IN PARTNERSHIP WITH:
CNRS

Université de Lorraine

Activity Report 2014

Project-Team ORPAILLEUR

Knowledge discovery, knowledge representation, reasonning

IN COLLABORATION WITH: Laboratoire lorrain de recherche en informatique et ses applications (LORIA)

RESEARCH CENTER
Nancy - Grand Est

THEME
Data and Knowledge Representation and Processing

Table of contents

1. Members **1**
2. Overall Objectives **2**
3. Research Program **3**
  3.1. From KDD to KDDK 3
  3.2. Knowledge Discovery guided by Domain Knowledge 3
  3.3. Text Mining 4
  3.4. Knowledge Systems and Semantic Web 4
4. Application Domains **5**
  4.1. Biology and Chemistry 5
  4.2. Medicine 5
  4.3. Cooking 6
  4.4. Agronomy 6
5. New Software and Platforms **7**
  5.1. Generic Symbolic KDD Systems 7
    5.1.1. The Coron Platform 7
    5.1.2. Orion: Skycube Computation Software 7
  5.2. Stochastic systems for knowledge discovery and simulation 7
    5.2.1. The CarottAge System 7
    5.2.2. The ARPEnTAge System 7
  5.3. KDD in Systems Biology 8
    5.3.1. IntelliGO Online 8
    5.3.2. WAFOBI: KNIME Nodes for Relational Mining of Biological Data 8
    5.3.3. MOdel-driven Data Integration for Mining (MODIM) 8
  5.4. Knowledge-Based Systems and Semantic Web Systems 9
    5.4.1. The Kasimir System for Decision Knowledge Management 9
    5.4.2. Taaable: a System for Retrieving and Creating New Cooking Recipes by Adaptation 9
    5.4.3. Tuuurbine: a Generic Ontology Guided Case-Based Inference Engine 9
    5.4.4. BeGoood: a Generic System for Managing Non-Regression Tests on Knowledge Bases 10
    5.4.5. Revisor: a Library of Revision Operators and Revision-Based Adaptation Operators 10
6. New Results **10**
  6.1. Highlights of the Year 10
  6.2. The Mining of Complex Data 11
    6.2.1. FCA and Variations: RCA, Pattern Structures and Biclustering 11
    6.2.2. Sequence Mining 12
    6.2.3. Mining and Understanding Healthcare Trajectories 12
    6.2.4. Video Game Analytics 12
    6.2.5. KDDK in Text Mining 12
  6.3. KDDK in Life Sciences 13
    6.3.1. Inductive Logic Programming for Mining Linked Open Data 13
    6.3.2. Analysis of biomedical data annotated with ontologies 13
  6.4. Structural Systems Biology 14
    6.4.1. The Hex Protein Docking Program 14
    6.4.2. KBDOCK: Protein Docking Using Knowledge-Based Approaches 14
    6.4.3. Kpax: A New Algorithm for Multiple Flexible Protein Structure Alignments 14
    6.4.4. Polypharmacology: Developing New Uses for Old Drugs 15
  6.5. Around the Taaable research project 15
  6.6. Some Results in Graph Theory 15
    6.6.1. Structural and extremal graph theory 16
    6.6.2. Graph theory and other fields 16

6.6.3. Algorithmic Graph Theory and Clustering 16
6.6.4. Structural and Algebraic Graph Theory 17
7. Bilateral Contracts and Grants with Industry 17
7.1. The BioIntelligence Project 17
7.2. The Quaero Project 17
8. Partnerships and Cooperations 18
8.1. International Initiatives 18
8.1.1. Participation In International Programs 18
8.1.1.1. Ciência Sem Fronteiras 18
8.1.1.2. Pronex Brasilia 18
8.1.2. Inria Associate Team: Snowflake 18
8.1.3. Explorer Programme 19
8.1.4. Miscellaneous 19
8.2. National Initiatives 20
8.2.1. ANR 20
8.2.1.1. HEREDIA 20
8.2.1.2. Hybride 20
8.2.1.3. ISTEX 20
8.2.1.4. Kolflow 20
8.2.1.5. PEPSI: Polynomial Expansions of Protein Structures and Interactions 21
8.2.1.6. Termith 21
8.2.1.7. Trajcan: a study of patient care trajectories 21
8.2.2. Other National Initiatives and Collaborations 21
8.2.2.1. Towards the discovery of new nonribosomal peptides and synthetases 21
8.2.2.2. FUI Poqemon 22
8.3. Regional Initiatives 22
8.3.1. Le Bois Santé (LBS) 22
8.3.2. PEPS Mirabelle EXPLOD-Biomed 22
8.3.3. Hydreos 22
8.3.4. Contrat Plan État Région” (CPER) 22
9. Dissemination 23
9.1. Promoting Scientific Activities 23
9.2. Teaching – Supervision – Committees 24
10. Bibliography 24

Project-Team ORPAILLEUR

Keywords: Knowledge Discovery, Data Mining, Ontologies, Knowledge Representation, Reasoning

Creation of the Project-Team: 2008 January 01.

1. Members

Research Scientists

Amedeo Napoli [Team leader, CNRS, Senior Researcher, HdR]
Marie-Dominique Devignes [CNRS, Researcher, HdR]
Bernard Maigret [CNRS, Senior Researcher, HdR]
Chedy Raïssi [Inria, Researcher]
David Ritchie [Inria, Senior Researcher, HdR]
Jean-Sébastien Sereni [CNRS, Researcher]
Yannick Toussaint [Inria, Researcher, HdR]

Faculty Members

Miguel Couceiro [Univ. Lorraine, Professor, from Sep. 2014, HdR]
Adrien Coulet [Univ. Lorraine, Associate Professor]
Nicolas Jay [Univ. Lorraine, Professor, HdR]
Florence Le Ber [ENGEES Strasbourg, Professor (Team Associate), HdR]
Jean Lieber [Univ. Lorraine, Associate Professor, HdR]
Jean-François Mari [Univ. Lorraine, Professor, HdR]
Emmanuel Nauer [Univ. Lorraine, Associate Professor]
Malika Smaïl-Tabbone [Univ. Lorraine, Associate Professor, HdR]
Mario Valencia [Univ. Paris XIII, Associate Professor, HdR]

Engineers

Jérémie Bourseau [ADT Inria and CNRS]
Sami Ghadfi [Inria, Hybride, from Nov. 2014]
Luis-Felipe Melo [Univ. Lorraine]
Matthieu Osmuk [Inria, BioIntelligence and Poqemon]
Nicolas Pépin-Hermann [Inria, Biointelligence, until Feb. 2014]
Mickaël Zehren [Inria, Poqemon, from Nov. 2014]

PhD Students

Mehwish Alam [Univ. Lorraine]
Seyed Ziaeddin Alborzi [Inria, from Oct. 2014]
Aleksey Buzmakov [Univ. Lorraine]
Victor Codocedo [Univ. Lorraine]
Valmi Dufour-Lussier [Univ. Lorraine, until Oct. 2014]
Elias Egho [Univ. Lorraine, until Aug. 2014]
Emmanuelle Gaillard [Univ. Lorraine]
Gabin Personeni [Univ. Lorraine]
Sébastien Da Silva [Univ. Lorraine]
Mohsen Sayed [Inria]
My Thao Tang [Inria]

Post-Doctoral Fellows

Emmanuel Bresso [CNRS, Jan. – Feb. 2014]
Melisachew Chekol [Inria, until Apr. 2014, BioIntelligence]
Alice Hermann [Univ. Lorraine, until Oct. 2014]

Ioanna Lykourentzou [Univ. Luxembourg, until Apr. 2014]
Olfa Makkaoui [Inria, from Nov. 2014]

Administrative Assistants

Antoinette Courrier [CNRS]
Emmanuelle Deschamps [Inria]
Sylvie Musilli [Univ. Lorraine]

Others

Nicolas Bourseau [Inria, stagiaire IUT, from Mar. until Jun. 2014]
Quentin Brabant [Univ. Lorraine, stagiaire M1, from Jun. until Jul. 2014]
Rémi de Joannis de Verclos [ENS Lyon, stagiaire M2, from Feb. until Jul. 2014]
Nicolas Denis [stagiaire M2, from Feb. until Aug. 2014]
Ghania Khensous [Univ. Algérie, visitor, until Jun 2014]
Alex Herve [CHU Nancy, stagiaire, from Apr. until Jul 2014]
Juliette Krebs [Inria, stagiaire L3, from May until Jul. 2014]
Artuur Leeuwenberg [Inria, stagiaire M1, from Jun. until Jul. 2014]
Jimmy Nex [Inria, stagiaire L3, from May 2014 until Aug. 2014]
William Philbert [Univ. Lorraine, stagiaire M2, from Apr. until Sep. 2014]
Guillaume Pierson [Univ. Lorraine, stagiaire M1, from Apr. until Jun. 2014]
Justine Reynaud [Univ. Lorraine, stagiaire M2, from Feb. until Jun. 2014]
Anastasia Shimorina [Univ. Lorraine, stagiaire M1, from Jun. until Aug. 2014]
Sara Van de Moosdijk [Inria, stagiaire M2, from Feb. until Jun. 2014]
Daniel Vantroys [ENS Cachan, stagiaire, from Sep. 2014]
Sylvain Vissiere Guerinet [INRA, stagiaire, from Jun. until Aug. 2014]
Jan Volec [Univ. Lorraine, visitor, until Apr. 2014]

2. Overall Objectives

2.1. Introduction

Knowledge discovery in databases (KDD) consists in processing large volumes of data in order to discover knowledge units that are significant and reusable. Assimilating knowledge units to gold nuggets, and databases to lands or rivers to be explored, the KDD process can be likened to the process of searching for gold. This explains the name of the research team: in French “orpailleur” denotes a person who is searching for gold in rivers or mountains. The KDD process is based on three main operations: data preparation, data mining and interpretation of the extracted units as knowledge units. Moreover, the KDD process is iterative, interactive, and generally controlled by an expert of the data domain, called the analyst. The analyst selects and interprets a subset of the extracted units for obtaining knowledge units having a certain plausibility.

As a person searching for gold may have a certain experience about the task and the location, the analyst may use general and domain knowledge for improving the whole KDD process. Accordingly, the KDD process may be related to domain ontologies (or knowledge bases) relative to the domain of data for implementing *knowledge discovery guided by domain knowledge* or KDDK. In the KDDK process, the extracted units have “a life” after the interpretation step: they are represented as knowledge units using a knowledge representation formalism and integrated within an ontology to be reused for problem-solving needs. In this way, knowledge discovery extends and updates existing ontologies, reifying the complementarity of knowledge discovery and knowledge representation.

3. Research Program

3.1. From KDD to KDDK

Keywords: knowledge discovery in databases, knowledge discovery in databases guided by domain knowledge, data mining

Knowledge discovery in databases is a process for extracting from large databases knowledge units that can be interpreted and reused. From an operational point of view, a KDD system includes databases, data mining modules, and interfaces for interactions, e.g. editing and visualization. The KDD process is based on three main operations: selection and preparation of the data, data mining, and finally interpretation of the extracted units.

The process of "knowledge discovery in databases guided by domain knowledge" extends the KDD cycle with a fourth step, where extracted units are represented within a knowledge base to be reused. The KDDK process –as implemented in the research work of the Orpailleur team– is based on *data mining methods* that are either symbolic or numerical:

- Symbolic methods are based on frequent itemsets search, association rule extraction, Formal Concept Analysis and extensions [113].
- Numerical methods are based on higher order stochastic models, namely second-order Hidden Markov Models (HMM2) and Hidden Markov fields (HMRF), which are especially designed for an efficient modeling of space and time [12].

The principle summarizing KDDK can be understood as a process going from complex data to knowledge units being guided by domain knowledge. Two original aspects can be underlined: (i) the knowledge discovery process is guided by domain knowledge at each step of the process, and (ii) the extracted units are embedded within knowledge-based systems for problem solving purposes.

One main operation in the research work of Orpailleur on KDDK is *classification*, which is a polymorphic process involved in modeling, mining, representing, and reasoning tasks. Moreover, the KDDK process is intended to feed knowledge-based systems working in application domains, e.g. agronomy, biology, chemistry, cooking and medicine, and also in the context of semantic web, text mining, information retrieval, and ontology engineering.

3.2. Knowledge Discovery guided by Domain Knowledge

Keywords: knowledge discovery, data mining, formal concept analysis, classification, frequent itemset search, association rule extraction, second-order Hidden Markov Models

Classification problems can be formalized by means of a class of objects (or individuals), a class of attributes (or properties), and a binary correspondence between the two classes, indicating for each individual-property pair whether the property applies to the individual or not. The properties may be features that are present or absent, or the values of a property that have been transformed into binary variables. Formal Concept Analysis (FCA) relies on the analysis of such binary tables and may be considered as a symbolic data mining technique to be used for extracting a set of formal concepts then organized within a concept lattice [113] (concept lattices are also known as "Galois lattices" [103]).

In parallel, the search for frequent itemsets and the extraction of association rules are well-known symbolic data mining methods, related to FCA (actually searching for frequent itemsets can be understood as traversing a concept lattice). Both processes usually produce a large number of items and rules, leading to the associated problems of "mining the sets of extracted items and rules". Some subsets of itemsets, e.g. frequent closed itemsets (FCIs), allow to find interesting subsets of association rules, e.g. informative association rules. This is why several algorithms are needed for mining data depending on specific applications [45].

Among useful patterns extracted from a database, frequent itemsets are usually thought to unfold “regularities” in the data, i.e. they are the witnesses of recurrent phenomena and they are consistent with the expectations of the domain experts. In some situations however, it may be interesting to search for “rare” itemsets, i.e. itemsets that do not occur frequently in the data (contrasting frequent itemsets). These correspond to unexpected phenomena, possibly contradicting beliefs in the domain. In this way, rare itemsets are related to “exceptions” and thus may convey information of high interest for experts in domains such as biology or medicine.

From the numerical point of view, a Hidden Markov Model (HMM2) is a stochastic process aimed at extracting and modeling a sequence of stationary distributions of events. Such models can be used for data mining purposes, especially for spatial and temporal data as they show good capabilities to locate patterns both in time and space domains.

Moreover, stochastic models have been designed to mine temporal sequences having a spatial dimension, for example the succession of land uses in a territory. One main Markovian assumption states that the temporal event succession in a given place depends only on the temporal event successions in neighboring points. By means of stochastic models such as hierarchical hidden Markov models and Markov random fields, it is possible to perform an unsupervised clustering of a spatial territory for discovering “patches” characterized by time and space regularities in their temporal successions.

3.3. Text Mining

Keywords: knowledge discovery form large collection of texts, text mining, information extraction, document annotation, ontologies

The objective of a text mining process is to extract useful knowledge units from large collections of texts [110]. The text mining process shows specific characteristics due to the fact that texts are complex objects written in natural language. The information in a text is expressed in an informal way, following linguistic rules, making text mining a particular task. To avoid information dispersion, a text mining process has to take into account –as much as possible– paraphrases, ambiguities, specialized vocabulary and terminology. This is why the preparation of texts for text mining is usually dependent on linguistic resources and methods.

From a KDDK perspective, text mining is aimed at extracting “interesting units” (nouns and relations) from texts with the help of domain knowledge encoded within an ontology (also useful for text annotation). Text mining is especially useful in the context of semantic web for ontology engineering [105]. In the Orpailleur team, the focus is put on the mining of real-world texts in application domains such as biology and medicine, using mainly symbolic data mining methods, and especially Formal Concept Analysis. Accordingly, the text mining process may be involved in a loop used to enrich and to extend linguistic resources. In turn, linguistic and ontological resources can be exploited to guide a “knowledge-based text mining process”.

