into the proposed format.

We then had an agreement with ANR to withdraw from the project.

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

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

The MIDAS project "Mining Data Streams", granted by ANR, has 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 new elements;
- The summaries must cover the whole stream and enable to build summaries of any past part of the history of a stream.

In 2009, we have coordinated and participated in the final version of a delivrable written by all partners on their work on the topic of "Summarizing data streams by means of patterns extraction" [52].

7.2.4. Reference National Center in Health

Participants: Bernard Senach, Brigitte Trousse.

AxIs has contributed to the proposal for the call for a national reference center in Health which involved CIU Santé, Pôle SCS, CHU Nice, Inria, etc. This proposal was accepted. AXIS contributions are related to its know-how on experimentations and knowledge discovery and management and also to a potential cooperations between the two living labs "ICT Usage lab" and the Limousin living lab.

7.2.5. EGC Association

AxIS members participated 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 as invited editor with O. Boussaid and P. Ganscarky has prepared a special issue of RNTI on *Fouille de données complexes: avancées récentes* (planned for the beginning of 2010).
- Y. Lechevallier, F. Masseglia and B. Trousse are members of the scientific committee of the workshop of *Fouille de données complexes* at EGC09 (January).
- A. Da Silva made a presentation on her thesis's results [48] at the meeting *Complexité due aux données multiples* of the EGC-FDC group (CNAM, Paris, 18 juin).

7.2.6. Other Collaborations

Academic Collaborations

- Telecom Paris Tech: we collaborated with Georges Hébrail (ENST) via the MIDAS project and the BiLab laboratory
- LIRMM: F. Masseglia collaborated with M. Teisseire, A. Laurent et P. Poncelet via the MIDAS project and C. Fiot, J. Marlier, B. Senach and B. Trousse collaborated with J. Sallantin via the Intermed project.
- CEMAGREF (Montpellier): B. Senach and B. Trousse collaborated with N. Ferrand and others colleagues via the Intermed project.
- CNAM: Y. lechevallier collaborated with G. Saporta for the supervision of the Charrad's thesis [36], [28] and with M-A. Aufaure in the context of the STIC France-Tunisia [30], [29]. F. de A. T. de Carvalho made an invited talk [49] in november at CNAM..
- Paris Descartes: M. Csernel collaborated regularly with Francois Patte on aspects concerning the Sanskrit.
- University of Bordeaux 1 and 2 (MAP laboratory): Y. Lechevallier collaborated with M. Chavent [37].
- University of Rennes: T. Despeyroux collaborated with M. Arnoux [26].
- INRETS: we collaborated in the context of a FUI9 proposal with G. Scemama, Farida Saad and Samuel Sellam related to the experimental aspects and user driven innovation of the NAVTI proposal (FUI9).

Industrial Collaborations

Via our two proposals to FUI9 (NAvTI and e-transports), we had new industrial contacts such as Continental Automotive (Rambouillet France SAS) and VEOLIA transports and also from Orange labs and various SMEs.

7.3. European Intiatives

7.3.1. FP 7 calls and ENoLL: European Network of Living Labs

Participants: Brigitte Trousse, Bernard Senach.

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

The living Lab “ICT Usage lab” has written a *Letter of Support* for **two proposals** from ENOLL community which have been accepted

- FP7 CIP: the EU-funded project Apollon starts november targeting cross-border networks of Living Labs to provide a model for successful Living Lab cooperation in several domains: Energy efficiency, E-health, eParticipation and eManufacturing.
- FP7 Transportation - Galileo. 2008. 4.3.3 Research and Innovation in GNSS: the GAINS project (Galileo Advanced INnovation Services).

As a member of the living lab “ICT Usage Lab” of the ENOLL network, Inria (AxIS) participated with other partners to three proposals for FP7 Call 5 ICT

- Objective 1.3 STREP ELLIOT “Experiential Living Labs for the Internet of Things” (Coordinator: TXT polymedia (Italy)) Proposition of an use case “Green services” in relation with environmental data issues from captors (watch, vehicules) or fixed city captors + planned collaboration with AtmoPACA (a non-profit making association in charge of pollution control)..
- Objective 1.6 STREP UDI House “User-Driven Innovation in Harmonized process Of generation and Usability of new Services and systems in Europe” (Coordinator: IBBT (Belgium). There are 10 partners including Yokosuko Telecom Research Park (Japon), INSIGHt (Taiwain).
- Objective 1.6 CSA FIREBALL “Future Internet Research and Experimentation By Adopting LivingLabs - Toward Smart Cities”. Coordinator: LTU-CDT (Sweden) 17 Partners. Mr Estrosi for the City of Nice and the”communauté urbaine de Nice Côte d’Azur” (NCA) wrote a *Letter of Support* for this CSA.

7.3.2. ERCIM

Participants: Brigitte Trousse, Bernard Senach.

In collaboration with J. Chailloux and C. Marchand (ERCIM), AxIS has worked on the proposal of a new ERCIM Working group which aims at focusing the research of the ERCIM institutions with Living Lab experiences on a series of activities for fostering the European research and development on usage analysis of new technologies, methodologies for co-design, user-driven innovation and users involvement.

