IN PARTNERSHIP WITH:
CNRS

Université de Lorraine

Activity Report 2015

Project-Team ORPAILLEUR

Knowledge discovery, knowledge engineering

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

Project-Team ORPAILLEUR

Creation of the Project-Team: 2008 January 01

Keywords:

Computer Science and Digital Science:

3. - Data and knowledge
3.1.1. - Modeling, representation
3.1.7. - Open data
3.2.1. - Knowledge bases
3.2.2. - Knowledge extraction, cleaning
3.2.4. - Semantic Web
3.2.5. - Ontologies
3.3.2. - Data mining
3.3.3. - Big data analysis
3.4.1. - Supervised learning
3.4.2. - Unsupervised learning
4. - Security and privacy
4.1. - Threat analysis
8. - Artificial intelligence
8.1. - Knowledge
8.2. - Machine learning
8.6. - Decision support

Other Research Topics and Application Domains:

1. - Life sciences
1.1.2. - Molecular biology
1.1.5. - Genetics
1.1.6. - Genomics
1.2.1. - Biodiversity
2. - Health
2.3. - Epidemiology
3.1. - Sustainable development
9. - Society and Knowledge

1. Members

Research Scientists

Amedeo Napoli [Team leader, CNRS, Senior Researcher, HdR]
Esther Galbrun [Inria, Researcher, from Oct 2015]
Chedy Raïssi [Inria, Researcher]
Jean-Sébastien Sereni [CNRS, Researcher]
Yannick Toussaint [Inria, Researcher, HdR]

Faculty Members

Miguel Couceiro [Univ. Lorraine, Professor, 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, until August 2015, HdR]
Sébastien Da Silva [Univ. Lorraine, ATER]

Engineers

Sami Ghadfi [Inria, until Oct 2015]
Thi Nhu Nguyen Le [Inria, from Feb 2015]
Luis-Felipe Melo [Univ. Lorraine]
Matthieu Osmuk [Inria, until Sep. 2015]
Mickaël Zehren [Inria, granted by Bpifrance Financement]

PhD Students

Mehwish Alam [Univ. Lorraine, ATER]
Quentin Brabant [Univ. Lorraine]
Aleksey Buzmakov [Univ. Lorraine, ATER (until October 15)]
Victor Codocedo [Univ. Lorraine ATER and Engineer INSA Lyon (until September 15)]
Emmanuelle Gaillard [Univ. Lorraine]
Justine Reynaud [Inria]
Mohsen Sayed [Inria, granted by Conseil Régional de Lorraine]
My Thao Tang [Univ. Lorraine, granted by ANR HYBRIDE project]

Post-Doctoral Fellows

Dhouha Grissa [INRA]
Olfa Makkaoui [Inria, until Oct 2015]

Visiting Scientist

Luciano Grippo [Univ. Lorraine, Student, from Mar 2015 until Jun 2015]

Administrative Assistants

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

Others

Aurore Alcolei [ENS Lyon, Student, from Sep 2015]
Kévin Dalleau [Univ. Lorraine, Student, from Jun 2015 until Aug 2015]
Jane Hung [Inria, Student, from Apr 2015 until Jul 2015]
Benjamin Maurice [Inria, Student, from Mar 2015 until Jul 2015]
Yelen Per [CNRS, Student, from Sep 2015]
Kenny Rivalin [Univ. Lorraine, Student, from Mar 2015 until Aug 2015]
Alibek Sailanbayev [Inria, Student, from Jun 2015 until Jul 2015]
Daniel Vantroys [ENS Cachan, Student, until Sep 2015]

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. Knowledge Discovery guided by Domain Knowledge

Keywords: knowledge discovery in databases, knowledge discovery in databases guided by domain knowledge, data mining formal concept analysis, classification, pattern mining second-order Hidden Markov Models

Knowledge discovery in databases (KDD) is aimed at discovering patterns in large databases. These patterns can then be interpreted as knowledge units to be reused in knowledge systems. From an operational point of view, the KDD process is based on three main steps: (i) selection and preparation of the data, (ii) data mining, (iii) interpretation of the discovered patterns. The KDD process –as implemented in the Orpailleur team– is based on data mining methods which are either symbolic or numerical. Symbolic methods are based on pattern mining (e.g. mining frequent itemsets, association rules, sequences...), Formal Concept Analysis (FCA [93]) and extensions of FCA such as Pattern Structures [65] and Relational Concept Analysis (RCA [101]). Numerical methods are based on probabilistic approaches such as second-order Hidden Markov Models (HMM [98]), which are well adapted to the mining of temporal and spatial data.