3.4. Knowledge Systems and Semantic Web

Keywords: knowledge representation, ontology, description logics, classification-based reasoning, case-based reasoning, semantic web, information retrieval

Usually, people try to take advantage of the web by searching for information (navigation, exploration), and by querying documents using search engines (information retrieval). Then people try to analyze the obtained results, a task that may be difficult and tedious. Semantic web is an attempt for guiding search for information with the help of software agents, that are in charge of asking questions, searching for answers, classifying and interpreting the answers. However, a software agent may be able to read, understand, and manipulate information on the web, if and only if the knowledge necessary for achieving those tasks is available, and this is why ontologies are of main importance. Thus, there is a need for knowledge representation languages for annotating documents, describing the content of documents and giving a semantics to this content.

In particular, the knowledge representation language used for designing ontologies is the OWL language, which is based on description logics (DLs [100]). In OWL, knowledge units are represented within concepts (or classes), with attributes (properties of concepts, or relations, or roles), and individuals. The hierarchical organization of concepts (and relations) relies on a subsumption relation (i.e. a partial ordering).

The inference services are based on subsumption, concept and individual classification, two tasks related to “classification-based reasoning”. Furthermore, classification-based reasoning can be extended into case-based reasoning (CBR), which relies on three main operations: retrieval, adaptation, and memorization. Given a target problem, retrieval consists in searching for a source (memorized) problem similar to the target problem. Then, the solution of the source problem is adapted to fulfill the constraints attached to the target problem, and possibly memorized for further reuse.

4. Application Domains

4.1. Biology and Chemistry

Participants: Mehwish Alam, Aleksey Buzmakov, Adrien Coulet, Marie-Dominique Devignes, Elias Egho, Nicolas Jay, Bernard Maigret, Amedeo Napoli, Nicolas Pépin-Hermann, Gabin Personeni, David Ritchie, Mohsen Sayed, Malika Smaïl-Tabbone, Yannick Toussaint.

Keywords: knowledge discovery in life sciences, bioinformatics, biology, chemistry, genomics

One major application domain which is currently investigated by the Orpailleur team is related to life sciences, with particular emphasis on biology, medicine, and chemistry. The understanding of biological systems provides complex problems for computer scientists, and the developed solutions bring new research ideas or possibilities for biologists and for computer scientists as well. Accordingly, the Orpailleur team includes biologists, chemists, and a physician, making Orpailleur a very original EPI at Inria. Indeed, the interactions between researchers in biology and researchers in computer science improve not only knowledge about systems in biology, chemistry, and medicine, but knowledge about computer science as well.

Knowledge discovery is gaining more and more interest and importance in life sciences for mining either homogeneous databases such as protein sequences and structures, or heterogeneous databases for discovering interactions between genes and environment, or between genetic and phenotypic data, especially for public health and pharmacogenomics domains. The latter case appears to be one main challenge in knowledge discovery in biology and involves knowledge discovery from complex data depending on domain knowledge.

On the same line as biological data, chemical data are presenting important challenges w.r.t. knowledge discovery, for example for mining collections of molecular structures and collections of chemical reactions in organic chemistry. The mining of such collections is an important task for various reasons among which the challenge of graph mining and the industrial needs (especially in drug design, pharmacology and toxicology). Molecules and chemical reactions are complex data that can be modeled as undirected labeled graphs. One objective for guiding computer-based synthesis in organic chemistry is to discover general synthesis methods (i.e. kinds of “meta-reactions”) from currently available chemical reaction databases for designing generic and reusable synthesis plans.

Graph mining methods may play an important role in this framework as illustrated in [125], but Formal Concept Analysis (FCA) can also be used in an efficient and well-founded way [101]. Combining supervised methods –with a training sets where objects are tagged– and unsupervised methods, “jumping emerging patterns” can be detected that characterize classes of interest, e.g. toxic molecules or inhibitors. Then, a hybrid classification method based on FCA can be used for building a concept lattice where some of the concepts can be used as reference classes for classifying unknown objects, for recognition and prediction tasks. Graph mining in the framework of FCA is a very important task on which we are actively working, whose results can be transferred to text mining as well.

4.2. Medicine

Participants: Aleksey Buzmakov, Adrien Coulet, Elias Egho, Nicolas Jay, Jean Lieber, Amedeo Napoli, Matthieu Osmuk, Chedy Raïssi, Yannick Toussaint, Mickaël Zehren.

Keywords: knowledge representation, description logics, classification-based reasoning, case-based reasoning, semantic web, formal concept analysis, sequence mining, text mining

We are working on several applications in medicine, mainly in knowledge management and analysis of patient trajectories as sequences. In the first case, the Kasimir research project is about decision support and knowledge management for the treatment of cancer. This is a multidisciplinary research project in which participate researchers in computer science (Orpailleur), experts in oncology ("Institut de Cancérologie de Lorraine Alexis Vautrin" in Vandœuvre-lès-Nancy), Oncolor (a healthcare network in Lorraine involved in oncology), and A2Zi (a company working in Web technologies and involved in several projects in the medical informatics domain, http://www.a2zi.fr/). For a given cancer localization, a treatment is based on a protocol, which is applied in 70% of the cases and provides a treatment. The 30% remaining cases are "out of the protocol", e.g. contraindication, treatment impossibility, etc. and the protocol should be adapted, based on discussions among specialists. This adaptation process is modeled in Kasimir thanks to CBR, where the semantic Web technologies are used and adapted in the Kasimir project for several years.

Another work is in concern with the analysis of patient trajectories, i.e. the "path" of a patient during illness (chronic illnesses and cancer), considered as sequences. It is important to understand these sequence data and temporal data mining methods are good candidate tools for that. However, these methods should be adapted for addressing the complex nature of medical events. Thus, there is an ongoing work on the analysis of trajectories with different levels of granularity and w.r.t. external domain ontologies. In addition, it is also important to be able to compare and classify trajectories according to their content. This is why there is also a work on the definition of a similarity measure able to take into account the complex nature of trajectories and that can be efficiently implemented for allowing quick and reliable classifications.

4.3. Cooking

Participants: Valmi Dufour-Lussier, Emmanuelle Gaillard, Florence Le Ber, Jean Lieber, Amedeo Napoli, Emmanuel Nauer.

Keywords: cooking, knowledge representation, knowledge discovery, case-based reasoning, semantic wiki

The origin of the Taaable project is the Computer Cooking Contest (CCC). A contestant to CCC is a system that answers queries about recipes, using a recipe base; if no recipe exactly matches the query, then the system adapts another recipe. Taaable is a case-based reasoning system based on various technologies from semantic web, knowledge discovery, knowledge representation and reasoning. From a research viewpoint the system enables to test scientific results and to study the complementarity of various research trends in an application domain which is simple to understand and which raises complex issues at the same time. Taaable has been at the origin of the ANR CONTINT project Kolflow, whose application domain is WikiTaaable, the semantic wiki of Taaable.

4.4. Agronomy

Participants: Sébastien Da Silva, Florence Le Ber [contact person], Jean-François Mari.

Keywords: simulation, Markov model, Formal Concept Analysis, graph

In September, Sébastien da Silva has defended his PhD thesis [13]. His research was conducted in the framework of an Inria-INRA collaboration, which takes place in the INRA research network PAYOTE about landscape modeling. The thesis, supervised both by Claire Lavigne (DR in ecology, INRA Avignon) and Florence Le Ber, was concerned with the characterization and the simulation of hedgerows structures in agricultural landscapes, based on Hilbert-Peano curves and Markov models [6] [13], [66], [98].

An on-going research work about the representation of peasant knowledge is involved within a collaboration with IRD in Madagascar [40]. Sketches drawn by peasants were transformed into graphs and compared thanks to Formal Concept Analysis.

5. New Software and Platforms

5.1. Generic Symbolic KDD Systems

5.1.1. The Coron Platform

Participants: Jérémie Bourseau, Aleksey Buzmakov, Victor Codocedo, Adrien Coulet, Amedeo Napoli [contact person], Yannick Toussaint.

Keywords: data mining, frequent itemset, closed itemset, generator, association rule, rare itemset

The Coron platform [133], [120] is a KDD toolkit organized around three main components: (1) Coron-base, (2) AssRuleX, and (3) pre- and post-processing modules. The software was registered at the "Agence pour la Protection des Programmes" (APP) and is freely available (see http://coron.loria.fr).

The Coron-base component includes a complete collection of data mining algorithms for extracting itemsets such as frequent itemsets, closed itemsets, generators and rare itemsets. In this collection we can find APriori, Close, Pascal, Eclat, Charm, and, as well, original algorithms such as ZART, Snow, Touch, and Talky-G [45]. AssRuleX generates different sets of association rules (from itemsets), such as minimal non-redundant association rules, generic basis, and informative basis. In addition, the Coron system supports the whole life-cycle of a data mining task and proposes modules for cleaning the input dataset, and for reducing its size if necessary.

The Coron toolkit is developed in Java, is operational, and was already used in several research projects.

5.1.2. Orion: Skycube Computation Software

Participant: Chedy Raïssi [contact person].

Keywords: skyline, skycube

This program implements the algorithms described in a research paper published at VLDB 2010 [127]. The software provides a list of four algorithms discussed in the paper in order to compute skycubes. This is the most efficient –in term of space usage and runtime– implementation for skycube computation (see https://github.com/leander256/Orion).

5.2. Stochastic systems for knowledge discovery and simulation

5.2.1. The CarottAge System

Participants: Florence Le Ber, Jean-François Mari [contact person].

Keywords: Hidden Markov Models, stochastic process

The system CarottAge is based on Hidden Markov Models of second order and provides a non supervised temporal clustering algorithm for data mining and a synthetic representation of temporal and spatial data [92]. CarottAge is currently used by INRA researchers interested in mining the changes in territories related to the loss of biodiversity (projects ANR BiodivAgrim and ACI Ecoger) and/or water contamination. CarottAge is also used for mining hydromorphological data. Actually a comparison was performed with three other algorithms classically used for the delineation of river continuum and CarottAge proved to give very interesting results for that purpose [121].

CarottAge is freely available under GPL license (see http://www.loria.fr/~jfmari/App/). A special effort is currently aimed at designing interactive visualization tools to provide the expert a user-friendly interface.

5.2.2. The ARPEnTAge System

Participant: Jean-François Mari [contact person].

Keywords: Hidden Markov Models, stochastic process

ARPEnTAge, for "*Analyse de Régularités dans les Paysages : Environnement, Territoires, Agronomie*" (http://www.loria.fr/~jfmari/App/) is a software based on stochastic models (HMM2 and Markov Field) for analyzing spatio-temporal data-bases [124]. ARPEnTAge is built on top of the CarottAge system to fully take into account the spatial dimension of input sequences. It takes as input an array of discrete data in which the columns contain the annual land-uses and the rows are regularly spaced locations of the studied landscape. It performs a Time-Space clustering of a landscape based on its time dynamic Land Uses (LUS). Displaying tools and the generation of Time-dominant shape files have also been defined.

ARPEnTAge is freely available (GPL license) and is currently used by INRA researchers interested in mining the changes in territories related to the loss of biodiversity (projects ANR BiodivAgrim and ACI Ecoger) and/or water contamination. In these practical applications, CarottAge and ARPEnTAge aim at building a partition –called the hidden partition– in which the inherent noise of the data is withdrawn as much as possible. The estimation of the model parameters is performed by training algorithms based on the Expectation Maximization and Mean Field theories. The ARPEnTAge system takes into account: (i) the various shapes of the territories that are not represented by square matrices of pixels, (ii) the use of pixels of different size with composite attributes representing the agricultural pieces and their attributes, (iii) the irregular neighborhood relation between those pixels, (iv) the use of shape files to facilitate the interaction with GIS (geographical information system).

ARPEnTAge and CarottAge were used for mining decision rules in a territory showing environmental issues. They provide a way of visualizing the impact of farmers decision rules in the landscape and revealing new extra hidden decision rules [132].

5.3. KDD in Systems Biology

Participants: Marie-Dominique Devignes [contact person], Malika Smaïl-Tabbone.

5.3.1. IntelliGO Online

The IntelliGO measure computes semantic similarity between terms from a structured vocabulary (Gene Ontology: GO) and uses these values for computing functional similarity between genes annotated by sets of GO terms [104]. The IntelliGO measure is available on line (http://plateforme-mbi.loria.fr/intelligo/) to be used for evaluation purposes. It is possible to compute the functional similarity between two genes, the intra-set similarity value in a given set of genes, and the inter-set similarity value for two given sets of genes.

5.3.2. WAFOBI: KNIME Nodes for Relational Mining of Biological Data

KNIME (for "Konstanz Information Miner") is an open-source visual programming environment for data integration, processing, and analysis. The KNIME platform aims at facilitating the data mining experiment settings as many tests are required for tuning the mining algorithms. Various KNIME nodes were developed for supporting relational data mining using the ALEPH program (http://www.comlab.ox.ac.uk/oucl/research/areas/machlearn/Aleph/aleph.pl). These nodes include a data preparation node for defining a set of first-order predicates from a set of relation schemes and then a set of facts from the corresponding data tables (learning set). A specific node allows to configure and run the ALEPH program to build a set of rules. Subsequent nodes allow to test the first-order rules on a test set and to perform configurable cross validations.

5.3.3. MOdel-driven Data Integration for Mining (MODIM)

The MODIM software (MOdel-driven Data Integration for Mining) is a user-friendly data integration tool which can be summarized along three functions: (i) building a data model taking into account mining requirements and existing resources; (ii) specifying a workflow for collecting data, leading to the specification of wrappers for populating a target database; (iii) defining views on the data model for identified mining scenarios.

Although MODIM is domain independent, it was used so far for biological data integration in various internal research studies and for organizing data about non ribosomal peptide syntheses. The sources can be downloaded at https://gforge.inria.fr/projects/modim/.

5.4. Knowledge-Based Systems and Semantic Web Systems

5.4.1. The Kasimir System for Decision Knowledge Management

Participants: Nicolas Jay, Jean Lieber [contact person], Amedeo Napoli.

Keywords: classification-based reasoning, case-based reasoning, decision knowledge management, knowledge edition, knowledge base maintenance, semantic portal

The objective of the Kasimir system is decision support and knowledge management for the treatment of cancer. A number of modules have been developed within the Kasimir system for editing treatment protocols, visualization, and maintenance. Kasimir is developed within a semantic portal, based on OWL. KatexOWL (Kasimir Toolkit for Exploiting OWL Ontologies, http://katexowl.loria.fr) was developed in a generic way and is applied to Kasimir. In particular, the user interface EdHibou of KatexOWL is used for querying the protocols represented within the Kasimir system In [109], this research is presented, together with an extension of Kasimir for multi-viewpoint case-based reasoning.

CabamakA (case base mining for adaptation knowledge acquisition) is a module of the Kasimir system. This system performs case base mining for adaptation knowledge acquisition and provides information units to be used for building adaptation rules. Actually, the mining process in CabamakA is based on a frequent close itemset extraction module from the Coron platform (see §5.1.1).

The Oncologik system is a collaborative editing tool aiming at facilitating the management of medical guidelines. Based on a semantic wiki, it allows the acquisition of formalized decision knowledge also includes a graphical decision tree editor called KcatoS. A version of Oncologik was released in 2012 (http://www.oncologik.fr/).

5.4.2. Taaable: a System for Retrieving and Creating New Cooking Recipes by Adaptation

Participants: Valmi Dufour-Lussier, Emmanuelle Gaillard, Florence Le Ber, Jean Lieber, Amedeo Napoli, Emmanuel Nauer [contact person].

Keywords: knowledge acquisition, ontology engineering, semantic annotation, case-based reasoning, hierarchical classification, text mining

Taaable is a system whose objectives are to retrieve textual cooking recipes and to adapt these retrieved recipes whenever needed [4]. Suppose that someone is looking for a “leek pie” but has only an “onion pie” recipe: how can the onion pie recipe be adapted?

The Taaable system combines principles, methods, and technologies such as case-based reasoning (CBR), ontology engineering, text mining, text annotation, knowledge representation, and hierarchical classification. Ontologies for representing knowledge about the cooking domain, and a terminological base for binding texts and ontology concepts, were built from textual web resources. These resources are used by an annotation process for building a formal representation of textual recipes. A CBR engine considers each recipe as a case, and uses domain knowledge for reasoning, especially for adapting an existing recipe w.r.t. constraints provided by the user, holding on ingredients and dish types.