A kickoff meeting is under preparation at the beginning of 2010.

7.3.3. Other Collaborations

- Italy, University of Napoli II: Y. lechevallier pursued his collaborationwith Prof. R. Verde, A. Irpino and A. Balzanella [27]. This year, our visitor F. de A. T. de Carvalho presented there a seminar about “Recent clustering methods for interval data”(november).
- Italy - INTERREG IV ALCOTRA: we have been involved related to a user centered approach of designing and evaluating new services in the context of the proposal called MyMed (resp. L. Liquori) which has been accepted and will star next year..
- Germany: B. Trousse is involved to the project "Core Technology Cluster" of the AII program "QUAERO" in the multimedia domain (ontology construction, personalization) http://www.quaero.org/.

7.4. International Initiatives

7.4.1. France-Brazil

Participants: Yves Lechevallier, Marc Csernel, Alzennyr Da Silva.

We continue our collaboration on clustering and web usage mining and we start a collaboration on social network data analysis with F.A.T. De Carvalho from Federal University of Pernambuco (Recife) and his team.

- A scientific project submitted by F. De Carvalho and Y. Lechevallier has been accepted by FACEPE and INRIA. The project starts on 01/2010 and ends on 12/2011. Researches and students are concerned by this project from AxIS and CIn-UFPE side. It aims at developing methods of clustering of relational data and social network data analysis tools.
- Y. Lechevallier has visited the UFPE. During its stay, in collaboration with F. de A.T. de Carvalho, they continued 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. Three complete papers on this subject have been published this year on "Pattern Recognition" [23], "Pattern Recognition Letters" [21], and "IEEE Transactions on System Man and Cybernetics, Part A" [22], journals.
- M. Csernel, F. de A. T. de Carvalho and K. Silva published a work [33] issued from her Master Thesis (about the adaptation of Fuzzy clustering on symbolic data, specially in presence of rules as background knowledge) in the proceedings of the 2009 International Joint Conference on Neural Networks published by IEEE computer society.
- During *L'année de la France au Brésil* the French-Brazilian workshop on data mining happenned in Recife form 5 to 7 May 2009 (http://www-roc.inria.fr/axis/modulad//Workshop_Franco_Bresilien/index.htm). The program of the workshop included 3 tutorials, 9 invited conferences and 16 scientific presentations. There were about 100 partitipants. F. de A. T. de Carvalho and Y. Lechevallier have participated actively on the organisation of this workshop.
- Y. Lechevallier gave an invited talk [53] on some clustering methods on dissimilarity or similarity matrices at the workshop franco-brésilien (Récife).

7.4.2. France-Morocco

Participants: Bernard Senach, Brigitte Trousse.

The project (2009-2013) with Casablanca University and ENSAM (Meknes) about the WRUM proposal (Web Redesign based on Usage Mining) accepted in may 2008 started this year for 4 years (2009-2013). This project received fundings from the Research and Education Ministery and is managed by the CSPT commission. In this project close to the three topics of AxIS, the chosen application domain is the e-learning.

Hicham Behja, Prof à ENSAM Meknès is in charge of the WRUM project with the support of Prof. A. Marzark (Univ Casablanca).

B. Senach and B. Trousse had their first visit to Casablanca in the context of the WRUM project. During their visit, Two meetings were organised: one at the University of Casablanca in order to have a demonstration of the e-learning platform used at OFPPT by M. Naamary (OFPPT) and a second one (october 26) with OFPPT (*Office de la Formation professionnelle et de la Promotion du travail*) at Casablanca in order to present the WRUM project and to identify future collaborations.

H. Behja and A. Marzark made two visits in France in the context of the WRUM project in order to work on the different thesis subjects: september 23 - october 2 and december 7-13.

7.4.3. France-Tunisia

Participant: Yves Lechevallier.

Y. Lechevallier is involved in the STIC program between France-Tuisia. With M-A. Aufaure (CNAM), they supervised 4 masters and 2 thesis (Riadi Lab, ENSI Tunis) in this project. These masters and thesis subjects are about web mining (usage, content and structure, using different methods) and ontology construction from heterogeneous sources. This year M. R. Haddad [30], [29] and M. Hadjouni [72], [71] participated actively to the themes of our STIC on personalization and an integration of space and accessibility in Web personalization with the co-supervision of H. Baazaoui Zghal, M.-A. Aufaure, C. Claramunt and Y. Lechevallier.

In this STIC project, we welcomed from ENSI (Tunisia) at Rocquencourt: M.R Haddad (January-February), M. Hadjouni (May to July, November), N. Ben Mustapha and I. Megdiche (May to July).

7.5. Exterior Research Visitors

F. de A. T. de Carvalho was a visiting professor at AxIS project during three months (October to December). He collaborated actively with Y. Lechevallier on partitional clustering algorithms based on multiple dissimilarity matrices in order to submit a paper to the COMPSTAT conference (to be held on Paris, August 22-27 2010) and an extended version to a international journal.