Domain knowledge, when available, can improve and guide the KDD process, materializing the idea of *Knowledge Discovery guided by Domain Knowledge* or KDDK. In KDDK, domain knowledge plays a role at each step of KDD: the discovered patterns can be interpreted as knowledge units and reused for problem-solving activities in knowledge systems, implementing the operational sequence "mining, interpreting (modeling), representing, and reasoning". In this way, knowledge discovery appears as a core task in knowledge engineering, with an impact in various semantic activities, e.g. information retrieval, recommendation and ontology engineering. Moreover, it is used in application domains such as agronomy, astronomy, biology, chemistry, medicine. Accordingly, the Orpailleur team includes biologists, chemists, and a physician, making Orpailleur a very original team at Inria Nancy Grand Est.

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. 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 [93] (concept lattices are also known as "Galois lattices" [81]).

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 [103].

3.2. Text Mining

Keywords: text mining, knowledge discovery form collection of texts, annotation, ontology engineering from texts

The objective of a text mining process is to extract useful knowledge units from large collections of texts [90]. 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. 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".

3.3. Knowledge Systems and Web of Data

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

The web of data constitutes a good platform for experimenting ideas on knowledge engineering and knowledge discovery, in relation with the principles of semantic web. 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: this is why domain knowledge and ontologies are of main importance. The knowledge representation language recommended by W3C to design ontologies and knowledge bases is OWL, which is based on description logics (DLs [79]). In OWL, knowledge units are represented by classes (or concepts) having properties (attributes) and instances. Concepts are organized within a partial order based on a subsumption relation, and the inference services are based on classification-based reasoning and case-based reasoning (CBR).

Actually, there are many interconnections between concept lattices in FCA and ontologies, e.g. the partial order underlying an ontology can be supported by a concept lattice. Moreover, a pair of implications within a concept lattice can be adapted for designing concept definitions in ontologies. Accordingly, we are interested here in two main challenges: how the web of data, as a set of potential knowledge sources (e.g. DBpedia, Wikipedia, Yago, Freebase...) can be mined for helping the design of definitions and knowledge bases and how knowledge discovery techniques can be applied for providing a better usage of the web of data (e.g. LOD classification).

Accordingly, a part of the research work in Knowledge Engineering is oriented towards knowledge discovery in the web of data, as, with the increased interest in machine processable data, more and more data is now published in RDF (Resource Description Framework) format. Particularly, we are interested in the completeness of the data and their potential to provide concept definitions in terms of necessary and sufficient conditions [1]. We have proposed a novel technique based on FCA which allows data exploration as well as the discovery of definition (bidirectional implication rules).

4. Application Domains

4.1. Biology and Chemistry

Participants: Mehwish Alam, Aleksey Buzmakov, Adrien Coulet, Nicolas Jay, Amedeo Napoli, 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 and Formal Concept Analysis can also be used in an efficient and well-founded way [34]. Combining supervised methods –with a training set 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, 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 researchers in computer science (Orpailleur) and experts in oncology are participating. 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: Emmanuelle Gaillard, Jean Lieber, 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.

4.4. Agronomy

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

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

Sébastien da Silva has defended his PhD thesis [87] in September 2014. This research was conducted in the framework of an Inria-INRA collaboration, taking 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 [88].

An on-going research work about the representation of peasant knowledge is involved within a collaboration with IRD in Madagascar [94]. Sketches drawn by peasants were transformed into graphs and compared thanks to Formal Concept Analysis.

4.5. Digital Humanities

Participant: Jean Lieber.

Keywords: digital humanities, semantic web, SPARQL, approximate search, case-based reasoning

Recent contacts with the digital humanity community have occurred, in particular, with a group of researchers in the domain of the history and philosophy of science and technologies (located in Brest, Montpellier and Nancy) willing to benefit from semantic Web technologies in order to provide better accesses to their corpora. A first paper based on this starting collaboration has been published [51], in which we proposed an approach to exact and approximate search in RDFS-annotated corpora based on the SPARQL technology and on case-based reasoning principles.