The Taaable system is available on line since 2008 at http://taaable.fr, and is constantly evolving. This year, a new version of Taaable has been implemented in order to participate to the 7th Computer Cooking Contest which held during the International Case-Based Reasoning, in Cork, Ireland. The new version of Taaable is based on Tuuurbine, a generic ontology guided CBR engine over RDFS (see Section 5.4.3), and Revisor, an adaptation engine implementing various revision operators (see Section 5.4.5). In particular, Revisor is used to compute ingredient substitutions and to adjust the ingredient quantities.

5.4.3. Tuuurbine: a Generic Ontology Guided Case-Based Inference Engine

Participants: Jean Lieber, Emmanuel Nauer [contact person].

Keywords: case-based reasoning, inference engine, knowledge representation, ontology engineering, semantic web

The experience acquired since 5 years with the Taaable system conducted to the creation of a generic cased-based reasoning system, whose reasoning procedure is based on a domain ontology [63]. This new system, called Tuuurbine (http://tuuurbine.loria.fr/), takes into account the retrieval step, the case base organization, and also an adaptation procedure which is not addressed by other generic case-based reasoning tools. Moreover, Tuuurbine is built over semantic web standards that will ensure facilities for being plugged over data available on the web. The domain knowledge is represented in an RDF store, which can be interfaced with a semantic wiki, for collaborative edition and management of the knowledge involved in the reasoning system (cases, ontology, adaptation rules). The development of Tuuurbine was supported by an Inria ADT funding until October 2013. Tuuurbine is distributed under an Affero GPL License and is available from http://tuuurbine.loria.fr/.

5.4.4. BeGoood: a Generic System for Managing Non-Regression Tests on Knowledge Bases

Participant: Emmanuel Nauer [contact person].

Keywords: tests, non-regression, knowledge evolution

BeGoood is a system allowing to define test plans, independent of any application domain, and usable for testing any system answering queries by providing results in the form of sets of strings. BeGoood provides all the features usually found in test systems, such as tests, associated queries, assertions, and expected result sets, test plans (sets of tests) and test reports. The system is able to evaluate the impact of a system modification by running again test plans and by evaluating the assertions which define whether a test fails or succeeds. The main components of BeGoood are (1) the "test database" that stores every test artifacts, (2) the "remote query evaluator" which evaluates test queries, (3) the "assertion engine" which evaluates assertions over the expected and effective query result sets, (4) the "REST API" which offers the test functionalities as web services, and finally (5) the "Test controller" and (6) the "Test client".

BeGoood is available under a AGPL license on github [1]. BeGoood is used by the Taaable system (see Section 5.4.2) for managing the evolution of the knowledge base used by the CBR system.

5.4.5. Revisor: a Library of Revision Operators and Revision-Based Adaptation Operators

Participants: Valmi Dufour-Lussier, Alice Hermann, Florence Le Ber, Jean Lieber [contact person], Emmanuel Nauer, Gabin Personeni.

Keywords: belief revision, adaptation, revision-based adaptation, case-based reasoning, inference engines, knowledge representation

Revisor is a library of inference engines dedicated to belief revision and to revision-based adaptation for case-based reasoning [3]. It is open source, under a GPL license and available on the web (http://revisor.loria.fr/). It gathers several engines developed during the previous years for various knowledge representation formalisms (propositional logic—with or without the use of adaptation knowledge [93]—conjunction of linear constraints, and qualitative algebras [61], [75], [87], [14]). Some of these engines are already used in the Taaable system. Current developments on Revisor aim at defining new engines in other formalisms.

6. New Results

6.1. Highlights of the Year

As highlights of the year, we would like to mention several elements, an award in a competition and a best paper. In addition we would like to also mention the importance gained by two other papers.

---

[1] https://github.com/kolflow/begoood

- Yen Low, a postdoctoral fellow from Stanford and Adrien Coulet (Orpailleur team) jointly developed a prototype named *Whypothesis?* whose goal is to provide explanations on drug side effects for which the molecular mechanism remains unknown. This prototype won the "Best Application Award" at the 2014 NCBO Hackathon (National Center for Biomedical Ontology), held at Stanford University, April 26-27 (http://www.bioontology.org/2014_NCBO_Hackathon).
- The paper [2] describing a first and original proposition for combining pattern structures and relational concept analysis won the best paper award at the International Conference on Formal Concept Analysis in Cluj-Napoca, Romania.
- The paper [10] published in Nucleic Acids Research describes the latest version of KBDOCK, which has had over 12,000 non-duplicate visitors since 2011.
- The paper [44] on polypharmacology represents a nice collaboration with Harmonic Pharma, and it was used for the cover issue of Journal Chemical Information (http://pubs.acs.org/toc/jcisd8/54/3).

BEST PAPER AWARD :

[56] **A Proposition for Combining Pattern Structures and Relational Concept Analysis in Formal Concept Analysis - 12th International Conference - Proceedings**. V. CODOCEDO, A. NAPOLI.

6.2. The Mining of Complex Data

Participants: Mehwish Alam, Aleksey Buzmakov, Melisachew Chekol, Victor Codocedo, Adrien Coulet, Elias Egho, Nicolas Jay, Florence Le Ber, Ioanna Lykourentzou, Luis-Felipe Melo, Amedeo Napoli, Chedy Raïssi, Mohsen Sayed, My Thao Tang, Yannick Toussaint.

Keywords: formal concept analysis, relational concept analysis, pattern structures, pattern mining, association rule, graph mining, sequence mining, biclustering

Formal Concept Analysis and pattern mining are suitable symbolic methods for KDDK, that may be used for real-sized applications. Global improvements are carried on the scope of applicability, the ease of use, the efficiency of the methods, and on the ability to fit evolving situations. Accordingly, the team is extending these symbolic data mining methods for working on complex data (e.g. textual documents, biological, chemical or medical data), involving objects with multi-valued attributes (e.g. domains or intervals), n-ary relations, sequences, trees and graphs.

6.2.1. FCA and Variations: RCA, Pattern Structures and Biclustering

There are a few extensions of FCA for handling contexts involving complex data formats, e.g. graphs or relational data. Among them, Relational Concept Analysis (RCA) is a process for analyzing objects described both by binary and relational attributes [2] [131]. The RCA process takes as input a collection of contexts and of inter-context relations, and yields a set of lattices, one per context, whose concepts are linked by relations. RCA can play has an important role in KDDK, especially in text mining [105].

Another extension of FCA is based on Pattern Structures (PS) [112], which allows to build a concept lattice from complex data, e.g. nominal, numerical, and interval data [119]. Since then, we worked on some experiments involving pattern structures, namely sequence mining [107], information retrieval and recommendation [58], [22], functional dependencies [50], [17] and biclustering [69], [41]. One of the next step is the adaptation of pattern structures to graph mining.

Moreover, the notion of similarity between objects is also closely related to pattern structures [102]: two objects are similar as soon as they share the same attributes (binary case) or attributes with similar values or the same description (at least in part). Combination of similarity and pattern structures is also under study, in particular for solving information retrieval and annotation problems.

In pattern mining as in FCA, one main problem is the volume of the output. One general idea is to extract patterns which show a "good behavior" w.r.t. a given measure. Such patterns or concepts are expected to have good characteristics and to provide effective knowledge. We have conducted in the framework of FCA a series of experiments on the so-called "stability measure", showing that this measure is able to detect significant patterns [54], [53].

Finally, there is also an on-going work relating FCA and semantic web. This work focuses on the classification within a concept lattice of the answers returned by SPARQL queries. The concept lattice is then used as an index for navigating and ranking the answers w.r.t. their content and interest for a given objective [47].

6.2.2. Sequence Mining

Sequence data is widely used in many applications. Consequently, mining sequential patterns and other types of knowledge from sequence data became an important data mining task. In the team, the main emphasis is on developing efficient mining algorithms for pattern classification problems. The most frequent sequences generally provide trivial information. When analyzing the set of frequent sequences with a low minimum support, the user is overwhelmed by millions of patterns.

In our recent work, we studied the notion of δ-freeness for sequences. While this notion has extensively been discussed for itemsets, our work is the first to extend it to sequences. We defined an efficient algorithm devoted to the extraction of δ-free sequential patterns. We presented the advantage of the δ-free sequences and highlighted their importance when building sequence classifiers, and we showed how they can be used to address the feature selection problem in statistical classifiers which optimizes both accuracy and earliness of predictions [68].

6.2.3. Mining and Understanding Healthcare Trajectories

With the increasing burden of chronic illnesses, administrative health care databases hold valuable information that could be used to monitor and assess the processes shaping the trajectory of care of chronic patients. In this context, temporal data mining methods are promising tools, though lacking flexibility in addressing the complex nature of medical events. In the thesis work of Elias Egho [15], new algorithms were designed to extract patient trajectory patterns with different levels of granularity by relying on external taxonomies [62], [34]. The algorithms rely on the general FCA framework to formalize the general notion of multidimensional healthcare trajectories. There was also another work focusing on the similarity measure among sequences. An efficient and original similarity measure was design for that purpose [8].

6.2.4. Video Game Analytics

The video game industry has grown enormously over the last twenty years, bringing new challenges to the artificial intelligence and data analysis communities. We tackled this year the problem of automatic discovery of strategies in real-time strategy games through pattern mining. Such patterns are the basic units for many tasks such as automated agent design, but also to build tools for the professionally played video games in the electronic sports scene. We presented a new formalism within a sequential pattern mining approach and a novel measure, the balance measure, telling how a strategy is likely to win [51]. We experimented our methodology on a real-time strategy game that is professionally played in the electronic sport community and laid plans on a future collaboration with the MIT Game Lab.

6.2.5. KDDK in Text Mining

Ontologies help software and human agents to communicate by providing shared and common domain knowledge, and by supporting various tasks, e.g. problem-solving and information retrieval. In practice, building an ontology depends on a number of “ontological resources” having different types: thesaurus, dictionaries, texts, databases, and ontologies themselves. We are currently working on the design of a methodology based on FCA and RCA for ontology engineering from heterogeneous ontological resources. This methodology is based on both FCA and RCA, and was previously successfully applied in domains such as astronomy and biology.

In the framework of the ANR Hybride project (see 8.2.1.2), an engineer is implementing a robust system based on these previous research results, for preparing the way to new research directions involving trees and graphs. Moreover, we led a first successful experiment on extracting drug-drug interactions applying “lazy pattern structure classification” to syntactic trees. In addition, in his thesis work, Mohsen Sayed focused on extracting relations between named entities using graph mining methods applied to dependency graphs [67].

We are currently investigating how this approach can be generalized, i.e. how to detect a relation between complex expressions which are not previously recognized as named entities.

The notion of "Jumping Emerging Patterns" (JEP) previously used in chemistry [101], was updated and adapted in the context of text mining within the ANR Termith project. The objective is to design a learning method for filtering candidate terms within a full text and to decide whether an occurrence should be tagged as a term, i.e. a positive example, or as a simple word, i.e. a negative example. The method extracts from a training set all JEPs which are considered as hypotheses. To reduce the number of JEPs and to retain only the more significant JEPs from a linguistic point of view, JEPs are weighted and a constraint solver is used to verify the maximal coverage of the positive examples. Results are currently under evaluation.

6.3. KDDK in Life Sciences

Participants: Adrien Coulet, Marie-Dominique Devignes, Bernard Maigret, Gabin Personeni, David Ritchie, Malika Smaïl-Tabbone.

The Life Sciences constitute a challenging domain for KDDK. Biological data are complex from many points of views, e.g. voluminous, high-dimensional and deeply inter-connected. Analyzing such data is a crucial issue in health care, environment and agronomy. Besides, many bio-ontologies are available and can be used to enhance the knowledge discovery process. Accordingly, the research work of the Orpailleur team in KDDK applied to Life Sciences is in concern with the use of bio-ontologies to improve KDDK, and as well information retrieval, access to "Linked Open Data" (LOD) and data integration.

6.3.1. Inductive Logic Programming for Mining Linked Open Data

Increasing amounts of biomedical data provided as LOD offer novel opportunities for knowledge discovery in biomedicine. We proposed and published an approach for selecting, integrating, and mining LOD with the goal of discovering genes responsible for a disease [11]. The selection step relies on a set of choices made by a domain expert to isolate relevant pieces of LOD. Because these pieces are potentially not linked, an integration step is required to connect unlinked pieces. The resulting graph is subsequently mined using Inductive Logic Programming (ILP) that presents two main advantages. First, the input format compliant with ILP (first order logic) is close to the format of LOD (RDF triples). Second, domain knowledge can be added to this input and used during the induction step. We have applied this approach to the characterization of genes responsible for intellectual disability. For this real-world use case, we could evaluate ILP results and assess the contribution of domain knowledge. Our ongoing efforts explore how the combination of rules coming from distinct theories can improve the prediction accuracy [70] [16].

6.3.2. Analysis of biomedical data annotated with ontologies

Annotating data with concepts of an ontology is a common practice in the biomedical domain. Resulting annotations define links between data and ontologies that are key for data exchange, data integration and data analysis. Since 2011, we collaborate with the National Center for Biomedical Ontologies (NCBO) to develop a large repository of annotations named the NCBO Resource Index [118]. This repository contains annotations of 36 biomedical databases annotated with concepts of more than 200 ontologies of the BioPortal (http://bioportal.bioontology.org/). In the preceding years, we compared the annotations of a database of biomedical publications (Medline) with two databases of scientific funding (Crisp and ResearchCrossroads) to profile disease research [122]. One main challenge remains to develop a knowledge discovery approach able to mine correlations between annotations based on BioPortal ontologies, i.e. is it possible to discover interesting knowledge units within these annotations?

Then, we proposed an adaptation of FCA techniques, namely pattern structures, to explore the annotations of biomedical databases [108]. We considered documents of biomedical databases annotated with sets of ontological concepts as objects in a pattern structure. Corresponding annotations have been classified according to several dimensions, where a dimension is related to a particular aspect of domain knowledge. The pattern structure formalism was applied to classify these annotations, allowing to discover correlations between annotations but also lacks of completion in the annotations that could be fixed afterward. This adaptation of pattern structures opens many perspectives in term of ontology reengineering and knowledge discovery.

6.4. Structural Systems Biology

Participants: Marie-Dominique Devignes, Bernard Maigret, David Ritchie, Malika Smaïl-Tabbone.

Keywords: bioinformatics, chemistry, docking, knowledge discovery, screening, systems biology

Structural systems biology aims to describe and analyze the many components and interactions within living cells in terms of their three-dimensional (3D) molecular structures. We are currently developing advanced computing techniques for molecular shape representation, protein-protein docking, protein-ligand docking, high-throughput virtual drug screening, and knowledge discovery in databases dedicated to protein-protein interactions.

6.4.1. The Hex Protein Docking Program

Our *Hex* protein docking software is being more widely used than ever before. The unique polar Fourier correlation approach used in *Hex* [129] allows the expensive FFT part of its calculations to be greatly accelerated on modern graphics processors (GPUs) [130]. *Hex* is freely available for download for academic users at http://hex.loria.fr. A public GPU-powered server has also been created (http://hexserver.loria.fr) [123]. In the last four years, the server has performed some 63,700 docking runs, and the program has had some 37,000 downloads. The latest version of the program has been used successfully to dock symmetric dimers (unpublished results) in the international “CAPRI” docking experiment [115]. A manuscript on performing polar Fourier docking using symmetry constraints is in preparation with the Nano-D team at Inria Grenoble.