Additionally, he presented a communication about "Holt's exponential smoothing model for interval-valued time series" during the Workshop in Symbolic Data Analysis, 18-20 October 2009, Wienerwaldhof (Austria).

He was also invited by Prof. R. Verde (Second Naples University) from 10 to 15 of November 2009 where he presented a seminar about "Recent clustering methods for interval data".

He was invited by A. Napoli (Projet Orpailleur, Laboratoire LORIA, Nancy) where F. de A. T. de Carvalho presented a talk about "Recent Advances in Partitioning Clustering Algorithms for Interval-Valued Data" [50] during the First SaarLorLux Workshop on Systems Biology 2009 (https://wiki.mpi-inf.mpg.de/SysBio/Workshop2009/Schedule?action=print).

He also presented seminars on the CNAM, Paris [49] and Télécom Paris-Tech in the context of Bilab [51].

8. Dissemination

8.1. Promotion of the Scientific Communnity

8.1.1. Journals and Books

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

- B. Trousse was co-editor (with O. Boussaïd and P. Gancarsky) of the second special issue of the RNTI journal on *Fouille de données complexes*: avancées récentes":
- 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: Y. Lechevallier is one of the four editors.

AxIS members belong to editorial boards of three international journals and one special issue of a national journal:

- the Journal of Symbolic Data Analysis (JSDA) (Editor: E. Diday, electronic journal: Y. Lechevallier
- the Co-Design Journal (Editor: Janet McDonnel University of the Arts London - Publisher: Swets & Zeitlinger): B. Trousse
- IJDST - International Journal of Design Sciences & Technology (Editors R. besheshti & K. Zreik - Publisher: Europia productions),
- RIA Special Issue *Intelligence Artificielle et Web Intelligence* (Publisher, Hermes-Lavoisier) série RIA Vol. 23 N°1 (January-Février) Editions, editors-in-chief: Y. Demazeau and L. Vercouter: B. Trousse

AxIS members were reviewers for two journals and 1 book:

- DKE International Journal on Data Knowledge and Engineering (Elsevier): F. Masseglia
- TSI journal *Technique et Science informatiques* (Publisher, Hermes-Lavoisier) Special issue *L'informatique à l'interface de l'activité humaine et sociale*: B. Trousse
- AKDM, Advances in Knowledge Discovery and Management, post-proceedings of EGC'09 (Springer): F. Masseglia

8.1.2. National Conferences/Workshops

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

- IC'09: 20èmes journées francophones d'Ingénierie des Connaissances modèles, représentations, raisonnements, gestion et usages des connaissances, Tunisia, http://ic2009.inria.fr/: B. Trousse
- RàPC'09: 17ème atelier de Raisonnement à Partir de Cas: 29-30 juin 2009 à Paris. http://liris.cnrs.fr/~rapc/mediawiki/index.php/Rapc2009: B. Trousse
- EGC'09: 9èmes Journées Francophones *Extraction et Gestion des Connaissances* (Strasbourg, January 27-30): Y. Lechevallier, F. Masseglia, B. Trousse
- Ateliers EGC'09: Sophia-Antipolis, France, 26 January 2009
  - *Fouille de Données Complexes dans un processus d'extraction des connaissances*: Y. Lechevallier, F. Masseglia, B. Trousse
  - Fouille de données temporelles et analyse de flux de données: Y. Lechevallier, F. Masseglia.
  - Modélisation utilisateur et personnalisation d'interfaces Web: Y. Lechevallier
- H2PTM'09: 10th international conference on "Hypertexte et Hypermédia Produits, Outils et Méthodes", September 30, october 2nd 2009, Université Paris 8, France Université Paris 8: B. Trousse
- EPIQUE'09: Nice, September 28-30. B. Senach was also discutant for the symposium "Interactions and cooperations inside on-line communities and social networks".

8.1.3. International Conferences/Workshops

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

- CSCWD'09:13th International Conference on Computer Supported Cooperative Work in Design (Santiago, Chile, April 22-24, 2009): B. Trousse
- ICCR'09: 8th International Conference on Case-Based Reasoning, held in Seattle, Washington, USA from 20 July to 23 July 2009: B. Trousse
- WebCBR'09: ICCBR Workshop on Reasoning with People's Experiences on the Web, held in Seattle, Washington, USA on 21 to 23 July 2009: B. Trousse
- SYNASC'09, 11th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, Timisoara, Romania, September 26-29: B. Trousse
- ISICA'09, 4th International Symposium on Intelligent Computation and Applications (China, October 23-25 2009): F. Masseglia
- EPIA'09, Fourteenth Portuguese Conference on Artificial Intelligence (Aveiro, Portugal, October 12-15, 2009): F. Masseglia
- DS'09, Discovery Science (Porto, Portugal, 3-5 October 2009): F. Masseglia

8.1.4. Organization of Conferences / Workshops

- B. Trousse was *Publicity Chair for Europe* of the Fifteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD'09**), organised for the first time in Europe in France at Paris (June 28 - July 1, 2009). As the premier international conference on Data Mining and Knowledge Discovery, KDD 2009 provides a forum for academic researchers and industry and government innovators to share their results and experiences. Researchers and practitioners gather to present academic and industrial papers, panels, implemented software demos, posters, workshops, tutorials, and insights from the popular KDD Cup competition.