5. Highlights of the Year

5.1. Highlights of the Year

- Aleksey Buzmakov was nominated at the 13th International Conference on Formal Concept Analysis (ICFCA, Nerja Málaga, Spain, June 23-26 2015) as the "best promising researcher in Formal Concept Analysis" and won the best student paper award [53].
- Two (very) young researchers have made a stay in the team, Artuur Leeuwenberg in Spring 2014 and Alibek Sailanbayev in Spring 2015. Both young researchers have done a very good work which was rewarded by two conference publications, [66] and [46]. The Orpailleur team is particularly proud of the very good results of these young researchers.
- Three PhD students, namely Alam Mehwish, Aleksey Buzmakov and Victor Codocedo, have joined their efforts in their last year of thesis preparation for working on a common topic, the completion of web of data. This very good and very uncommon research work was rewarded by a publication in the very highly selective IJCAI 2015 Conference [1].
- The paper "Miguel Couceiro, Lucien Haddad, Karsten Schölzel, Tamas Waldhauser. Relation graphs and partial clones on a 2-element set. 44th IEEE International Symposium on Multiple-Valued Logic (ISMVL 2014), IEEE Computer Society, 161-166." was awarded the "Outstanding Contributed Paper Award" at the conference ISMVL 2015 (IEEE Computer Society).
- The Taaable system won 3 of the 5 prizes of the 8th "Computer Cooking Contest", which was held during the International Conference on Case-Based Reasoning, in Bad Homburg, Germany (http://ccc2015.loria.fr/?id=rules): the prize of the best cocktail system according to the jury, based on the technical/scientific paper reviews and on the comparison of the results of the systems on a same set of queries, the prizes of the public for the cocktail and sandwich systems, based on the vote after tasting.

5.1.1. Awards

BEST PAPER AWARD:

[53]
A. BUZMAKOV, S. O. KUZNETSOV, A. NAPOLI. *Revisiting Pattern Structure Projections*, in "International Conference in Formal Concept Analysis - ICFCA 2015", Nerja, Spain, J. BAIXERIES, C. SACAREA, M. OJEDA-ACIEGO (editors), Lecture Notes in Computer Science, Springer International Publishing, June 2015, vol. 9113, pp. 200–215 [*DOI :* 10.1007/978-3-319-19545-2_13], https://hal.archives-ouvertes.fr/hal-01186719

6. New Software and Platforms

6.1. Symbolic KDD Systems

6.1.1. The Coron Platform

- Contact: Amedeo Napoli
- URL: http://coron.loria.fr/site/index.php
- KEYWORDS: Data mining, Closed itemset, Frequent itemset, Generator, Association rule, Rare itemset

FUNCTIONAL DESCRIPTION.

The Coron platform [102], [96] 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 [103]. 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.

6.1.2. Orion: Skycube Computation Software

- Contact: Chedy Raissi
- URL: https://github.com/leander256/Orion
- KEYWORDS: Skyline, skycube.

FUNCTIONAL DESCRIPTION.

This program implements the algorithms described in a research paper published at VLDB 2010 [100]. 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.

6.1.3. OrphaMine – Data mining platform for orphan diseases

- Partners: INSERM - MoDYCo CNRS - Délégation régionale Ile-de-France, secteur ouest et nord - Greyc Université de Caen - Basse-Normandie
- Contact: Chedy Raissi
- URL: http://webloria.loria.fr/~mosmuk/orphamine/
- KEYWORDS: Bioinformatics, data mining, biology, health, data visualization, drug development.

FUNCTIONAL DESCRIPTION.

The OrphaMine platform, developed as part of the ANR Hybrid project, enables visualization, data integration and in-depth analytics. The data at the heart of the platform is about orphan diseases and is extracted from the OrphaData ontology (http://www.orpha.net).

We aim to build a true collaborative portal that will serve the different actors of the Hybrid project: (i) A general visualization of OrphaData data for physicians working, maintaining and developing this knowledge database about orphan diseases. (ii) The integration of analytics (data mining) algorithms developed by the different academic actors. (iii) The use of these algorithms to improve our general knowledge of rare diseases.