6.4.2. KBDOCK: Protein Docking Using Knowledge-Based Approaches

In order to explore the possibilities of using structural knowledge of protein-protein interactions, Anisah Ghoorah recently developed the KBDOCK system as part of her doctoral thesis project [116]. KBDOCK is available at http://kbdock.loria.fr. KBDOCK combines coordinate data from the Protein Data Bank [106] with the Pfam protein domain family classification [111] in order to describe and analyze all known protein-protein interactions for which the 3D structures are available. We have demonstrated the utility of KBDOCK [114] for template-based docking using 73 complexes from the Protein Docking Benchmark [117]. We recently presented results obtained using KBDOCK at the CAPRI conference on protein docking in Utrecht [115]. In late 2013, we updated KBDOCK with the latest data from Pfam and the Protein Data Bank. In 2014, an article describing the new version of KBDOCK was published in the special Database Issue of Nucleic Acids Research [10]. Since the KBDOCK web site (http://kbdock.loria.fr) was created in 2011, it has had over 12,000 distinct visitors.

6.4.3. Kpax: A New Algorithm for Multiple Flexible Protein Structure Alignments

We recently developed a new protein structure alignment approach called Kpax [128]. The approach exploits the fact that each amino acid residue has a carbon atom with a highly predictable tetrahedral geometry. This allows the local environment of each residue to be transformed into a canonical orientation, thus allowing easy comparison between the canonical orientations of residues within pairs of proteins using a novel scoring function based on Gaussian overlaps. The overall approach is two or three orders of magnitude faster than most contemporary protein structure alignment algorithms, while still being almost as accurate as the state-of-the-art TM-Align approach [134]. Kpax is now used heavily by the KBDOCK web server [10] to find structural templates for docking which might be beyond the reach of sequence-based homology modeling approaches. The Kpax program is also available for download at http://kpax.loria.fr/.

In 2014, the Kpax algorithm has been extended to allow flexible alignment and superposition of protein backbones and to perform multiple structure alignments, in analogy with multiple protein sequence alignments. Our early results show that incorporating backbone flexibility leads to much higher quality multiple alignments than can be achieved with existing algorithms.

6.4.4. Polypharmacology: Developing New Uses for Old Drugs

In 2010, Violeta Pérez-Nueno joined the Orpailleur team thanks to a Marie Curie Intra-European Fellowship (IEF) award to develop new virtual screening algorithms (DOVSA). The aim of this project was to advance the state of the art in computational virtual drug screening by developing a novel consensus shape clustering approach based on spherical harmonic (SH) shape representations [126].

In 2012, Violeta joined Harmonic Pharma, a LORIA spin-out company for drug re-purposing, and we have since continued our collaborations to develop new algorithms for drug discovery and drug re-purposing. The observation that many existing drugs may be used to treat more than one disease is often referred to as "polypharmacology." Our latest work on predicting polypharmacology uses a Gaussian clustering approach to identify groups molecules with similar three-dimensional shapes. This work was published in the Journal of Chemical Information and Modeling [44]. An illustration from this article was used to provide the cover page for the March 2014 issue of the journal (http://pubs.acs.org/toc/jcisd8/54/3).

6.5. Around the Taaable research project

Participants: Valmi Dufour-Lussier, Emmanuelle Gaillard, Florence Le Ber, Jean Lieber, Amedeo Napoli, Emmanuel Nauer.

Keywords: knowledge representation, description logics, classification-based reasoning, case-based reasoning, belief revision, semantic web

The Taaable project was originally created as a challenger of the Computer Cooking Contest (ICCBR Conference) [4] (http://taaable.fr). A candidate to this contest is a system whose goal is to solve cooking problems.

Beyond its participation to the CCC challenges, the Taaable project aims at federating various research themes: case-based reasoning (CBR), information retrieval, knowledge acquisition and extraction, knowledge representation, minimal change theory, ontology engineering, semantic wikis, text-mining, etc. CBR performs adaptation of recipes w.r.t. user constraints. The reasoning process is based on a cooking domain ontology (especially hierarchies of classes) and adaptation rules. The knowledge base is encoded within a semantic wiki containing the recipes, the domain ontology and adaptation rules.

Minimal change theory and belief revision can be used as tools to support adaptation in CBR, i.e. the source case is modified to be consistent with the target problem using a revision operator. Belief revision was applied to Taaable to compute ingredient substitutions and to adjust the ingredient quantities [65] using engines included in the Revisor library (see § 5.4.5).

As acquiring knowledge from experts is costly, a new approach was proposed to allow a CBR system to use partially reliable, non expert, knowledge from the Web for reasoning. This approach is based on a meta-knowledge model to manage knowledge reliability. This model represents notions such as belief, trust, reputation and quality, as well as their relationships and rules to evaluate knowledge reliability. The reliability estimation is used to filter knowledge with high reliability as well as to rank the results produced by the CBR system. Performing CBR with knowledge resulting from an e-community is improved by taking into account the knowledge reliability [64].

Taaable won in 2014 the CCC originality challenge for all the open resources that the Taaable team developed during the last years for the CBR community: WikiTaaable, a semantic wiki containing cooking domain knowledge, Tuuurbine, a generic ontology guided CBR engine over RDFS (see § 5.4.3), and Revisor, an adaptation engine implementing various revision operators (see § 5.4.5).

6.6. Some Results in Graph Theory

Participants: Miguel Couceiro, Amedeo Napoli, Chedy Raïssi, Jean-Sébastien Sereni, Mario Valencia.

Keywords: graph theory, extremal graph theory, coloring, clustering

6.6.1. Structural and extremal graph theory

Regarding graph coloring, a conjecture of Gera, Okamoto, Rasmussen and Zhang on set coloring was solved. A *set coloring* of a graph $G = (V, E)$ is a function $c : V \to \{1, ..., k\}$ such that whenever u and v are adjacent vertexes, it holds that $\{c(x) : x \text{ neighbor of } u\} \neq \{c(x) : x \text{ neighbor of } v\}$. In other words, there must be at least one neighbor of u that has a color not assigned to a neighbor of v, or *vice-versa*. The smallest k such that G admits a set coloring is the *set coloring number* $\chi_s(G)$. We confirmed the conjecture by proving that $\chi_s(G) \geq \lceil \log_2 \chi(G) \rceil + 1$, where $\chi(G)$ is the (usual) chromatic number of G. This bound is tight.

Works have been started on a 12-year-old conjecture by Heckman and Thomas about the fractional chromatic number of graphs with no triangles and maximum degree at most 3. This conjecture is actually a natural generalization of a fact established by Staton in 1979. Heckman and Thomas posits that in every graph with no triangles, maximum degree at most 3 and arbitrary weights on the vertexes, there exists an independent set of weight at least $5/14$ times the total weight of the graph.

Regarding extremal graph theory, two results have been obtained. The first one deals with permutation snarks, while the second one reads as follows.
For every 3-coloring of the edges of the complete graph on n vertexes, there is a color c and a set X of 4-vertexes such that at least $2n/3$ vertexes are linked to a vertex in X by an edge of color c.
This theorem is motivated by a conjecture of Erdős, Faudree, Gould, Gyárfás, Rousseau and Schelp from 1989, which asserts that X can be of size 3 only. However, they were only able to prove that X can be of size 22. Recently, Rahil Baber and John Talbot managed to build upon our work in a very nice article: adding a new idea to our argument, they managed to confirm the conjecture.

6.6.2. Graph theory and other fields

Interactions of graph theory with other topics (theoretical computer science, number theory, group theory, sociology and chemistry) have been considered. Most of them are still in progress and some are published. For instance, regarding distributed computing, the purpose of our work was to question the global knowledge each node is assumed to start with in many distributed algorithms (both deterministic and randomized). More precisely, numerous sophisticated local algorithm were suggested in the literature for various fundamental problems. Noticeable examples are the MIS algorithms and the $(\Delta + 1)$-coloring algorithms. Unfortunately, most known local algorithms are *non-uniform*, that is, they assume that all nodes know good estimations of one or more global parameters of the network, e.g., the number of nodes n. Our work provides a rather general method for transforming a non-uniform local algorithm into a uniform one. Furthermore, the resulting algorithm enjoys the same asymptotic running time as the original non-uniform algorithm. Our method applies to a wide family of both deterministic and randomized algorithms. Specifically, it applies to almost all of the state of the art non-uniform algorithms regarding MIS and Maximal Matching, as well as to many results concerning the coloring problem.

6.6.3. Algorithmic Graph Theory and Clustering

Since September 2013, Mario Valencia has obtained a two years invitation (namely Inria "Délégation”) for working at Inria Nancy – Grand Est, in the Orpailleur team, on graph theoretical aspects and data clustering. This research work consists in studying the modular decomposition techniques on the threshold graphs issues of the clustering process. The principal studied problem is known as the *Cluster Deletion Problem*: given a graph with real non negative edge weights, partition the vertexes into clusters (in this case cliques) in order to minimize the total weight of edges out of the clusters. Two papers were submitted to journals in 2014. In [94], we discovered a one-to-one correspondence between potential solutions of the cluster deletion problem and the minimum sum coloring problem, and use it to obtain a polynomial time algorithm to solve the cluster deletion problem in a special family of graphs called P_4-reducible graphs.

In [95], we studied the complexity of the cluster deletion problem on subclasses of chordal graphs and cographs. In particular, it is shown that the cluster deletion problem is NP-hard for unweighted chordal graphs and weighted cographs. Some polynomial-time solvable cases are also identified.

Moreover, the paper "b-coloring is NP-hard on co-bipartite graphs and polytime solvable on tree-cographs", has been accepted for publication in the journal *Algorithmica* [1].

6.6.4. Structural and Algebraic Graph Theory

We have also worked on the following topics. Golumbic, Lipshteyn and Stern proved that every graph can be represented as the edge intersection graph of paths on a grid, i.e., one can associate to each vertex of the graph a nontrivial path on a grid such that two vertexes are adjacent if and only if the corresponding paths share at least one edge of the grid. For a non-negative integer k, B_k-EPG graphs are defined as graphs admitting a model in which each path has at most k bends. Circular-arc graphs are intersection graphs of open arcs of a circle. It is easy to see that every circular-arc graph is B_4-EPG, by embedding the circle into a rectangle of the grid. We proved also that every circular-arc graph is B_3-EPG (paper submitted).

We have studied the k-tuple chromatic number of the Cartesian product of two graphs G and H in [96]. We have shown that there exist graphs G and H such that $\chi_k(G \Box H) > \max\{\chi_k(G), \chi_k(H)\}$ for $k \geq 2$. Moreover, we have also shown that there exist graph families such that, for any $k \geq 1$, the k-tuple chromatic number of their Cartesian product is equal to the maximum k-tuple chromatic number of its factors.

7. Bilateral Contracts and Grants with Industry

7.1. The BioIntelligence Project

Participants: Mehwish Alam, Aleksey Buzmakov, Melisachew Chekol, Adrien Coulet, Marie-Dominique Devignes, Amedeo Napoli [contact person], Nicolas Pépin-Hermann, Malika Smaïl-Tabbone.

The objective of the “BioIntelligence” project is to design an integrated framework for the discovery and the development of new biological products. This framework takes into account all phases of the development of a product, from molecular to industrial aspects, and is intended to be used in life science industry (pharmacy, medicine, cosmetics, etc.). The framework has to propose various tools and activities such as: (1) a platform for searching and analyzing biological information (heterogeneous data, documents, knowledge sources, etc.), (2) knowledge-based models and process for simulation and biology in silico, (3) the management of all activities related to the discovery of new products in collaboration with the industrial laboratories (collaborative work, industrial process management, quality, certification). The “BioIntelligence” project is led by “Dassault Systèmes” and involves industrial partners such as Sanofi Aventis, Laboratoires Pierre Fabre, Ipsen, Servier, Bayer Crops, and two academics, Inserm and Inria. An annual meeting of the project usually takes place in Sophia-Antipolis at the beginning of July.

Two theses related to “BioIntelligence” are currently in preparation within the Orpailleur team. A first thesis is related to the mining of complex biological data using FCA and RCA techniques [47], [48], [49]. The objective is to take advantage of Linked Open Data in biology for helping the biologist for querying and navigating complex data. There are needs to integrate data and knowledge from several web biological resources. At present, some experiments are conducted on designing practical interfaces based on sophisticated visualization tools for allowing human agents to have an easy and quick access to interesting patterns .

A second thesis is based on an extension of FCA involving Pattern Structures on complex data such as sequences and graphs [107]. The idea is to extend the formalism of pattern structures to these complex data for being able to classify complex structures such as patient trajectories or molecular structures. The classification results (e.g. concept lattices) are expected to help practitioners in information retrieval tasks and specific problem solving. In addition, a theoretical and practical work was conducted on the evaluation of interest measures for selecting the best concepts to be analyzed by a human agent in a concept lattice, and especially the stability measure in FCA. This led to a series of original and pioneering experiments on this probably underestimated research subject [20], [54], [53].

7.2. The Quaero Project

Participants: Victor Codocedo [contact person], Ioanna Lykourentzou, Amedeo Napoli.

The Quaero project (http://www.quaero.org) is a program aimed at promoting research and industrial innovation on technologies for automatic analysis and classification of multimedia and multilingual documents (the project ended at the beginning of 2014). The partners collaborate on research and the realization of advanced demonstrators and prototypes of innovating applications and services for access and usage of multimedia information, such as spoken language, images, video and music.

In this framework, the Orpailleur team worked on information retrieval, document annotation and recommendation. The objective was to define methods and algorithms for achieving these complex tasks, based on KDDK techniques and especially the FCA technology.

Part of the thesis of Victor Codocedo was prepared in this context, focusing on information retrieval guided by domain knowledge, recommendation and classification of documents w.r.t. sets of annotations using FCA and pattern structures [2] [58], [22].

8. Partnerships and Cooperations

8.1. International Initiatives

8.1.1. Participation In International Programs

8.1.1.1. Ciência Sem Fronteiras

Participant: Amedeo Napoli [contact person].

Program “Ciência Sem Fronteiras” is a Brazilian research fellowship which provides a funding for the stay of a visiting French researcher in Brazil at Universidade Federal Pernambuco Recife for three years. The on-going project is called “Formal Concept Analysis as a Support for Knowledge Discovery” and is aimed at combining FCA methods with numerical clustering methods used by Brazilian colleagues. This project is supervised in Brazil by Professor Francisco de A.T. de Carvalho (CIn/UFPE).

The project aims at developing and comparing classification and clustering algorithms for complex data (especially interval and multi-valued data). Two families of algorithms are studied, namely “clustering algorithms” based on the use of a similarity or a distance for comparing the objects, and “classification algorithms in Formal Concept Analysis (FCA)” based on attribute sharing between objects. The objectives here are to combine the facilities of both families of algorithms for improving the potential of each family in dealing with more complex and voluminous datasets.

8.1.1.2. Pronex Brasilia

Participant: Bernard Maigret [contact person].

In this research project, the goal is to identify, using virtual screening techniques that we developed, new compounds against tropical diseases (e.g. trypanosome, dengue and mycosis) in collaboration with several Brazilian laboratories among which the Department of Biology at University of Brasilia, together with the Harmonic Pharma start-up. Through this collaboration, several PhD and postdocs came to the lab for one year training with our home-developed virtual screening engine (VSM-G). This project is in part supported by the Brazilian CNPq agency. Fruitful results were already obtained leading to several papers in preparation and patents. These patents concern the discovery of new putative treatment of strong mycosis due to fungi particularly virulent in South America. These patents were funded by the University of Brasilia, Embrapa and Harmonic Pharma.

8.1.2. Inria Associate Team: Snowflake

Participants: Adrien Coulet [contact person], Gabin Personeni, Malika Smaïl-Tabbone.

Snowflake (http://snowflake.loria.fr/) is an Inria Associate Team which started in 2014. It is aimed at facilitating the collaboration between researchers from the Inria ORPAILLEUR team and the Stanford Center for Biomedical Informatics Research, Stanford University, USA. The main objective of Snowflake is to improve biomedical knowledge discovery by connecting Electronic Health Records (EHRs) with LOD (Linked Open Data). Such a connection could allow to complete domain knowledge w.r.t. EHRs. The initial focus of Snowflake is the identification and characterization of groups of patients w.r.t. reactions to drugs. Identifies features associated with such groups of patients could be used as predictors of over- or under-reactions to some drugs. The considered use case is related to pharmacogenomics drugs, i.e., drugs known to cause variable effects depending on the genetic profile of patients. Data associated with pharmacogenomics drugs and their mechanisms are available in LOD and, once connected to EHRs, they can be used to classify drugs and then patients presenting a specific reaction profile to a given group of drugs.