- B. Senach is member of the scientific committee and of the local organization committee of the **EPIQUE 2009** Colloquium which held in Nice (September 28-30).
- Y. Lechevallier is involved in the organisation of the 19th COMPSTAT symposium, **COMPSTAT 2010** (August 22nd-27th 2010), which is the conference on the Statistical Computing and sponsored by the European Regional Section of the IASC (International Association for Statistical Computing). IASC is a section of ISI (International Statistical Institute). COMPSTAT2010 is locally organized by the CNAM and the INRIA (AxIS).
- Y. Lechevallier with F. de A. T. De Carvalho have participated actively in the organisation of the **French-Brazilian Workshop on Data Mining** (may 5-7 may 2009) which held during *l'année de la France au Brésil* (França.Br 2009) with 100 participants.

8.1.5. Other Reviewing Activities

B. Trousse was involved as expert-reviewer in four different contexts::

- Call for proposals *Projets structurant* (with 3 themes) proposed by the scientific orientation and coordination commmittee *CCOS Lorrain* of a grouping of 4 universities + 4 EPST (including Inria) from the Region: review of a proposal on the *Industry of Knowledge* theme (June).
- Cluster ISLE: participation as expert to the scientific days of the cluster ISLE (27-28 october, Lyon) and review of the project "Web intelligence (november)
- ANR Technologies logicielles: participation as expert to the review organised for 4 ANR projects on "Technologies Logicielles 2006 and 2007" (Paris, Feb 4): CADI (Techlog 2007), SEPTIA (Techlog 2006, revue finale), Procogec (Techlog 2006) and Dynamo (Techlog 2007).
- Pacalabs call 1: review of one proposal related to the Pacalabs call 1

F. Masseglia was expert for the CONTINT call for project in 2009.

8.1.6. Activities of General Interest

- Regional Council: B. Trousse belongs as Inria representative to the expert committee of the Pacalabs Calls and to the strategic orientation committee of the Pacalabs instrument of the Regional Council.
- Living Lab "ICT Usage lab" and Campus STIC (cf. section 7.1.1):
  - B. Senach and B. Trousse have pursued their efforts this year in coordinating the "restructuration" of the living lab "ICT Usage Lab" (labelled by ENoLL in 2006). A coordination team for the living lab with 1 or 2 representative members of each current founding members (Orange labs, Inria, CSTB, UNSA) was created. Its activities were: a first definition of the main objectives, research activities and services of the living lab in relation to the campus STIC, the PACA clusters and main actors related to innovation, the identification of the main avaible infrastructures and the first projects and finally a draft of the governance model and business model
  - B. Trousse is the corresponding person with Servane Crave (Orange Labs) and Alain Zarli (CSTB) of the Living lab "ICT Usage Lab" inside the ENoLL network.
- KIC ICT Labs (EIT): AxIS (B. Trousse and B. Senach) participated to a local contribution with others EPIs (Inria Sophia Antipolis) for a proposal at the EIT call for KICs where Inria is involved at the national level. The proposal called *ICT labs* has been accepted.
- AGOS: T. Despeyroux is involved (30 %) as president of AGOS (Inria Works Council).
- Inria web site: B. Senach participated to the redesign of the Inria web site and to the service provider selection.
- French national "Science celebration": F. Masseglia gave three sessions of one hour to students of Audiberti High School of Antibes. The goal of these sessions was to explain the research work in Computer Science in general and more particularly in Data Mining.

- IEEE : B. Trousse is member of the technical committee on Computer Supported Work in Design in the context of the IEEE SMCS (Systems, Man & Cybernetics Society)

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 (UNSA, Univ Paris IX-Dauphine, CNAM, ENSAE):

- Master IFI (resp. B. Martin) at Polytech'Sophia: Lecture (6h) on *Data Mining: Sequences, Streams and Security*: F. Masseglia.
- 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 Ingénierie de la Statistique (resp: G. Saporta) of CNAM (12h) on *Méthodes neuronales*: Y. Lechevallier.
- ENSAE (*Ecole Nationale de la Statistique et de l'Administration Ecnoomique*): Tutorial (18h) on *"Data Mining"*: Y. Lechevallier.

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

one H.D.R defense this year:

- **F. Masseglia**, "Extraction de connaissances: l'évolution dans les données et dans les résultats de la fouille", [20] University of Nice Sophia Antipolis, november 27th.

Three Ph.D. defenses this year:

- **H. Behja**, (start: 2003), "Gestion de points de vues multiples dans l'analyse d'un observatoire sur le Web", [17] (director: A. Marzark - Univ. of Casablanca with the participation of B. Trousse).
- **A. Marascu**, (start: October 2005), "Extraction de Motifs Séquentiels dans les flux de données", [19], University of Nice-Sophia Antipolis (director: Y. Lechevallier, with the participation of F. Masseglia), septembre 14.
- **A. Da Silva**, (start: October 2005), "Modélisation de données agrégées ou complexes par l'approche symbolique, application au Web Usage Mining", [18], University of Paris IX Dauphine (directors: Edwin Diday and Y. Lechevallier).