6.1.4. PoQeMON Analytics: Platform for Quality Evaluation of Mobile Netwoks

- Partners: Altran, DataPublica, GenyMobile, HEC, Inria Nancy-Grand Est, IP-Label, Next Interactive Media, Orange, Université Paris-Est Créteil
- Contact: Chedy Raissi
- URL: https://members.loria.fr/poqemon/
- KEYWORDS: Data mining, data visualization.

FUNCTIONAL DESCRIPTION.

PoQEMoN is a quality evaluation platform for mobile phone networks. The quality measures include the coverage, availability and network performances. Multiple methods are implemented in this platform, either in visualization or in data anonymization to make on-line analytics as simple as possible.

6.2. Stochastic systems for knowledge discovery and simulation

6.2.1. The CarottAge System

- Contact: Jean-François Mari
- URL: http://www.loria.fr/~jfmari/App/index_in_english.html
- KEYWORDS: Stochastic process, Hidden Markov Models.

FUNCTIONAL DESCRIPTION.

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 [97]. 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 proved to give very interesting results for that purpose.

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.

6.2.2. The ARPEnTAge System

- Contact: Jean-François Mari
- URL: http://www.loria.fr/~jfmari/App/index_in_english.html
- KEYWORDS: Stochastic process, Hidden Markov Models.

FUNCTIONAL DESCRIPTION.

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 [98]. 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.

6.2.3. *The GenExp System*

- Contact: Florence Le Ber
- URL: http://orpailleur.loria.fr/index.php/GenExp-LandSiTes:_KDD_and_simulation
- KEYWORDS: Simulation, Hidden Markov Models.

FUNCTIONAL DESCRIPTION.

In the framework of the project "Impact des OGM" initiated by the French Ministry of Research, we have developed a software called GenExp-LandSiTes for simulating bidimensional random landscapes, and then studying the dissemination of vegetable transgenes. The GenExp-LandSiTes system is linked to the CarottAge system, and is based on computational geometry and spatial statistics. The simulated landscapes are given as input for programs such as "Mapod-Maïs" or "GeneSys-Colza" for studying the transgene diffusion. Other landscape models based on tessellation methods are under studies. The last version of GenExp allows an interaction with R and deals with several geographical data formats.

6.3. KDD systems in Biology

6.3.1. *IntelliGO Online*

- Contact: Malika Smaïl-Tabbone
- URL: http://plateforme-mbi.loria.fr/intelligo/
- KEYWORDS: Bioinformatics, genomics.

FUNCTIONAL DESCRIPTION.

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 [82]. 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.

6.3.2. *WAFOBI: KNIME Nodes for Relational Mining of Biological Data*

- Contact: Malika Smaïl-Tabbone
- KEYWORDS: Bioinformatics, genomics.

FUNCTIONAL DESCRIPTION.

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.

6.3.3. *MODIM: MOdel-driven Data Integration for Mining*

- Contact: Malika Smaïl-Tabbone
- URL: https://gforge.inria.fr/projects/modim/
- KEYWORDS: Data integration, workflow, data modeling.

FUNCTIONAL DESCRIPTION.

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.

6.4. Knowledge Systems in Health and Cooking

6.4.1. The Kasimir System for Decision Knowledge Management

- Contact: Jean Lieber
- KEYWORDS: Classification-based reasoning, case-based reasoning, decision knowledge management, knowledge edition, knowledge base maintenance, semantic portal

FUNCTIONAL DESCRIPTION.

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 [86], 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 §6.1.1).

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/).

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

- Contact: Emmanuel Nauer
- URL: http://intoweb.loria.fr/taaable3ccc/
- KEYWORDS: Knowledge acquisition, ontology engineering, semantic annotation, case-based reasoning, hierarchical classification, text mining.

FUNCTIONAL DESCRIPTION.

The objectives of the Taaable system are to retrieve textual cooking recipes and to adapt these retrieved recipes whenever needed [84]. 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://intoweb.loria.fr/taaable3ccc/, and is constantly evolving. Since 2014, Taaable is based on Tuuurbine, a generic ontology guided CBR engine over RDFS, and Revisor, an adaptation engine implementing various revision operators. This year, new features have been added to the Taaable system in order to address the new challenges of the 8th Computer Cooking Contest at ICCBR 2015. Firstly, FCA was used to improve the ingredient substitution, by taking into account ingredient combinations in a large set of recipes. Secondly, an approach based on mixed linear optimization has been used to adapt ingredient quantities, in order to be more realistic with a real cooking setting.