8.1.3. Explorer Programme

Participant: Chedy Raïssi [contact person].

Chedy Raïssi visited the MIT Lab (Massachusetts Institute of Technology, MIT http://web.mit.edu/) during one month between July and August 2014. The objective of this research visit was the following.

Over the last decade, the annual turnovers generated by the electronic entertainment industry went beyond those of both cinema and music industries, making video game production a highly profitable business. In parallel with the game industry growth, watching video-game live streams is becoming an increasingly popular way of entertainment. Given the shared interests, between Orpailleur team and the GameLab at MIT, in emerging broadcasting platforms for games and work on analytics, Chedy Raïssi started a productive collaboration with researchers from the GameLab at MIT where we he explored for one month the potential for future collaborations between the MIT and Inria on this interesting new topic.

8.1.4. Miscellaneous

Participants: Mehwish Alam, Aleksey Buzmakov, Melisachew Chekol, Victor Codocedo, Adrien Coulet, Elias Egho, Ioanna Lykourentzou, Amedeo Napoli [contact person], Chedy Raïssi, Jean-Sébastien Sereni, Mario Valencia.

- The team had a Fapemig – Inria research project between 2009 and 2013, called “Incorporating Knowledge Models into Scalable Data Mining Algorithms” (IKMSDM). The IKMSDM project involved researchers at Universidade Federal de Minas Gerais in Belo Horizonte –a group led by Prof. Wagner Meira– and the Orpailleur team at Inria Nancy Grand Est. In this project we were interested in the mining of large amount of data with two relevant application scenarios: text mining and graph mining. This year, contact and work was going on, focusing on the preparation of a joint publication on the notion of skylines for tensor data.
- An on-going collaboration involves the Orpailleur team and Sergei Kuznetsov at Higher School of Economics in Moscow (HSE). Amedeo Napoli visited HSE laboratory several times (with the support of HSE) while Sergei Kuznetsov visited Inria Nancy Grand Est several times too. The collaboration is materialized by the joint supervision of the thesis of Aleksey Buzmakov and the organization of scientific events, and in particular the workshop FCA4AI whose fourth edition will take place this year in July at IJCAI 2014 (see http://www.fca4ai.hse.ru).
- LEA STRUCO is an “Associated International Laboratory” of CNRS between IÚUK, Prague, and LIAFA, Paris. It focuses on high-level study of fundamental combinatorial objects, with a particular emphasis on comprehending and disseminating the state-of-the-art theories and techniques developed. The obtained insights shall be applied to obtain new results on existing problems as well as to identify directions and questions for future work. Jean-Sébastien Sereni is the contact person for LEA STRUCO which was initiated when Jean-Sébastien was a member of LIAFA.
- At present, Mario Valencia is the international coordinator of the MathAmSud project 13MATH-07 “Structural an algebraic problems on graph theory” (2013–2015). This project is funded by the following research institutes: CNRS in France, MinCyT in Argentina, CAPES in Brazil and CMM in Chile.

8.2. National Initiatives

8.2.1. ANR

8.2.1.1. HEREDIA

Participant: Jean-Sébastien Sereni [contact person].

HEREDIA (http://www.liafa.univ-paris-diderot.fr/~sereni/Heredia/) is an ANR JCJC ("Jeunes Chercheurs") focusing on hereditary properties of graphs, which provide a general perspective to study graph properties. Several important general theorems are known and the approach offers an elegant way of unifying notions and proof techniques. Further, hereditary classes of graphs play a central role in graph theory. Besides their theoretical appeal, they are also particularly relevant from an algorithmic point of view. With Jean-Sébastien Sereni, the HEREDIA project involves Pierre Charbit (LIAFA, Paris), Louis Esperet (G-SCOP, Grenoble) and Nicolas Trotignon (LIP, Lyon).

8.2.1.2. Hybride

Participants: Adrien Coulet, Luis-Felipe Melo, Amedeo Napoli, Matthieu Osmuk, Chedy Raïssi, My Thao Tang, Mohsen Sayed, Yannick Toussaint [contact person].

The Hybride research project (http://hybride.loria.fr/) aims at combining Natural Language Processing (NLP) and Knowledge Discovery in Databases (KDD) for text mining. A key idea is to design an interacting and convergent process where NLP methods are used for guiding text mining and KDD methods are used for guiding the analysis of textual documents. NLP methods are mainly based on text analysis and extraction of general and temporal information. KDD methods are based on pattern mining, e.g. patterns and sequences, formal concept analysis and graph mining. In this way, NLP methods applied to texts extract "textual information" that can be used by KDD methods as constraints for focusing the mining of textual data. By contrast, KDD methods extract patterns and sequences to be used for guiding information extraction from texts and text analysis. Experimental and validation parts associated with the Hybride project are provided by an application to the documentation of rare diseases in the context of Orphanet.

The partners of the Hybride consortium are the GREYC Caen laboratory (pattern mining, NLP, text mining), the MoDyCo Paris laboratory (NLP, linguistics), the INSERM Paris laboratory (Orphanet, ontology design), and the Orpailleur team at Inria NGE (FCA, knowledge representation, pattern mining, text mining).

8.2.1.3. ISTEX

Participants: Luis-Felipe Melo, Amedeo Napoli, Yannick Toussaint [contact person].

ISTEX is a so-called "Initiative d'excellence" managed by CNRS and DIST ("Direction de l'Information Scientifique et Technique"). ISTEX aims at giving to the research and teaching community an on-line access to scientific publications in all the domains. Thus ISTEX is in concern with a massive acquisition of documentation such as journals, proceedings, corpus, databases...ISTEX-R is one research project within ISTEX in which is involved the Orpailleur team, with two other partners, namely the ATILF laboratory and the INIST Institute (both in Nancy). ISTEX-R aims at developing new tools for querying full-text documentation, analyzing content and extracting information. A platform is currently under development to provide robust NLP tools for text processing, as well as methods in text mining and domain conceptualization.

8.2.1.4. Kolflow

Participants: Jean Lieber [contact person], Alice Hermann, Amedeo Napoli, Emmanuel Nauer, My Thao Tang, Yannick Toussaint.

Kolflow (http://kolflow.univ-nantes.fr/) is a 3-year basic research project taking place from February 2011 to November 2014, funded by French National Agency for Research (ANR), program ANR CONTINT. The aim of the project is to investigate man-machine collaboration in continuous knowledge-construction flows.

Kolflow partners are GDD (LINA Nantes), Silex (LIRIS Lyon), Orpailleur (Inria NGE/LORIA), Coast (Inria NGE/LORIA), and Wimmics (Inria Sophia Antipolis).

8.2.1.5. PEPSI: Polynomial Expansions of Protein Structures and Interactions

Participants: David Ritchie [contact person], Marie-Dominique Devignes, Malika Smaïl-Tabbone, Seyed Ziaeddin Alborzi.

The PEPSI ("Polynomial Expansions of Protein Structures and Interactions") project is a collaboration with Sergei Grudinin at Inria Grenoble (project Nano-D) and Valentin Gordeliy at the Institut de Biologie Structurale (IBS) in Grenoble. This four-year project funded by the ANR "Modèles Numériques" program involves developing computational protein modeling and docking techniques and using them to help solve the structures of large molecular systems experimentally (http://pepsi.gforge.inria.fr).

8.2.1.6. Termith

Participants: Luis-Felipe Melo, Yannick Toussaint [contact person].

Termith (http://www.atilf.fr/ressources/termith/) is an ANR Project which involves the following laboratories: ATILF, LIDILEM, LINA, INIST, Inria Saclay and Inria Nancy Grand Est. It aims at indexing documents belonging to different domain of Humanities. Thus, the project focuses on extracting candidate terms (information extraction) and on disambiguation.

In the Orpailleur team, we are mainly concerned by information extraction using Formal Concept Analysis techniques, but also pattern and sequence mining. The objective is to define "contexts introducing terms", i.e. finding textual environments allowing a system to decide whether a textual element is actually a candidate term and its corresponding environment.

8.2.1.7. Trajcan: a study of patient care trajectories

Participants: Elias Egho, Nicolas Jay [contact person], Amedeo Napoli, Chedy Raïssi.

Since 30 years, many patient classification systems (PCS) have been developed. These systems aim at classifying care episodes into groups according to different patient characteristics. In France, the so-called "Programme de Médicalisation des Systèmes d'Information" (PMSI) is a national wide PCS in use in every hospital. It systematically collects data about millions of hospitalizations. Though it is used for funding purposes, it includes useful information for public health domains such as epidemiology or health care planning.

The objective of the Trajcan project was to represent and analyze "patient care trajectories" (patient suffering from cancer limited to breast, colon, rectum, and lung cancers) and the associated healthcares (it should be noticed that the Trajcan Project ended at the beginning of 2014). The data are related to patients receiving hospital cares in the "Bourgogne" region and using data from PMSI. Such an analysis involves various data, e.g. type of cancer, number of visits, type of stays, hospitalization services, therapies used, and demographic factors such as age, gender, place of residence.

Elias Egho defended a Phd thesis on this subject in July 2014 [15]. Combining knowledge discovery and knowledge representation methods for improving the definition of patients as temporal objects (sequential data mining), he successfully developed different approaches for characterizing Patient Care Trajectories (PCT). A first characterization is based on sequential pattern structures, extending Formal Concept Analysis techniques to multidimensional sequential data. A second one, involves an algorithm called MMISP for "Mining Multidimensional Itemsets Sequential Patterns" and makes use of external knowledge to improve the mining process and discover sequential patterns at different levels of granularity [62]. Finally, a new similarity measure was developed for comparing sequences of itemsets and for applying clustering methods to classify patients having similar healthcare trajectories. This later work was distinguished by a forthcoming publication in Data Mining and Knowledge Discovery.

8.2.2. Other National Initiatives and Collaborations

8.2.2.1. Towards the discovery of new nonribosomal peptides and synthetases

We have initiated a collaboration with researchers from the LIFL and Université Lille Nord de France on the NRPS toolbox. Data was cleaned and integrated from various public and specific analysis programs. The resulting database should facilitate the process of knowledge discovery of new nonribosomal peptides and synthetases. Actual results of this research collaboration were published in [21].

8.2.2.2. FUI Poqemon

Participant: Chedy Raïssi [Contact Person].

The POQEMON project aims at developing new pattern mining methods and tools for guiding knowledge discovery from mobile phone networks for monitoring purposes. The main idea is to develop sound approaches that handle the trade-off between privacy of data and the power of analysis.

8.3. Regional Initiatives

8.3.1. Le Bois Santé (LBS)

Participants: Marie-Dominique Devignes [contact person], Malika Smaïl-Tabbone.

The project "LBS – Le Bois Santé – #38017" is funded by the European Regional Development Fund (FEDER) and the French "Fonds Unique Interministériel (FUI)" in the framework of the BioProLor consortium. This project is coordinated by "Harmonic Pharma", a start-up specialized in the identification of active principles in natural products. The aim of LBS is to exploit wood products in the pharmaceutical and nutriment domains. Concerned people in the team are working on data management and knowledge discovery about new therapeutic applications.

8.3.2. PEPS Mirabelle EXPLOD-Biomed

Participants: Adrien Coulet [contact person], Marie-Dominique Devignes, Gabin Personeni, Malika Smaïl-Tabbone.

This project has initiated a collaboration with geneticists from the Hospital of Nancy, namely Philippe Jonveaux and Céline Bonnet. The aim of the EXPLOD-Biomed project is to propose novel knowledge discovery methods applied to Linked Open Data for discovering gene that could be responsible for intellectual deficiencies. Linked Open Data are available on-line, interconnected and encoded in a format which can be straightforwardly mapped to ontologies. Thus they offer novel opportunities for knowledge discovery in biomedical data. Here, geneticists play the role of experts and guide the knowledge discovery process at different steps.

8.3.3. Hydreos

Participant: Jean-François Mari [contact person].

Hydreos is a state organization –actually a so-called "Pôle de compétitivité"– aimed at evaluating the quality of water (http://www.hydreos.fr/fr). Actually, water resources rely on many agronomic variables, including land use successions. Accordingly, one objective of our participation in Hydreos is to have a better understanding of the changes in the organization of a territory. The data to be analyzed are obtained by surveys or by satellite images and describe the land use at the level of the agricultural parcel. Then there is a search for detecting changes in land use and for correlating these changes to groundwater quality.

The systems ARPEnTAge (see § 5.2.2) and CarottAge (see § 5.2.1) are used in this context, especially by agronomists of INRA (ASTER Mirecourt http://www6.nancy.inra.fr/sad-aster.

This year, our research work focused on implementing various display tools to have a better understanding of the clustering results that a stochastic modeling provide http://www.loria.fr/~jfmari/App/Arpentage/Yar.avi.

8.3.4. Contrat Plan État Région" (CPER)

A part of the links between the Regional Administration and Inria Nancy Grand Est/LORIA are materialized through the so-called "Contrat Plan État Région" (CPER) which is running from 2015 to 2020. There is an associated scientific program in which the Orpailleur team is involved.

Some members of the Orpailleur team participated to the definition of a project in one of the two tracks of the interdisciplinary scientific program called “Santé et Vieillissement”. The other track called “Innovations Technologiques, Modélisation et Médecine Personnalisée (IT2MP)” is coordinated by Pr. Zannad (CHU-Nancy). We proposed a project called “Simulation et Modélisation pour l’Extraction de Connaissances (SMEC)” which gathers physicians, bio-statisticians, chemists and computer scientists. The objective of this project is to design innovative methodologies for analyzing cohort data and make progress towards personalized medicine.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organization, General Chairs, Scientific Chairs

- “ECCB 2014”. Several members of the team were involved in the organization of the 13th edition of the European Conference on Computational Biology held in Strasbourg on September 2014. This conference attracted over 1100 participants (http://www.eccb14.org). Marie-Dominique Devignes was the scientific chair of the conference and the scientific editor of the proceedings [7] [76].
- “ECML-PKDD 2014”. Amedeo Napoli and Chedy Raïssi were the Conference chairs of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2014) which was held in Nancy between September 15th and September 19th, http://www.ecmlpkdd2014.org/, in conjunction with ILP 2014 (September 14-16, http://dtai.cs.kuleuven.be/events/ilp2014/). There were roughly 560 persons attending the ECML-PKDD Conference (50 of which were shared with ILP).
- “CCC”. Emmanuel Nauer co-organized the "Computer Cooking Contest" at ICCBR 2014 which held this year in Cork (Ireland) (see http://liris.cnrs.fr/ccc/ccc2014/).
- “FCA4AI 2014”. Amedeo Napoli organized with Sergei O. Kuznetsov (HSE Moscow) and Sebastian Rudolph (TU Dresden) the third workshop FCA4AI (“What can do FCA for Artificial Intelligence”) which was associated with the ECAI Conference in Prague (Czech Republic, August 2014, see http://www.fca4ai.hse.ru/2014 and http://ceur-ws.org/Vol-1257).
- “ICCBR Workshops”. Jean Lieber with David B. Leake was in charge of the organization of the ICCBR-2014 workshops [82] (see http://www.iccbr.org/iccbr14/).
- “LD4KD”. Nicolas Jay organized with Mathieu D’Aquin and Ilaria tiddi (Open University, UK) the first workshop on Linked Data for Knowledge Discovery (LD4KD) collocated with the ECML/PKDD conference in Nancy, September 2014 (http://events.kmi.open.ac.uk/ld4kd2014/).