Three PhD in progress: 2 (at inria) and 1 (under a France-Tunisia cooperation)

1. **E. Smirnova** (start: September 2008), "Mining social networks", University of Nice-Sophia Antipolis & INRIA, (co-directors: M. Rueher, B. Trousse). Let us note that E. Smirnova made an internship at Google (Paris) may-august.
2. **C. Zhang** (start: october 2008), "Mining data streams: clustering and pattern extraction", University of Nice-Sophia Antipolis & INRIA, (director: F. Masseglia).
3. **M. Charrad** (start: end of 2005), "Design and development a Web mining environment coupling content mining and usage mining (with biclustering technics), CNAM. and University La Manouba (Tunisia). Y. Lechevallier is co-superviser with G. Saporta (CNAM) and in *co-tutelle* with Prof M. Benhamed (ENSI, Tunisia).

In the context of the WRUM project, B. Trousse and B. Senach participate in the thesis committees of three Ph-D students:

- **M. Naamany** from the University of Casablanca supervised by A. Marzark.
- **N. Sael** from the University of Casablanca supervised by A. Marzark.
- **E Zemmouri** from the University of Fes and ENSAM (Meknès, Morocco) supervised by H. Behja and A. Marzark.

Let us note that B. Senach and B. Trousse have been also involved in the supervision of the PhD thesis of **Julie Marlier** via the Intermed project, from november 2008 until march 2009, thesis supervised by B. Conein (UNSA).and N. Ferrand (Cemagref Montpellier).

AxIS researchers were members of Ph.D. committees in 2009:

- **F. Masseglia**, H.D.R "Extraction de connaissances: l'évolution dans les données et dans les résultats de la fouille", University of Nice Sophia Antipolis, November 27th: Y. Lechevallier, F. de A. T. de Carvalho
- **A. Marascu**, PhD, "Extraction de motifs séquentiels dans les flux de données" (supervisors: Y. Lechevallier and F. Masseglia) University of Nice Sophia Antipolis, September: Y. Lechevallier, F. Masseglia, B. Trousse
- **M. Vincent Labbé**, PhD, "Modélisation et apprentissage des préférences appliqués à la recommandation dans les systèmes d'impression" (supervisors: B. Bouchon-Meunier and N. Labroche), LIP6, Paris, September: B. Trousse
- **A. Gomes Da Silva**, Ph.D, Analyse des données évolutives: application aux données d'usage du Web, September, University Paris-Dauphine: Y. Lechevallier, F. de A. T. de Carvalho
- **E. Prudhomme**, Ph.D, Représentation et fouille de données volumineuses, June, University of Lyon II: Y. Lechevallier
- **L. El Moubarki**, Ph.D, Validation d'un partitionnement: proposition d'une nouvelle approche fondée sur la décomposition en isolation-cohésion de la stabilité des classes d'une partition, December, University Paris-Dauphine: Y. Lechevallier

8.2.3. Internships

We welcomed six students this year:

- **C. Isai** (supervisor B. Trousse), AxIS Sophia Antipolis, Master 2, University of Timisoara, Roumania, structuration of annotations from debates of a group and implementation of an extended version of AWLH (until September 2009)
- **C. De Albuquerque Melo**, Master in Human-Machine Interaction, February-June on community detection in social networks.
- **4 students** from the STIC program Tunisia -Inria on spatial Web and personnalisation: M. R. Haddad (January-February) [30], [29], M. Hadjouni (May to July, November) [72], [71], N. Ben Mustapha and I. Megdiche (May to July).

8.3. Participation to Workshops, Conferences, Seminars, Invitations

Furthermore we attended the following conferences or workshops:

- the Lift Conference, Marseille, June.19-20: B. Senach (invited by the FING)
- Imagina, Monaco, February 19-20: B. Senach and B. Trousse (invited by the living lab *La fabrique du Futur*)
- KDD 2009, Paris, June 28 - July 1st: C. Zhang

- SAC 2009: C. Zhang (he made the presentation of [32]).
- EPIQUE 2009: B. Trousse
- World Usability Day 2009, CICA Sophia Antipolis, November 12: B. Senach, B. Trousse
- Workshop *Les Transports du futur, La mobilité des personnes* organised by ADEME Sophia Antipolis, December 8: B. Senach, B. Trousse
- Conference JFO 2009, Poitiers, December 3-4: Th. Despeyroux
- Seminars Jean-Pierre Fénelon on Data Analysis: Y. Lechevallier
- Bilab Seminars: Y. Lechevallier, A. Da Silva
- Réunion *Complexité due aux données multiples* du groupe EGC-FDC, CNAM, Paris, 18 juin: Y. Lechevallier

Y. Lechevallier made an invited talk at MSDM'09 (1st Meeting on Statistics and Data Mining, March 4-6, Hammamet, Tunisia) and at LFA 2009 [40] (Rencontres Francophones sur la Logique Floue et ses Applications, November 5-6, Annecy).