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

- Contact: Emmanuel Nauer
- URL: http://tuuurbine.loria.fr/
- KEYWORDS: case-based reasoning, inference engine, knowledge representation, ontology engineering, semantic web

FUNCTIONAL DESCRIPTION.

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 [91]. 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.

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

- Contact: Emmanuel Nauer
- URL: https://github.com/kolflow/begoood
- KEYWORDS: Tests, non-regression, knowledge evolution.

FUNCTIONAL DESCRIPTION.

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. BeGoood is used by the Taaable system for managing the evolution of the knowledge base used by the CBR system.

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

- Contact: Jean Lieber
- URL: http://revisor.loria.fr/
- KEYWORDS: Belief revision, adaptation, revision-based adaptation, case-based reasoning, inference engines, knowledge representation.

FUNCTIONAL DESCRIPTION.

Revisor is a library of inference engines dedicated to belief revision and to revision-based adaptation for case-based reasoning. 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, conjunction of linear constraints, and qualitative algebras [89]). Some of these engines are already used in the Taaable system. Current developments on Revisor aim at defining new engines in other formalisms. In particular, a study on a revision operator in the propositional closure of linear constraints (with integer and real number variables) has been carried out [70]: definition, properties and algorithm.

7. New Results

7.1. The Mining of Complex Data

Participants: Mehwish Alam, Aleksey Buzmakov, Victor Codocedo, Miguel Couceiro, Adrien Coulet, Esther Galbrun, Nicolas Jay, Florence Le Ber, 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

Pattern mining and Formal Concept Analysis are suitable symbolic methods for KDDK, that may be used for real-sized applications. Global improvements are carried out 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.

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

Advances in data and knowledge engineering have emphasized the needs for pattern mining tools working on complex data. In particular, FCA, which usually applies to binary data-tables, can be adapted to work on more complex data. In this way, we have contributed to two main extensions of FCA, namely Pattern Structures and Relational Concept Analysis. Pattern Structures (PS [92]) allow to build a concept lattice from complex data, e.g. numbers, sequences, trees and graphs. Relational Concept Analysis (RCA) is able to analyze objects described both by binary and relational attributes [101] and can play an important role in text classification and text mining. Following this way, and regarding itemset and association rule discovery, we improved standard algorithms for building lattices from large data and for completing the algorithm collection of the Coron platform [103].

Many developments were carried out in pattern mining and FCA for improving data mining algorithms and their applicability, and for solving some specific problems such as information retrieval, discovery of functional dependencies and biclustering. We designed new information retrieval methods based on FCA where the concept lattice is considered as an index space for answering disjunctive queries [54]. We developed also a whole line of work on pattern structures for the discovery of functional dependencies [80], text classification and heterogeneous pattern structures [83], and pattern structures for structured attribute sets [46]. FCA can also be considered as a clustering method and we adapted pattern structures to clustering for analyzing numerical datatables supporting recommendation problems [13]. Projections can be associated with pattern structures for leveraging the volume and the complexity of the computation [53]. We designed also a quasi-polynomial algorithm for mining top patterns w.r.t. measures satisfying special properties in a FCA framework [52]. We also proposed new visualization techniques and tools able to display important and useful information (e.g. stable concepts) from large concept lattices [49].

Still considering complex data, we worked on the analysis of molecular structures (or molecular graphs) [34]. The mining of molecular graphs is an important task for many reasons, among which the challenges it represents regarding knowledge discovery, life sciences and healthcare, and, as well, the industrial needs that can be met whenever substantial results are obtained (especially in pharmacology).

7.1.2. 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 or at least "ontological concept definitions" 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 [66]. In addition, in his thesis work, Mohsen Sayed focused on extracting relations between named entities using graph mining methods applied to dependency graphs. 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 [64].

The notion of "Jumping Emerging Patterns" (JEP) previously used in chemistry [12], was updated and adapted to 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. as a positive example, or as a simple word, i.e. as a negative example. The method extracts from a training set all JEPs which are considered as hypotheses [7]. To reduce the number of JEPs and to only retain the most significant from a linguistic point of view, JEPs are weighted and a constraint solver is used to check the maximal coverage of the positive examples. Results are currently under evaluation.