9.1.1.1. Scientific Animation

- The scientific animation in the Orpailleur team is based on two seminars, the Team Seminar and the BINGO seminar. The Team Seminar is held in general twice a month and is used either for general presentations of members of the team or for invited presentations of external researchers. The BINGO seminar is also held twice a month and is used for more specific presentations focusing on biological, chemical, and medical topics. Actually, both seminars are active and are useful instruments for researchers in the team.
- Members of the Orpailleur team are all involved, as members or as head persons, in various national research groups.
- The members of the Orpailleur team are involved in the organization of conferences and workshops, as members of conference program committees (ECAI, ECML-PKDD, ICCBR, ICDM, ICFCA, IJCAI, KDD...), as members of editorial boards, and finally in the organization of journal special issues.

9.2. Teaching – Supervision – Committees

- The members of the Orpailleur team are involved in teaching at all levels of teaching, mainly at University of Lorraine. Actually, most of the members of the Orpailleur team are employed on university positions.
- The members of the Orpailleur team are also involved in student supervision, at all university levels, from under-graduate until post-graduate students.
- Finally, the members of the Orpailleur team are involved in HDR and thesis defenses, being thesis referees or thesis committee members.

10. Bibliography

Major publications by the team in recent years

[1] F. BONOMO, O. SCHAUDT, M. STEIN, M. VALENCIA-PABON. *b-coloring is NP-hard on co-bipartite graphs and polytime solvable on tree-cographs*, in "Algorithmica", 2014, 17 p. , An extended abstract of this paper appears in the proceedings of ISCO symposium, LNCS 8596, pp. 100-11, 2014 [*DOI :* 10.1007/S00453-014-9921-5], https://hal.archives-ouvertes.fr/hal-01102516

[2] V. CODOCEDO, A. NAPOLI. *A Proposition for Combining Pattern Structures and Relational Concept Analysis*, in "Formal Concept Analysis - 12th International Conference, ICFCA 2014, Cluj-Napoca, Romania, June 10-13, 2014. Proceedings", Cluj-Napoca, Romania, June 2014, pp. 96 - 111 [*DOI :* 10.1007/978-3-319-07248-7_8], https://hal.inria.fr/hal-01095870

[3] J. COJAN, J. LIEBER. *Applying Belief Revision to Case-Based Reasoning*, in "Computational Approaches to Analogical Reasoning: Current Trends", Studies in Computational Intelligence, Springer, 2014, vol. 548, pp. 133 - 161 [*DOI :* 10.1007/978-3-642-54516-0_6], https://hal.inria.fr/hal-01095344

[4] A. CORDIER, V. DUFOUR-LUSSIER, J. LIEBER, E. NAUER, F. BADRA, J. COJAN, E. GAILLARD, L. INFANTE-BLANCO, P. MOLLI, A. NAPOLI, H. SKAF-MOLLI. *Taaable: a Case-Based System for personalized Cooking*, in "Successful Case-based Reasoning Applications-2", S. MONTANI, L. C. JAIN (editors), Studies in Computational Intelligence, Springer, January 2014, vol. 494, pp. 121-162 [*DOI :* 10.1007/978-3-642-38736-4_7], https://hal.inria.fr/hal-00912767

[5] M. COUCEIRO, E. LEHTONEN, K. SCHÖLZEL. *Hypomorphic Sperner Systems and Non-Reconstructible Functions*, in "Order", July 2014, 42 p. [*DOI :* 10.1007/S11083-014-9330-Z], https://hal.archives-ouvertes.fr/hal-01090540

[6] S. DA SILVA, F. LE BER, C. LAVIGNE. *Structure Analysis of Hedgerows with Respect to Perennial Landscape Lines in Two Contrasting French Agricultural Landscapes*, in "International Journal of Agricultural and Environmental Information Systems (IJAEIS)", 2014, vol. 5, n^{o} 1, 19 p. [*DOI :* 10.4018/IJAEIS.2014010102], https://hal.archives-ouvertes.fr/hal-01057108

[7] M.-D. DEVIGNES, Y. MOREAU. *ECCB 2014: The 13th European Conference on Computational Biology*, in "Bioinformatics", September 2014, vol. 30, n^{o} 17, 4 p. [*DOI :* 10.1093/BIOINFORMATICS/BTU512], https://hal.inria.fr/hal-01097339

[8] E. EGHO, C. RAÏSSI, T. CALDERS, N. JAY, A. NAPOLI. *On measuring similarity for sequences of itemsets*, in "Data Mining and Knowledge Discovery", July 2014, 33 p. [*DOI :* 10.1007/S10618-014-0362-1], https://hal.inria.fr/hal-01094383

[9] R. GERBIER, V. LEROUX, P. COUVINEAU, R. ALVEAR-PEREZ, B. MAIGRET, C. LLORENS-CORTES, X. ITURRIOZ. *New structural insights into the apelin receptor: identification of key residues for apelin binding*, in "The FASEB Journal", January 2015, vol. 29, n^{o} 1, pp. 314-322, https://hal.inria.fr/hal-01101857

[10] A. GHOORAH, M.-D. DEVIGNES, M. SMAÏL-TABBONE, D. RITCHIE. *KBDOCK 2013: A spatial classification of 3D protein domain family interactions*, in "Nucleic Acids Research", January 2014, vol. 42, n^{o} D1, pp. 389-395, https://hal.inria.fr/hal-00920612

[11] G. PERSONENI, S. DAGET, C. BONNET, P. JONVEAUX, M.-D. DEVIGNES, M. SMAÏL-TABBONE, A. COULET. *Mining Linked Open Data: A Case Study with Genes Responsible for Intellectual Disability*, in "Data Integration in the Life Sciences - 10th International Conference, DILS 2014", Lisbon, Portugal, E. R. HELENA GALHARDAS (editor), Lecture Notes in Computer Science, Springer, July 2014, vol. 8574, pp. 16 - 31 [*DOI :* 10.1007/978-3-319-08590-6_2], https://hal.inria.fr/hal-01095591

[12] Y. XIAO, C. MIGNOLET, J.-F. MARI, M. BENOÎT. *Modeling the spatial distribution of crop sequences at a large regional scale using land-cover survey data: A case from France*, in "Computers and Electronics in Agriculture", March 2014, vol. 012, pp. 51-63, https://hal.inria.fr/hal-00967890

Publications of the year

Doctoral Dissertations and Habilitation Theses

[13] S. DA SILVA. *Spatial data mining and modelling of hedgrows in agricultural landscapes*, Université de Lorraine, September 2014, https://hal.inria.fr/tel-01101424

[14] V. DUFOUR-LUSSIER. *Reasoning with Qualitative Spatial and Temporal Textual Cases*, Université de Lorraine, October 2014, https://hal.inria.fr/tel-01098087

[15] E. EGHO. *Mining Heterogeneous Multidimensional Sequential Data* , Université de Lorraine, July 2014, https://hal.inria.fr/tel-01094400

[16] M. SMAÏL-TABBONE. *Contribution to knowledge extraction from biological data*, Université de Lorraine, November 2014, Habilitation à diriger des recherches, https://hal.inria.fr/tel-01093943

Articles in International Peer-Reviewed Journals

[17] J. BAIXERIES, M. KAYTOUE, A. NAPOLI. *Characterizing functional dependencies in formal concept analysis with pattern structures*, in "Annals of Mathematics and Artificial Intelligence", 2014, vol. 72, pp. 129 - 149 [*DOI :* 10.1007/s10472-014-9400-3], https://hal.inria.fr/hal-01101107

[18] A. BERRY, A. GUTIERREZ, M. HUCHARD, A. NAPOLI, A. SIGAYRET. *Hermes: a simple and efficient algorithm for building the AOC-poset of a binary relation*, in "Annals of Mathematics and Artificial Intelligence", 2014, vol. ISSN, n^{o} 72, pp. 45 - 71 [*DOI :* 10.1007/s10472-014-9418-6], https://hal.inria.fr/hal-01101144

[19] F. BONOMO, O. SCHAUDT, M. STEIN, M. VALENCIA-PABON. *b-coloring is NP-hard on co-bipartite graphs and polytime solvable on tree-cographs*, in "Algorithmica", 2014, 17 p. , An extended abstract of this paper appears in the proceedings of ISCO symposium, LNCS 8596, pp. 100-11, 2014 [*DOI :* 10.1007/s00453-014-9921-5], https://hal.archives-ouvertes.fr/hal-01102516

[20] A. Buzmakov, S. O. Kuznetsov, A. Napoli. *Is Concept Stability a Measure for Pattern Selection?*, in "procedia computer science", 2014, vol. 31, pp. 918 - 927 [*DOI :* 10.1016/j.procs.2014.05.344], https://hal.inria.fr/hal-01095914

[21] T. Caradec, M. Pupin, A. Vanvlassenbroeck, M.-D. Devignes, M. Smaïl-Tabbone, P. Jacques, V. Leclère. *Prediction of Monomer Isomery in Florine: A Workflow Dedicated to Nonribosomal Peptide Discovery*, in "PLoS ONE", January 2014, vol. 9, n^{o} 1, e85667 [*DOI :* 10.1371/journal.pone.0085667], https://hal.archives-ouvertes.fr/hal-01090619

[22] V. Codocedo, I. Lykourentzou, A. Napoli. *A semantic approach to concept lattice-based information retrieval*, in "Annals of Mathematics and Artificial Intelligence", October 2014, vol. 72, pp. 169 - 195 [*DOI :* 10.1007/s10472-014-9403-0], https://hal.inria.fr/hal-01095859

[23] M. Couceiro, E. Lehtonen. *A survey on the arity gap*, in "Journal of Multiple-Valued Logic and Soft Computing", 2015, vol. 24, n^{o} 1–4, pp. 223–249, https://hal.archives-ouvertes.fr/hal-01093666

[24] M. Couceiro, E. Lehtonen. *On the arity gap of finite functions : results and applications*, in "Journal of Multiple-Valued Logic and Soft Computing", 2015, 15 p. , https://hal.archives-ouvertes.fr/hal-01093929

[25] M. Couceiro, E. Lehtonen, K. Schölzel. *Hypomorphic Sperner Systems and Non-Reconstructible Functions*, in "Order", July 2014, 42 p. [*DOI :* 10.1007/s11083-014-9330-z], https://hal.archives-ouvertes.fr/hal-01090540

[26] M. Couceiro, E. Lehtonen, K. Schölzel. *A complete classification of equational classes of threshold functions included in clones*, in "RAIRO Operations Research", 2015, vol. 49, n^{o} 1, pp. 39-66, forthcoming [*DOI :* 10.1051/ro/2014034], https://hal.archives-ouvertes.fr/hal-01090621

[27] M. Couceiro, E. Lehtonen, K. Schölzel. *Set-reconstructibility of Post classes*, in "Discrete Applied Mathematics", 2015, 8 p. , arXiv admin note: text overlap with arXiv:1306.5578, https://hal.archives-ouvertes.fr/hal-01090618

[28] M. Couceiro, E. Lehtonen, T. Waldhauser. *Additive decomposition schemes for polynomial functions over fields*, in "Novi Sad J. Math. ", 2014, vol. 44, n^{o} 2, 16 p. , https://hal.archives-ouvertes.fr/hal-01090554

[29] M. Couceiro, E. Lehtonen, T. Waldhauser. *On equational definability of function classes*, in "Journal of Multiple-Valued Logic and Soft Computing", 2015, vol. 24, n^{o} 1–4, pp. 203–222, https://hal.archives-ouvertes.fr/hal-01093668

[30] M. Couceiro, I. G. Rosenberg, L. Haddad. *Partial clones containing all Boolean monotone self-dual partial functions*, in "Journal of Multiple-Valued Logic and Soft Computing", 2015, 10 p. , https://hal.archives-ouvertes.fr/hal-01093942

[31] M. Couceiro, T. Waldhauser. *Pseudo-polynomial functions over finite distributive lattices*, in "Fuzzy Sets and Systems", 2014, vol. 239, pp. 21-34 [*DOI :* 10.1016/j.fss.2012.09.007], https://hal.archives-ouvertes.fr/hal-01090562

[32] S. Da Silva, F. Le Ber, C. Lavigne. *Structure Analysis of Hedgerows with Respect to Perennial Landscape Lines in Two Contrasting French Agricultural Landscapes*, in "International Journal of Agricultural and Envi-

ronmental Information Systems (IJAEIS)", 2014, vol. 5, n° 1, 19 p. [*DOI :* 10.4018/IJAEIS.2014010102], https://hal.archives-ouvertes.fr/hal-01057108

[33] M.-D. DEVIGNES, Y. MOREAU. *ECCB 2014: The 13th European Conference on Computational Biology*, in "Bioinformatics", September 2014, vol. 30, n° 17, 4 p. [*DOI :* 10.1093/BIOINFORMATICS/BTU512], https://hal.inria.fr/hal-01097339

[34] E. EGHO, N. JAY, C. RAÏSSI, D. IENCO, P. PONCELET, M. TEISSEIRE, A. NAPOLI. *A contribution to the discovery of multidimensional patterns in healthcare trajectories*, in "Journal of Intelligent Information Systems", April 2014, vol. 42, pp. 283 - 305 [*DOI :* 10.1007/S10844-014-0309-4], https://hal.inria.fr/hal-01094377

[35] E. EGHO, C. RAÏSSI, T. CALDERS, N. JAY, A. NAPOLI. *On measuring similarity for sequences of itemsets*, in "Data Mining and Knowledge Discovery", July 2014, 33 p. [*DOI :* 10.1007/S10618-014-0362-1], https://hal.inria.fr/hal-01094383

[36] B. FUCHS, J. LIEBER, A. MILLE, A. NAPOLI. *Differential adaptation: An operational approach to adaptation for solving numerical problems with CBR*, in "Knowledge-Based Systems", 2014, vol. 68, pp. 103 - 114 [*DOI :* 10.1016/J.KNOSYS.2014.03.009], https://hal.inria.fr/hal-01101145

[37] R. GERBIER, V. LEROUX, P. COUVINEAU, R. ALVEAR-PEREZ, B. MAIGRET, C. LLORENS-CORTES, X. ITURRIOZ. *New structural insights into the apelin receptor: identification of key residues for apelin binding*, in "The FASEB Journal", January 2015, vol. 29, n° 1, pp. 314-322, https://hal.inria.fr/hal-01101857

[38] A. GHOORAH, M.-D. DEVIGNES, M. SMAÏL-TABBONE, D. RITCHIE. *KBDOCK 2013: A spatial classification of 3D protein domain family interactions*, in "Nucleic Acids Research", January 2014, vol. 42, n° D1, pp. 389-395, https://hal.inria.fr/hal-00920612

[39] S. GUESSOUM, M. T. LASKRI, J. LIEBER. *RespiDiag: a Case-Based Reasoning System for the Diagnosis of Chronic Obstructive Pulmonary Disease*, in "Expert Systems with Applications", February 2014, vol. 41, pp. 267–273, https://hal.inria.fr/hal-00912641

[40] D. HERVÉ, J.-H. RAMAROSON, A. RANDRIANARISON, F. LE BER. *Comment les paysans du corridor forestier de Fianarantsoa (Madagascar) dessinent-ils leur territoire ? Des cartes individuelles pour confronter les points de vue*, in "Cybergeo : Revue européenne de géographie / European journal of geography", 2014, document 681, https://hal.archives-ouvertes.fr/hal-01057112

[41] M. KAYTOUE, S. O. KUZNETSOV, J. MACKO, A. NAPOLI. *Biclustering meets triadic concept analysis*, in "Annals of Mathematics and Artificial Intelligence", 2014, vol. 70, pp. 55 - 79 [*DOI :* 10.1007/S10472-013-9379-1], https://hal.inria.fr/hal-01101143

[42] A. NAPOLI. *Preface to the special issue on "Concept Lattice and their Applications" (CLA-2011)*, in "Annals of Mathematics and Artificial Intelligence", 2014, vol. 70, pp. 1 - 3 [*DOI :* 10.1007/S10472-013-9390-6], https://hal.inria.fr/hal-01101104