B. Trousse and R. Decorme (CSTB) made an invited talk on the ICT Usage Lab at the working days *Open Innovation & Living Labs*, april 6-7th Marseille organised by the PACA council..

9. Bibliography

Major publications by the team in recent years

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

[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] 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.

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

[5] 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.

[6] 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.

[7] A. DA SILVA, Y. LECHEVALLIER, F. A. T. 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.

[8] 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.

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

[10] 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.

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

[12] 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.

[13] 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.", 2007, DOI 10.1007/s10618-007-0080-z.

[14] 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.

[15] 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.

[16] 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

Doctoral Dissertations and Habilitation Theses

[17] H. BEHJA. *Plateforme objet d'évaluation orientée point de vue d'un système d'information*, University of Casablanca, Morocco, 2009, Ph. D. Thesis.

[18] A. DA SILVA. *Analysis of evolving data: application to the Web usage data (in French)*, University of Paris IX Dauphine and INRIA-France, 2009, Ph. D. Thesis.

[19] A. MARASCU. *Extraction de motifs séquentiels dans les flux de données*, University of Nice Sophia Antipolis, 2009, Ph. D. Thesis.

[20] F. MASSEGLIA. *Extraction de connaissances l'évolution dans les données et dans les résultats de la fouille*, Mémoire d'HDR, University of Nice Sophia-Antipolis, 2009, Ph. D. Thesis.

Articles in International Peer-Reviewed Journal

[21] F. A. T. DE CARVALHO, M. CSERNEL, Y. LECHEVALLIER. *Clustering constrained symbolic data*, in "Pattern Recognition Letters", vol. 30, n^{o} 11, 2009, p. 1037-1045 Br .

[22] F. A. T. DE CARVALHO, Y. LECHEVALLIER. *Dynamic Clustering of Interval-Valued Data Based on Adaptive Quadratic Distances*, in "IEEE Transactions on Systems, Man, and Cybernetics, Part A", vol. 39, n^{o} 6, 2009, p. 1295-1306 Br .

[23] F. A. T. DE CARVALHO, Y. LECHEVALLIER. *Partitional clustering algorithms for symbolic interval data based on single adaptive distances*, in "Pattern Recognition", vol. 42, n^{o} 7, 2009, p. 1223-1236 Fr Br .

[24] C. FIOT, F. MASSEGLIA, A. LAURENT, M. TEISSEIRE. *Evolution patterns and gradual trends*, in "International Journal of Intelligent Systems", vol. 24, n^{o} 10, 2009, p. 1013–1038.

[25] F. MASSEGLIA, P. PONCELET, M. TEISSEIRE. *Efficient mining of sequential patterns with time constraints: Reducing the combinations*, in "Expert Systems with Applications", vol. 36, n^{o} 2, 2009, p. 2677-2690.

International Peer-Reviewed Conference/Proceedings

[26] M. ARNOUX, T. DESPEYROUX. *Multi-représentation d'une ontologie : OWL, bases de données, systèmes de types et d'objets*, in "3ème édition des Journées Francophones sur les Ontologies (JFO'09), Poitiers, France", ACM, December 2009 Fr .

[27] A. BALZANELLA, Y. LECHEVALLIER, R. VERDE. *A new approach for clustering multiple streams of data*, in "Classification and Data Analysis 2009", S. INGRASSIA, R. ROCCI (editors), 2009, p. 417-420 Fr It .

[28] M. CHARRAD, Y. LECHEVALLIER, M. BEN AHMED, G. SAPORTA. *Block Clustering for Web Pages Categorization*, in "IDEAL", 2009, p. 260-267 Fr Tu .

[29] M. R. HADDAD, H. BAAZAOUI ZGHAL, M.-A. AUFAURE, C. CLARAMUNT, Y. LECHEVALLIER, H. BEN GHEZALA. *A User-oriented Accessibility Measure for Spatial Web Personalization*, in "GIIW/ECIR 2009 : Geographic Information on the Internet Workshop collocated with the 31st European Conference on Information Retrieval", 2009 Fr Tu .

[30] M. R. HADDAD, H. BAAZAOUI ZGHAL, M.-A. AUFAURE, C. CLARAMUNT, Y. LECHEVALLIER, H. BEN GHEZALA. *Towards an integration of space and accessibility in Web personalization*, in "W2GIS - Web and Wireless Geographical Information Systems", 2009 Fr Tu .

[31] A. MARASCU, F. MASSEGLIA. *A Multi-resolution Approach for Atypical Behaviour Mining*, in "Advances in Knowledge Discovery and Data Mining, 13th Pacific-Asia Conference, PAKDD 2009, Bangkok, Thailand, April 27-30", 2009, p. 899-906.

[32] A. MARASCU, F. MASSEGLIA. *Parameterless outlier detection in data streams*, in "ACM Symposium on Applied Computing (SAC), Honolulu, Hawaii, USA, March 9-12", 2009, p. 1491-1495.

[33] K. SILVA, F. A. T. DE CARVALHO, M. CSERNEL. *Clustering of Symbolic Data Using the Assignment-Prototype Algorithm*, in "IJCNN", 2009, p. 2936-2942 Fr Br .