7.1.3. Mining Sequences and Trajectories

Sequence data is widely used in many applications. Computing the similarity between sequences is a very important challenge for many different data mining tasks. There is a plethora of similarity measures for sequences in the literature, most of them being designed for sequences of items. In a recent work with Elias Egho, we study the problem of measuring the similarity between sequences of itemsets [32]. We focus on the notion of common subsequences as a way to measure similarity between a pair of sequences composed of a list of itemsets. In this work, we present new combinatorial results for efficiently counting distinct and common subsequences. These theoretical results are the cornerstone of an effective dynamic programming approach to deal with this problem. In addition, we develop an approximate method to speed up the computation process for long sequences. We have applied the method to various data sets: healthcare trajectories, on-line handwritten characters and synthetic data. The results confirm that the current similarity measure produces competitive scores and indicate that the method is relevant for large scale sequential data analysis.

Nowadays data sets are available in very complex and heterogeneous ways. Mining of such data collections is essential to support many real-world applications ranging from healthcare to marketing. In a recent work, we focused on the analysis of "complex sequential data" by means of interesting sequential patterns [19]. We approach the problem using FCA and pattern structures, where the subsumption relation ordering patterns is defined w.r.t. the partial order on sequences. We show how pattern structures along with projections, i.e. a data reduction of sequential structures, are able to enumerate more meaningful patterns and increase the computing efficiency of the approach. Finally, we demonstrate the applicability of the method for discovering and analyzing patient patterns from a French healthcare data set on cancer. The quantitative and qualitative results –with annotations and analysis from a physician– are reported in this use case which is one main motivation for this work.

7.1.4. Mining with Preferences

In the last decade, the pattern mining community has witnessed a sharp shift from efficiency-based approaches to methods which can extract more meaningful patterns. Recently, new methods adapting results from studies of economic efficiency and multi-criteria decision analysis such as Pareto efficiency, or skylines, have been studied. Within pattern mining, this novel line of research allows the easy expression of preferences according to a dominance relation. We have developed approaches that are useful from a user-preference point of view, tending to promote the use of pattern mining algorithms for non-experts. These approaches are based on the discovery of skyline patterns, or "skypatterns", in relation with condensed representations of patterns. This last relationship facilitates the computation of skypatterns, providing a flexible and efficient approach to mine skypatterns reusing a dynamic constraint satisfaction problems (CSP) framework [8].

7.1.5. Aggregation

Aggregation or consensus theory studies any process dealing the merging of several objects (numerical values, qualitative data, preferences, etc.) into a single (or several) object of similar type and that, in some way, is the best representation. The need to aggregate objects in a meaningful way has become more and more present in an increasing number of areas not only of mathematics, statistics or physics, but especially in applied fields such as engineering, computer science, social sciences and biology. In social choice and multicriteria decision aid, objects are preferences that are expressed by users, voters or criteria, and are modeled by order relations or utility functions. In cluster analysis, the objects to merge are classifications (such as partitions, hierarchies or trees) or related functions (such as similarity/dissimilarity measures).

With the proliferation of massive databases and new fields such as computational advertising, search engines and recommender systems, the need for information retrieval and knowledge discovery processes became emergent as well as the construction of user preference models for classification and prediction purposes. Also in biology and phylogenetics, aggregation is used to find consensus patterns among DNA sequences or finding consensus trees within taxonomies. As algorithms are often heuristic in such large datasets, they rarely produce the same output, highlighting the importance of finding means of aggregation to produce consensus structures. The difficulty in extracting such consensus structures comes down to define appropriate aggregation rules (e.g., counting and median procedures), and their impossibility is many times revealed by Arrowian results. A way to avoid such impossibility results is the consideration of alternative aggregation rules or the weakening of underlying structures, for instance weak hierarchies that allow overlapping clusters while keeping desirable tree-like properties.

We are working on a theoretical basis of a unified theory of consensus and to set up a general machinery for the choice and use of aggregation functions. This choice depends on properties specified by users or decision makers, the nature of the objects to aggregate as well as computational limitations due to prohibitive algorithmic complexity. This problem demands an exhaustive study of aggregation functions that requires an axiomatic treatment and classification of aggregation procedures as well as a deep understanding of their structural behavior. Moreover, Arrowian results are also envisioned since they constitute an important tool in the identification of reasonable algebraic/relational structures for representing data as well as in the identification of meaningful aggregation processes.