[43] P. POPOV, D. RITCHIE, S. GRUDININ. *DockTrina: Docking triangular protein trimers*, in "Proteins: Structure, Function, and Genetics", January 2014, vol. 82, n° 1, pp. 34-44 [*DOI :* 10.1002/PROT.24344], https://hal.inria.fr/hal-00880359

[44] V. PÉREZ-NUENO, A. S. KARABOGA, M. SOUCHET, D. RITCHIE. *GES Polypharmacology Fingerprints: A Novel Approach for Drug Repositioning*, in "Journal of Chemical Information and Modeling", February 2014, vol. 54, n^{o} 3, pp. 720-734 [*DOI :* 10.1021/CI4006723], https://hal.inria.fr/hal-01092928

[45] L. SZATHMARY, P. VALTCHEV, A. NAPOLI, R. GODIN, A. BOC, V. MAKARENKOV. *A fast compound algorithm for mining generators, closed itemsets, and computing links between equivalence classes*, in "Annals of Mathematics and Artificial Intelligence", 2014, vol. 70, pp. 81 - 105 [*DOI :* 10.1007/S10472-013-9372-8], https://hal.inria.fr/hal-01101140

[46] Y. XIAO, C. MIGNOLET, J.-F. MARI, M. BENOÎT. *Modeling the spatial distribution of crop sequences at a large regional scale using land-cover survey data: A case from France*, in "Computers and Electronics in Agriculture", March 2014, vol. 012, pp. 51-63, https://hal.inria.fr/hal-00967890

International Conferences with Proceedings

[47] M. ALAM, A. NAPOLI. *Defining Views with Formal Concept Analysis for Understanding SPARQL Query Results*, in "Proceedings of the Eleventh International Conference on Concept Lattices and Their Applications", Košice, Slovakia, October 2014, https://hal.inria.fr/hal-01089770

[48] M. ALAM, A. NAPOLI. *Lattice-Based View Access: A way to Create Views over SPARQL Query for Knowledge Discovery*, in "What can FCA do for Artificial Intelligence? (Third FCA4AI Workshop)", Prague, Czech Republic, August 2014, https://hal.inria.fr/hal-01089777

[49] M. ALAM, A. NAPOLI. *Lattice-Based Views over SPARQL Query Results*, in "Proceedings of the 1st Workshop on Linked Data for Knowledge Discovery co-located with European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2014)", Nancy, France, September 2014, https://hal.inria.fr/hal-01089774

[50] J. BAIXERIES, M. KAYTOUE, A. NAPOLI. *Characterization of Database Dependencies with FCA and Pattern Structures*, in "AIST - Third International Conference, Analysis of Images, Social Networks and Texts", Yekaterinburg, Russia, Communications in Computer and Information Science, Springer, 2014, vol. 436, pp. 3 - 14 [*DOI :* 10.1007/978-3-319-12580-0_1], https://hal.inria.fr/hal-01101147

[51] G. BOSC, M. KAYTOUE, C. RAÏSSI, J.-F. BOULICAUT, P. TAN. *Mining Balanced Sequential Patterns in RTS Games 1*, in "ECAI 2014 - 21st European Conference on Artificial Intelligence", Prague, Czech Republic, August 2014 [*DOI :* 10.3233/978-1-61499-419-0-975], https://hal.inria.fr/hal-01100933

[52] A. BUZMAKOV, S. O. KUZNETSOV, A. NAPOLI. *Concept Stability as a Tool for Pattern Selection*, in "FCA4AI 2014. What can FCA do for Artificial Intelligence?", Praque, Czech Republic, August 2014, https://hal.inria.fr/hal-01095903

[53] A. BUZMAKOV, S. O. KUZNETSOV, A. NAPOLI. *On Evaluating Interestingness Measures for Closed Itemsets*, in "7th European Starting AI Researcher Symposium (STAIRS 2014)", Prague, Czech Republic, J. L. ULLE ENDRISS (editor), Proceedings of the 7th European Starting AI Researcher Symposium (STAIRS 2014), 2014, vol. 264, pp. 71 - 80 [*DOI :* 10.3233/978-1-61499-421-3-71], https://hal.inria.fr/hal-01095927

[54] A. BUZMAKOV, S. O. KUZNETSOV, A. NAPOLI. *Scalable Estimates of Concept Stability*, in "12th International Conference on Formal Concept Analysis (ICFCA 2014)", Cluj-Napoca, Romania, C. S. CYN-

thia Vera Glodeanu (editor), Formal Concept Analysis, Springer, 2014, vol. 8478, pp. 157 - 172 [*DOI :* 10.1007/978-3-319-07248-7_12], https://hal.inria.fr/hal-01095920

[55] C. Chamard, E. Bresso, T. Boukobza, M.-D. Devignes, M. Smaïl-Tabbone, A. Chesnel, H. Dumond. *Long chain alkylphenol mixture promotes mammary epithelial cell metaplastic phenotype through an estrogen receptor alpha 36 mediated mechanism*, in "23rd Biennial Congress of the European Association for Cancer Research, EACR-23", Munich, Germany, European Journal of cancer, July 2014, vol. 50, n^{o} Supplement 5, S118 p. , Présentation Poster [*DOI :* 10.1016/S0959-8049(14)50435-2], https://hal.archives-ouvertes.fr/hal-01095429

[56] *Best Paper*
V. Codocedo, A. Napoli. *A Proposition for Combining Pattern Structures and Relational Concept Analysis*, in "Formal Concept Analysis - 12th International Conference - Proceedings", Cluj-Napoca, Romania, June 2014, pp. 96 - 111 [*DOI :* 10.1007/978-3-319-07248-7_8], https://hal.inria.fr/hal-01095870.

[57] V. Codocedo, A. Napoli. *Bicluster enumeration using Formal Concept Analysis*, in "What formal concept analysis can do for artificial intelligence? (FCA4AI 2014) Workshop at ECAI 2014", Prague, Czech Republic, August 2014, https://hal.inria.fr/hal-01095884

[58] V. Codocedo, A. Napoli. *Lattice-based biclustering using Partition Pattern Structures*, in "ECAI 2014 - 21st European Conference on Artificial Intelligence - Including Prestigious Applications of Intelligent Systems (PAIS) 2014", Prague, Czech Republic, August 2014 [*DOI :* 10.3233/978-1-61499-419-0-213], https://hal.inria.fr/hal-01095865

[59] M. Couceiro, L. Haddad, T. Waldhauser, K. Schölzel. *Relation Graphs and Partial Clones on a 2-Element Set*, in "IEEE 44th International Symposium on Multiple-Valued Logic (ISMVL)", Bremen , France, May 2014 [*DOI :* 10.1109/ISMVL.2014.36], https://hal.archives-ouvertes.fr/hal-01090638

[60] M. Couceiro, J.-L. Marichal. *Quasi-Lovász extensions on bounded chains*, in "IPMU - 15th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems", Montpellier, France, Communications in Computer and Information Science, Springer, July 2014, vol. 442 [*DOI :* 10.1007/978-3-319-08795-5_21], https://hal.archives-ouvertes.fr/hal-01090634

[61] V. Dufour-Lussier, A. Hermann, F. Le Ber, J. Lieber. *Belief revision in the propositional closure of a qualitative algebra* *, in "14th International Conference on Principles of Knowledge Representation and Reasoning", Vienne, Austria, AAAI Press, July 2014, 4 p. , https://hal.inria.fr/hal-01094264

[62] E. Egho, C. Raïssi, N. Jay, A. Napoli. *Mining Heterogeneous Multidimensional Sequential Patterns*, in "European Conference on Artificial Intelligence", Prague, Czech Republic, France, August 2014, vol. 263, 6 p. [*DOI :* 10.3233/978-1-61499-419-0-279], https://hal.inria.fr/hal-01094365

[63] E. Gaillard, L. Infante-Blanco, J. Lieber, E. Nauer. *Tuuurbine: A Generic CBR Engine over RDFS*, in "Case-Based Reasoning Research and Development", Cork, Ireland, September 2014, vol. 8765, pp. 140 - 154 [*DOI :* 10.1007/978-3-319-11209-1_11], https://hal.inria.fr/hal-01082372

[64] E. Gaillard, J. Lieber, E. Nauer, A. Cordier. *How Case-Based Reasoning on e-Community Knowledge Can Be Improved Thanks to Knowledge Reliability*, in "Case-Based Reasoning Research and Develop-

ment", Cork, Ireland, Ireland, Luc Lamontagne and Enric Plaza, September 2014, vol. 8765, pp. 155 - 169 [*DOI :* 10.1007/978-3-319-11209-1_12], https://hal.inria.fr/hal-01082369

[65] E. GAILLARD, J. LIEBER, E. NAUER. *Case-Based Cooking with Generic Computer Utensils: Taaable Next Generation*, in "Proceedings of the ICCBR 2014 Workshops", Cork, Ireland, David B. Leake and Jean Lieber, September 2014, n^{o} pp 89-100, 254 p. , https://hal.archives-ouvertes.fr/hal-01082683

[66] T. GUYET, F. LE BER, S. DA SILVA, C. LAVIGNE. *Comparaison des chemins de Hilbert adaptatif et des graphes de voisinage pour la caractérisation d'un parcellaire agricole*, in "Conférence Extraction et Gestion de Connaissances", Rennes, France, January 2014, https://hal.inria.fr/hal-00916964

[67] M. HASSAN, A. COULET, Y. TOUSSAINT. *Learning Subgraph Patterns from text for Extracting Disease–Symptom Relationships*, in "1st International Workshop on Interactions between Data Mining and Natural Language Processing", Nancy, France, P. CELLIER, T. CHARNOIS, A. HOTHO, S. MATWIN, M.-F. MOENS, Y. TOUSSAINT (editors), CEUR-WS, September 2014, vol. 1202, https://hal.inria.fr/hal-01095595

[68] P. HOLAT, M. PLANTEVIT, C. RAÏSSI, N. TOMEH, T. CHARNOIS, B. CRÉMILLEUX. *Sequence Classification Based on Delta-Free Sequential Pattern*, in "IEEE International Conference on Data Mining", Shenzhen, China, December 2014, https://hal.inria.fr/hal-01100929

[69] M. KAYTOUE, V. CODOCEDO, J. BAIXERIES, A. NAPOLI. *Three Related FCA Methods for Mining Biclusters of Similar Values on Columns*, in "Proceedings of the Eleventh International Conference on Concept Lattices and Their Applications", Kosice, Slovakia, October 2014, https://hal.inria.fr/hal-01095877

[70] G. PERSONENI, S. DAGET, C. BONNET, P. JONVEAUX, M.-D. DEVIGNES, M. SMAÏL-TABBONE, A. COULET. *ILP for Mining Linked Open Data: a biomedical Case Study*, in "The 24th International Conference on Inductive Logic Programming (ILP 2014)", Nancy, France, September 2014, https://hal.inria.fr/hal-01095597

[71] G. PERSONENI, S. DAGET, C. BONNET, P. JONVEAUX, M.-D. DEVIGNES, M. SMAÏL-TABBONE, A. COULET. *Mining Linked Open Data: A Case Study with Genes Responsible for Intellectual Disability*, in "DILS - 10th International Conference on Data Integration in the Life Sciences", Lisbon, Portugal, E. R. HELENA GALHARDAS (editor), Lecture Notes in Computer Science, Springer, July 2014, vol. 8574, pp. 16 - 31 [*DOI :* 10.1007/978-3-319-08590-6_2], https://hal.inria.fr/hal-01095591

[72] G. PERSONENI, A. HERMANN, J. LIEBER. *Adapting Propositional Cases Based on Tableaux Repairs Using Adaptation Knowledge*, in "International Conference on Case-Based Reasoning", Cork, Ireland, L. LAMONTAGNE, E. PLAZA (editors), Lecture Notes in Computer Science, Springer International Publishing, September 2014, vol. Case-Based Reasoning Research and Development, n^{o} 8765, pp. 390-404 [*DOI :* 10.1007/978-3-319-11209-1_28], https://hal.inria.fr/hal-01092210

[73] D. RIZZO, J.-F. MARI, E. MARRACCINI, E.-G. LAZRAK. *Agricultural landscape segmentation: a stochastic method to map heterogeneous variables*, in "Advances in Spatial Typologies: How to move from concepts to practice?", Lisbon, Portugal, IALE-Europe Thematic Workshop 2014, July 2014, https://hal.inria.fr/hal-01098402

National Conferences with Proceedings

[74] C. CHAMARD, E. BRESSO, T. BOUKOBZA, M.-D. DEVIGNES, M. SMAÏL-TABBONE, A. CHESNEL, H. DUMOND. *Long chain alkylphenol mixture promotes mammary epithelial cell metaplastic phenotype through an ERalpha36-mediated mechanism*, in "8e Forum du Canceropôle Grand Est", Strasbourg, France, November 2014, Présentation Poster, https://hal.archives-ouvertes.fr/hal-01095763

[75] V. DUFOUR-LUSSIER, A. HERMANN, F. LE BER, J. LIEBER. *Révision des croyances dans la clôture propositionnelle d'une algèbre qualitative*, in "JIAF-2014 – Huitièmes Journées de l'Intelligence Artificielle Fondamentale", Angers, France, JIAF - Huitièmes Journées de l'Intelligence Artificielle Fondamentale, June 2014, 10 p. , https://hal.inria.fr/hal-01093977

Scientific Books (or Scientific Book chapters)

[76] *Proceedings of ECCB 2014: The 13th European Conference on Computational Biology*, Bioinformatics, Oxford University PressStrasbourg, France, September 2014, vol. 30, n^{o} 17, https://hal.inria.fr/hal-01097346

[77] J. COJAN, J. LIEBER. *Applying Belief Revision to Case-Based Reasoning*, in "Computational Approaches to Analogical Reasoning: Current Trends", Studies in Computational Intelligence, Springer, 2014, vol. 548, pp. 133 - 161 [*DOI :* 10.1007/978-3-642-54516-0_6], https://hal.inria.fr/hal-01095344

[78] A. CORDIER, V. DUFOUR-LUSSIER, J. LIEBER, E. NAUER, F. BADRA, J. COJAN, E. GAILLARD, L. INFANTE-BLANCO, P. MOLLI, A. NAPOLI, H. SKAF-MOLLI. *Taaable: a Case-Based System for personalized Cooking*, in "Successful Case-based Reasoning Applications-2", S. MONTANI, L. C. JAIN (editors), Studies in Computational Intelligence, Springer, January 2014, vol. 494, pp. 121-162 [*DOI :* 10.1007/978-3-642-38736-4_7], https://hal.inria.fr/hal-00912767

[79] V. DUFOUR-LUSSIER, B. GUILLAUME, G. PERRIER. *Parsing Coordination Extragrammatically*, in "Human Language Technology. Challenges for Computer Science and Linguistics. 5th Language and Technology Conference, LTC 2011, Poznan, Poland, November 25-27, 2011, Revised Selected Papers", Z. VETULANI, J. MARIANI (editors), Springer International Publishing, July 2014, 12 p. [*DOI :* 10.1007/978-3-319-08958-4_5], https://hal.inria.fr/hal-00921033

Books or Proceedings Editing

[80] N. JAY, I. TIDDI, M. D'AQUIN (editors). *Proceedings of the Workshop on Linked Data for Knowledge Discovery (LD4KD 2014)*, CEUR Workshop ProceedingsNancy, France, September 2014, vol. 1232, n^{o} 3, https://hal.archives-ouvertes.fr/hal-01101149

[81] S. O. KUZNETSOV, A. NAPOLI, S. RUDOLPH (editors). *Proceedings of the International Workshop "What can FCA do for Artificial Intelligence?" (FCA4AI 2014)*, CEUR Proceedings, August 2014, vol. 1257, 100 p. , https://hal.inria.fr/hal-01101150

[82] D. B. LEAKE, J. LIEBER (editors). *Proceedings of the Workshops of the 22nd International Conference on Case-Based Reasoning*, September 2014, 260 p. , https://hal.inria.fr/hal-01097373