[34] G. SINGH, F. MASSEGLIA, C. FIOT, A. MARASCU, P. PONCELET. *Data Mining for Intrusion Detection: From Outliers to True Intrusions*, in "Advances in Knowledge Discovery and Data Mining, 13th Pacific-Asia Conference, PAKDD 2009, Bangkok, Thailand, April 27-30", 2009, p. 891-898.

[35] W. WANG, F. MASSEGLIA, T. GUYET, R. QUINIOU, M.-O. CORDIER. *A general framework for adaptive and online detection of web attacks*, in "the 18th International Conference on World Wide Web, WWW 2009, Madrid, Spain, April 20-24", 2009, p. 1141-1142.

National Peer-Reviewed Conference/Proceedings

[36] M. CHARRAD, Y. LECHEVALLIER, G. SAPORTA, M. BEN AHMED. *Détermination du nombre des classes dans l'algorithme CROKI de classification croisée*, in "Extraction et gestion des connaissances (EGC'2009), Actes des 9èmes journées Extraction et Gestion des Connaissances, Strasbourg, France, 27 au 30 janvier", 2009, p. 447-448 Fr .

[37] M. CHAVENT, Y. LECHEVALLIER, K. PETIT, F. VERNIER. *Classification divisive sous contraintes. Application au bassin de la Charente*, in "41èmes Journées de Statistique de la SFdS", 2009, p. 260-267 Fr .

[38] A. DA SILVA, Y. LECHEVALLIER, F. A. T. DE CARVALHO. *Approche pour le suivi des changements sur les données évolutives*, in "Journées de la Société Francophone de Classification", G. D'AUBIGNY (editor), 2009, p. 129-132 Fr .

[39] A. DA SILVA, Y. LECHEVALLIER, F. A. T. DE CARVALHO. *Vers la simulation et la détection des changements des données évolutives d'usage du Web*, in "Extraction et gestion des connaissances (EGC'2009), Actes des 9èmes journées Extraction et Gestion des Connaissances, Strasbourg, France, 27 au 30 janvier", 2009, p. 453-454 Fr Br .

[40] Y. LECHEVALLIER. *L'imprécision et la classification automatique*, in "Rencontres francophones sur la logique floue et ses applications", Cépadués-éditions, 2009, p. 11-13 Fr .

[41] A. MARASCU, F. MASSEGLIA. *Détection d'enregistrements atypiques dans un flot de données: une approche multi-résolution*, in "Extraction et gestion des connaissances (EGC'2009), Actes des 9èmes journées Extraction et Gestion des Connaissances, Strasbourg, France, 27 au 30 janvier", 2009, p. 455-456.

[42] G. SINGH, F. MASSEGLIA, C. FIOT, A. MARASCU, P. PONCELET. *Collaborative Outlier Mining for Intrusion Detection*, in "Extraction et gestion des connaissances (EGC'2009), Actes des 9èmes journées Extraction et Gestion des Connaissances, Strasbourg, France, 27 au 30 janvier", 2009, p. 313-324.

[43] N. VERMA, F. TROUSSET, P. PONCELET, F. MASSEGLIA. *Détection d'intrusions dans un environnement collaboratif sécurisé*, in "Extraction et gestion des connaissances (EGC'2009), Actes des 9èmes journées Extraction et Gestion des Connaissances, Strasbourg, France, 27 au 30 janvier", 2009, p. 301-312.

[44] W. WANG, T. GUYET, R. QUINIOU, M.-O. CORDIER, F. MASSEGLIA. *Online and adaptive anomaly Detection: detecting intrusions in unlabelled audit data streams*, in "Extraction et gestion des connaissances (EGC'2009), Actes des 9èmes journées Extraction et Gestion des Connaissances, Strasbourg, France, 27 au 30 janvier", 2009, p. 457-458.

Scientific Books (or Scientific Book chapters)

[45] A. DA SILVA, Y. LECHEVALLIER, F. ROSSI, F. A. T. DE CARVALHO. *Clustering Dynamic Web Usage Data*, in "Innovative Applications in Data Mining", N. NEDJAH, L. DE MACEDO MOURELLE, J. KACPRZYK (editors), Studies in Computational Intelligence, vol. 169/2009, chap. 4, Springer, 2009, p. 71–82 Fr Br .

[46] A. DA SILVA, Y. LECHEVALLIER, F. A. T. DE CARVALHO. *Comparing Clustering on Symbolic Data*, in "Intelligent Text Categorization and Clustering", Springer, 2009, p. 81-94 Fr Br .

Research Reports

[47] A. DA SILVA, Y. LECHEVALLIER. *Axis Tool for Web Usage Evolving Data Analysis (ATWEDA)*, INRIA-France, 2009.

Other Publications

[48] A. DA SILVA, Y. LECHEVALLIER. *Analyse exploratoire des indices pour la détermination du bon nombre de clusters : application aux données évolutives*, 2009, Groupe de Travail EGC sur la Fouille de Données Complexes (GT EGC-FDC) : Complexité liée aux données multiples. Br Fr .

[49] F. A. T. DE CARVALHO. *Nouveaux Modèles de Régression Linéaire pour les Variables de Type Intervalle*, 2009, Invited talk, 18 november,CNAM, Paris, France.

[50] F. A. T. DE CARVALHO. *Recent Advances in Partitioning Clustering Algorithms for Interval-Valued Data*, 2009, First SaarLorLux Workshop on Systems Biology 2009, Invited talk.

[51] F. A. T. DE CARVALHO. *Une approche hybride pour les prévisions sous forme d'intervalle de séries temporelles d'intervalles*, 2009, BiLab'seminar at Telecom ParisTech, 10 December, Paris, France, Invited talk.

[52] GET-ENST, LIRMM, CEREGMIA, INRIA, LGI2P, EDF R&D, FT R&D. *Summarizing data streams by means of patterns extraction*, n^{o} Tâche 1.3, 2009.

[53] Y. LECHEVALLIER. *Some clustering methods on dissimilarity or similarity matrices: Uncovering clusters in WEB Content, Structure and Usage*, 2009, French-Brezilian Workshop on data Mining 2009, Invited talk.

[54] B. TROUSSE, B. SENACH. *INTERMED: Rapport d'activités pour la période 1/3/2008-1/3/2009 - AxIS Inria Sophia Antipolis*, 2009 Fr .

[55] B. TROUSSE, B. SENACH. *Living Labs as an Open Innovation Ecosystem: Needs & Requirements for Collaboration Services viewed by the Living lab "ICT Usage Lab"*, 2009, ICE2009 workshop COIN4LL, june 24th, Invited talk Fr .

References in notes

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

[57] K. BALOG, T. BOGERS, L. AZZOPARDI, M. DE RIJKE, A. VAN DEN BOSCH. *Broad expertise retrieval in sparse data environments*, in "SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval", ACM, 2007, p. 551–558.

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

[59] D. BLEI, A. NG, M. JORDAN. *Latent Dirichlet Allocation*, in "Journal of Machine Learning Research", vol. 3, 2002, 2003.

[60] 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.

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

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

[63] 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.

[64] 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.

[65] F. A. T. 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.

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

[67] T. DESPEYROUX. *XML based information systems and formal semantics of programming languages*, Springer, Advanced Information and Knowledge Processing, April 2010, to appear Fr .

[68] 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.

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

[70] J. GRIFFITHS, F. JOHNSON, R. HARTLEY. *User satisfaction as a measure of system performance*, in "Journal of Librarianship and Information Science", vol. 39, n^{o} 3, 2007.

[71] M. HADJOUNI, M. R. HADDAD, H. BAAZAOUI ZGHAL, M.-A. AUFAURE, H. BEN GHEZALA. *A spatially enhanced web personalization approach*, in "ICWIT/ACM SIGAPP : In Second International Conference on Web and Information Technologies, in cooperation with ACM SIGAPP", 2009 Fr Tu .

[72] M. HADJOUNI, M. R. HADDAD, H. BAAZAOUI ZGHAL, M.-A. AUFAURE, H. BEN GHEZALA. *Personalized Information Retrieval Approach*, in "WISM/CAISE'09 : Sixth International Workshop on Web Information Systems Modeling, held in conjunction with CAISE'09", 2009 Fr Tu .

[73] T. HOFMANN. *Probabilistic Latent Semantic Analysis*, in "SIGIR", 1999, p. 50–57.

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

[75] H. LIU, H. MOTODA. *Computational Methods of Feature Selection (Chapman & Hall/Crc Data Mining and Knowledge Discovery Series)*, Chapman & Hall/CRC, 2007.

[76] A. MARASCU, F. MASSEGLIA. *Mining sequential patterns from data streams: a centroid approach*, in "J. Intell. Inf. Syst.", vol. 27, n^{o} 3, 2006, p. 291-307.

[77] L. C. MOLINA, L. BELANCHE, A. NEBOT. *Feature selection algorithms: a survey and experimental evaluation*, in "ICDM'02. Proceedings. 2002 IEEE International Conference on Data Mining", 2002, p. 306-313.

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

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

[80] 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.

[81] 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.

[82] 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.

[83] 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.

[84] 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)", September 2007, http://dx.doi.org/10.2390/biecoll-wsom2007-144, ISBN: 978-3-00-022473-7.

[85] 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.

[86] P. SERDYUKOV, D. HIEMSTRA. *Modeling Documents as Mixtures of Persons for Expert Finding*, in "ECIR", 2008, p. 309-320.

[87] D. TANASA. *Web Usage Mining: Contributions to Intersites Logs Preprocessing and Sequential Pattern Extraction with Low Support*, University of Nice Sophia Antipolis, june 2005, Ph. D. Thesis.

[88] 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, S. E. G. GONZALEZ CISARO, X. DANIEL HUGO (editors), chap. 3, Information Science Reference, 2007, p. 37-64.

[89] 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)", September 2007, http://dx.doi.org/10.2390/biecoll-wsom2007-139, ISBN: 978-3-00-022473-7.