[83] M. MINOR, E. NAUER (editors). *Proceedings of the Computer Cooking Contest Workshop at ICCBR 2014*, Proceedings of the Workshops of the 22nd International Conference on Case-Based Reasoning, September 2014, 48 p. , https://hal.inria.fr/hal-01095840

Research Reports

[84] M. Alam, A. Napoli. *Lattice-Based View Access: A way to Create Views over SPARQL Query for Knowledge Discovery*, September 2014, https://hal.archives-ouvertes.fr/hal-01059528

[85] S. Bessy, D. Gonçalves, J.-S. Sereni. *Two floor building needing eight colors*, May 2014, https://hal.archives-ouvertes.fr/hal-00996709

[86] M. W. Chekol. *Schema Query Containment*, February 2014, n^o RR-8484, 27 p. , https://hal.inria.fr/hal-00951201

[87] V. Dufour-Lussier, A. Hermann, F. Le Ber, J. Lieber. *Belief revision in the propositional closure of a qualitative algebra (extended version)*, Inria Nancy, July 2014, This is the extended version of an article originally presented at the 14th International Conference on Principles of Knowledge Representation and Reasoning, https://hal.inria.fr/hal-00954512

[88] Z. Dvořák, J.-S. Sereni, J. Volec. *Fractional coloring of triangle-free planar graphs*, February 2014, https://hal.archives-ouvertes.fr/hal-00950493

[89] K. Edwards, J. Van Den Heuvel, R. J. Kang, J.-S. Sereni. *Extension from Precoloured Sets of Edges*, June 2014, https://hal.archives-ouvertes.fr/hal-01024843

[90] E. Egho, C. Raïssi, N. Jay, A. Napoli. *Mining Heterogeneous Multidimensional Sequential Patterns*, April 2014, n^o RR-8521, https://hal.inria.fr/hal-00979804

[91] M. Loebl, J.-S. Sereni. *Potts partition function and isomorphisms of trees*, May 2014, https://hal.archives-ouvertes.fr/hal-00992104

[92] J.-F. Mari. *CarottAge Windows pour les données Teruti : manuel d'utilisation*, Loria & Inria Grand Est, February 2014, 43 p. , https://hal.inria.fr/hal-00951102

[93] G. Personeni, A. Hermann, J. Lieber. *Adapting propositional cases based on tableaux repairs using adaptation knowledge – extended report*, 2014, https://hal.archives-ouvertes.fr/hal-01011751

Other Publications

[94] F. Bonomo, G. Duran, A. Napoli, M. Valencia-Pabon. *A one-to-one correspondence between potential solutions of the cluster deletion problem and the minimum sum coloring problem, and its application to P 4 -sparse graphs*, May 2014, Paper submitted to a journal, 11/05/2014, https://hal.archives-ouvertes.fr/hal-01102515

[95] F. Bonomo, G. Duran, M. Valencia-Pabon. *Complexity of the cluster deletion problem on subclasses of chordal graphs and cographs* , October 2014, Paper submitted to a Journal, 28/10/2014, https://hal.archives-ouvertes.fr/hal-01102512

[96] F. Bonomo, I. Koch, P. Torres, M. Valencia-Pabon. *k-tuple chromatic number of the cartesian product of graphs*, November 2014, https://hal.archives-ouvertes.fr/hal-01103534

[97] M. Couceiro, J.-L. Marichal, B. Teheux. *Conservative median algebras*, 2014, https://hal.archives-ouvertes.fr/hal-01090614

[98] T. GUYET, S. DA SILVA, C. LAVIGNE, F. LE BER. *Caractérisation d'un parcellaire agricole : comparaison des sacs de noeuds obtenus par un chemin de Hilbert adaptatif et un graphe de voisinage*, January 2014, 13 p. , Atelier FST "fouille de données spatiales et temporelles" associé à la conférence EGC, Rennes, 28 janvier 2014, https://hal.inria.fr/hal-01100583

[99] G. PERSONENI, S. DAGET, C. BONNET, P. JONVEAUX, M.-D. DEVIGNES, M. SMAÏL-TABBONE, A. COULET. *Mining Linked Open Data: a Case Study with Genes Responsible for Intellectual Disability*, September 2014, ECCB'14 (European Conference on Computational Biology 2014), https://hal.inria.fr/hal-01092800

References in notes

[100] F. BAADER, D. CALVANESE, D. MCGUINNESS, D. NARDI, P. PATEL-SCHNEIDER (editors). *The Description Logic Handbook*, Cambridge University PressCambridge, UK, 2003

[101] Y. ASSES, A. BUZMAKOV, T. BOURQUARD, S. O. KUZNETSOV, A. NAPOLI. *A Hybrid Classification Approach based on FCA and Emerging Patterns - An application for the classification of biological inhibitors*, in "CLA'12: The Ninth International Conference on Concept Lattices and Their Applications - 2012", Fuengirola, Spain, L. SZATHMARY, U. PRISS (editors), October 2012, http://hal.inria.fr/hal-00761586

[102] J. BAIXERIES, M. KAYTOUE, A. NAPOLI. *Computing Similarity Dependencies with Pattern Structures*, in "The Tenth International Conference on Concept Lattices and their Applications - CLA 2013", La Rochelle, France, France, M. OJEDA-ACIEGO, J. OUTRATA (editors), CEUR Workshop Proceedings Vol. 1062, Karell Bertet, 2013, pp. 33-44, http://hal.inria.fr/hal-00922592

[103] M. BARBUT, B. MONJARDET. *Ordre et classification – Algèbre et combinatoire (2 tomes)*, HachetteParis, 1970

[104] S. BENABDERRAHMANE, M. SMAÏL-TABBONE, O. POCH, A. NAPOLI, M.-D. DEVIGNES. *IntelliGO: a new vector-based semantic similarity measure including annotation origin*, in "BMC Bioinformatics", December 2010, vol. 11, n^{o} 1, 588 p. [*DOI :* 10.1186/1471-2105-11-588], http://www.biomedcentral.com/1471-2105/11/588/abstract

[105] R. BENDAOUD, A. NAPOLI, Y. TOUSSAINT. *Formal Concept Analysis: A unified framework for building and refining ontologies*, in "Knowledge Engineering: Practice and Patterns - Proceedings of the 16th International Conference EKAW", A. GANGEMI, J. EUZENAT (editors), Lecture Notes in Computer Science 5268, 2008, pp. 156–171

[106] H. M. BERMAN, T. BATTISTUZ, T. N. BHAT, W. F. BLUHM, P. E. BOURNE, K. BURKHARDT, L. IYPE, S. JAIN, P. FAGAN, J. MARVIN, D. PADILLA, V. RAVICHANDRAN, B. SCHNEIDER, N. THANKI, H. WEISSIG, J. D. WESTBROOK, C. ZARDECKI. *The Protein Data Bank*, in "Acta Crystallographica Section D-Biological Crystallography", 2002, vol. 58, pp. 899–907

[107] A. BUZMAKOV, E. EGHO, N. JAY, S. O. KUZNETSOV, A. NAPOLI, C. RAÏSSI. *On Projections of Sequential Pattern Structures (with an application on care trajectories)*, in "The Tenth International Conference on Concept Lattices and Their Applications - CLA'13", La Rochelle, France, October 2013, http://hal.inria.fr/hal-00910300

[108] A. COULET, F. DOMENACH, M. KAYTOUE, A. NAPOLI. *Using Pattern Structures for Analyzing Ontology-Based Annotations of Biomedical Data*, in "International Conference on Formal Concept Analysis", Dresden, Germany, Springer, May 2013, http://hal.inria.fr/hal-00880643

[109] M. D'AQUIN, J. LIEBER, A. NAPOLI. *Decentralized case-based reasoning and Semantic Web technologies applied to decision support in oncology*, in "Knowledge Engineering Review", March 2013, vol. 28, n^{o} 4, pp. 425–449 [*DOI :* 10.1017/S0269888913000027], http://hal.inria.fr/hal-00922080

[110] R. FELDMAN, J. SANGER. *The Text Mining Handbook (Advanced Approaches in Analyzing Unstructured Data)*, Cambridge University Press, 2007

[111] R. D. FINN, J. MISTRY, J. TATE, P. COGGILL, A. HEGER, J. E. POLLINGTON, O. L. GAVIN, P. GUNASEKARAN, G. CERIC, K. FORSLUND, L. HOLM, E. L. L. SONNHAMMER, S. R. EDDY, A. BATEMAN. *The Pfam protein families database*, in "Nucleic Acids Research", 2010, vol. 38, pp. D211–D222

[112] B. GANTER, S. O. KUZNETSOV. *Pattern Structures and Their Projections*, in "Conceptual Structures: Broadening the Base, Proceedings of the 9th International Conference on Conceptual Structures, ICCS 2001, Stanford, CA", H. DELUGACH, G. STUMME (editors), Lecture Notes in Computer Science 2120, Springer, 2001, pp. 129–142

[113] B. GANTER, R. WILLE. *Formal Concept Analysis*, SpringerBerlin, 1999

[114] A. GHOORAH, M.-D. DEVIGNES, M. SMAÏL-TABBONE, D. RITCHIE. *Spatial clustering of protein binding sites for template based protein docking*, in "Bioinformatics", August 2011, vol. 27, n^{o} 20, pp. 2820-2827 [*DOI :* 10.1093/BIOINFORMATICS/BTR493], http://hal.inria.fr/inria-00617921

[115] A. GHOORAH, M.-D. DEVIGNES, M. SMAÏL-TABBONE, D. RITCHIE. *Protein Docking Using Case-Based Reasoning*, in "Proteins", October 2013, vol. 81, n^{o} 12, pp. 2150-2158 [*DOI :* 10.1002/PROT.24433], http://hal.inria.fr/hal-00880341

[116] A. GHOORAH. *Extraction de Connaissances pour la Modelisation tri-dimensionnelle de l'Interactome Structural*, Université de Lorraine, November 2012, http://hal.inria.fr/tel-00762444

[117] H. HWANG, T. VREVEN, J. JANIN, Z. WENG. *Protein-protein docking benchmark version 4.0*, in "Proteins: Structure Function and Bioinformatics", 2010, vol. 78, n^{o} 15, pp. 3111–3114

[118] C. JONQUET, P. LEPENDU, S. FALCONER, A. COULET, N. NOY, M. A. MUSEN, N. H. SHAH. *NCBO Resource Index: Ontology-Based Search and Mining of Biomedical Resources*, in "Journal of Journal of Web Semantics", Sep 2011, vol. 9, n^{o} 3, pp. 316–324, NIH Projet NCBO [*DOI :* 10.1016/J.WEBSEM.2011.06.005], http://hal-lirmm.ccsd.cnrs.fr/lirmm-00622155

[119] M. KAYTOUE, S. O. KUZNETSOV, A. NAPOLI, S. DUPLESSIS. *Mining Gene Expression Data with Pattern Structures in Formal Concept Analysis*, in "Information Science", 2011, vol. 181, n^{o} 10, pp. 1989–2001

[120] M. KAYTOUE, F. MARCUOLA, A. NAPOLI, L. SZATHMARY, J. VILLERD. *The Coron System*, in "8th International Conference on Formal Concept Analsis (ICFCA) - Supplementary Proceedings", L. BOUMEDJOUT, P. VALTCHEV, L. KWUIDA, B. SERTKAYA (editors), 2010, pp. 55–58

[121] T. LEVIANDIER, A. ALBER, F. LE BER, H. PIÉGAY. *Comparison of statistical algorithms for detecting homogeneous river reaches along a longitudinal continuum*, in "Geomorphology", 2012, vol. 138, n^{o} 1, pp. 130-144 [*DOI :* 10.1016/J.GEOMORPH.2011.08.031], http://hal.inria.fr/hal-00640698

[122] Y. LIU, A. COULET, P. LEPENDU, N. H. SHAH. *Using ontology-based annotation to profile disease research*, in "Journal of the American Medical Informatics Association", June 2012, vol. 19, n^{o} e1, pp. e177-e186 [*DOI :* 10.1136/AMIAJNL-2011-000631], http://hal.inria.fr/hal-00752101

[123] G. MACINDOE, L. MAVRIDIS, V. VENKATRAMAN, M.-D. DEVIGNES, D. RITCHIE. *HexServer: an FFT-based protein docking server powered by graphics processors*, in "Nucleic Acids Research", May 2010, vol. 38, pp. W445-W449 [*DOI :* 10.1093/NAR/GKQ311]

[124] J.-F. MARI, F. LE BER, E.-G. LAZRAK, M. BENOÎT, C. ENG, A. THIBESSARD, P. LEBLOND. *Using Markov Models to Mine Temporal and Spatial Data*, in "New Fundamental Technologies in Data Mining", K. FUNATSU, K. HASEGAWA (editors), Intech, 2011, pp. 561–584, http://hal.inria.fr/inria-00566801/en

[125] F. PENNERATH, G. NIEL, P. VISMARA, P. JAUFFRET, C. LAURENÇO, A. NAPOLI. *A graph-mining method for the evaluation of bond formability*, in "ACS Journal of Chemical Information and Modeling", 2010, vol. 50, n^{o} 2, pp. 221–239

[126] V. PÉREZ-NUENO, D. RITCHIE, J. BORRELL, J. TEIXIDÓ. *Clustering and Classifying Diverse HIV Entry Inhibitors Using a Novel Consensus Shape-Based Virtual Screening Approach: Further Evidence for Multiple Binding Sites within the CCR5 Extracellular Pocket*, in "Journal of Chemical Information and Modeling", 2008, vol. 48, n^{o} 11, pp. 2146–2165

[127] C. RAÏSSI, J. PEI, T. KISTER. *Computing Closed Skycubes*, in "Proceedings of the VLDB Endowment", September 2010, vol. 3, n^{o} 1, pp. 838–847, http://hal.inria.fr/inria-00610923/en

[128] D. RITCHIE, A. GHOORAH, L. MAVRIDIS, V. VENKATRAMAN. *Fast Protein Structure Alignment using Gaussian Overlap Scoring of Backbone Peptide Fragment Similarity*, in "Bioinformatics", October 2012, vol. 28, n^{o} 24, pp. 3274-3281 [*DOI :* 10.1093/BIOINFORMATICS/BTS618], http://hal.inria.fr/hal-00756813

[129] D. RITCHIE, G. KEMP. *Protein Docking Using Spherical Polar Fourier Correlations*, in "Proteins: Structure, Function and Genetics", 2000, vol. 39, n^{o} 2, pp. 178–194

[130] D. RITCHIE, V. VENKATRAMAN. *Ultra-fast FFT protein docking on graphics processors*, in "Bioinformatics", 2010, vol. 26, n^{o} 19, pp. 2398–2405 [*DOI :* 10.1093/BIOINFORMATICS/BTQ444], http://hal.inria.fr/inria-00537988/en/

[131] M. ROUANE-HACENE, M. HUCHARD, A. NAPOLI, P. VALTCHEV. *Relational Concept Analysis: Mining Concept Lattices From Multi-Relational Data*, in "Annals of Mathematics and Artificial Intelligence", January 2013, vol. 67, n^{o} 1, pp. 81-108 [*DOI :* 10.1007/S10472-012-9329-3], http://hal.inria.fr/lirmm-00816300

[132] N. SCHALLER, E.-G. LAZRAK, P. MARTIN, J.-F. MARI, C. AUBRY, M. BENOÎT. *Combining farmers' decision rules and landscape stochastic regularities for landscape modelling*, in "Landscape Ecology", March 2012, vol. 27, n^{o} 3, pp. 433-446 [*DOI :* 10.1007/S10980-011-9691-2], http://hal.inria.fr/hal-00656407

[133] L. SZATHMARY. *Symbolic Data Mining Methods with the Coron Platform*, Université Henri Poincaré (Nancy 1), 2006

[134] Y. ZHANG, J. SKOLNICK. *TM-align: a protein structure alignment algorithm based on TM-score*, in "Nucleic Acids Research", 2005, vol. 33, n^{o} 7, pp. 2302–2309