Direct applications of this theory are preference learning and cluster analysis. In the first case, preferences are represented by global utility functions and alternatives with higher utilities are preferred. Moreover, simplified versions of this model will be explored in the context of feature selection for both dimension reduction of data as well as classifier design. In the second case, we consider median structures that include several ordered/relational structures (trees, graphs, orders) and that allow several consensus procedures. This is particularly useful in a context of classification that takes into account evolutionary relations between classes, for instance, in taxonomical biology and phylogenetics.

7.1.6. Video Game Analytics

The video game industry has enormously grown over the last twenty years, bringing new challenges to the artificial intelligence and data analysis communities. We are studying the 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. Continuing our joint collaboration with researchers from the MIT GameLab we successfully extended our previous work to a journal paper that will be published in 2016.

7.2. Knowledge Discovery in Healthcare and Life Sciences

Participants: Miguel Couceiro, Adrien Coulet, Amedeo Napoli, Chedy Raïssi, Mohsen Sayed, Malika Smaïl-Tabbone, Yannick Toussaint.

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 healthcare, 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.

7.2.1. Ontology-based Clustering of Biological Linked Open Data

Increasing amounts of biomedical data provided as Linked Open Data (LOD) offer novel opportunities for knowledge discovery in bio-medicine. We proposed an approach for selecting, integrating, and mining LOD with the goal of discovering genes responsible for a disease [99]. We are currently working on the integration of LOD about known phenotypes and genes responsible for diseases along with relevant bio-ontologies. We are also defining a corpus-based semantic distance. One possible application of this work is to build and compare possible diseaseomes, i.e. global graphs representing all diseases connected according to their pairwise similarity values.

7.2.2. Suggesting Valid Pharmacogenes by Mining Linked Open Data and Electronic Health Records

A standard task in pharmacogenomics research is identifying genes that may be involved in drug response variability and called "pharmacogenes". As genomic experiments in this domain tend to generate many false positives, computational approaches based on background knowledge may generate more valuable results. Until now, the later have used only molecular networks databases or biomedical literature. We are studying and working on a novel method that take advantage of an eclectic set of linked data sources to validate uncertain drug–gene relationships, i.e. pharmacogenes [3]. One advantage relies on the standard implementation of linked data that facilitates the joint use of various sources and makes easier the consideration of features of various origins. Accordingly, we proposed an initial selection of linked data sources relevant to pharmacogenomics. We formatted these data to train a random forest algorithm, producing a model that classify drug–gene pairs as related or not, thus validating candidate pharmacogenes.

With this same motivation of validating state-of-the-art knowledge in pharmacogenomics, a new ANR project called "PractiKPharma" will be initiated in 2016 and will rely on similar ideas. The originality of "PractiKPharma" is to use "Electronic Health Records" to constitute cohorts of patients that are then mined for validating extracted pharmacogenomics knowledge units (http://practikpharma.loria.fr/).

7.2.3. Biological Data Aggregation for Knowledge Discovery

During this year, in collaboration with the Capsid Team, we contributed to write up two multi-disciplinary projects with a group of clinicians from the Regional University Hospital (CHU Nancy) and bio-statisticians from the Maths Lab (IECL). The first project, entitled ITM2P [1] lying in the so-called CPER 2015–2020 framework, was accepted and granted. The funding is mainly intended for medical and computing equipments and will be used to set up four scientific platforms. We are involved in the SMEC platform as a support for "Simulation, Modeling and Knowledge Extraction from Bio-Medical Data".

[1] "Innovations Technologiques, Modélisation et Médecine Personnalisée"

The second project is a RHU [2] project entitled *Fight Heart Failure* (FHF) and was accepted as a so-called "investissement d'avenir" and granted. We are in charge of a workpackage which will give us the opportunity of exploring important research questions. Among these questions, one is to define "data aggregation" mechanisms with a twofold objective: (i) the definition of pairwise patient similarity given that patients are described by complex dimensions involving relations and time and (ii) the efficient clustering of patients based on this similarity measure. Each cluster should correspond to a bioprofile, i.e. a subgroup of patients sharing the same form of the disease and thus the same diagnosis and care strategy. For doing that, we are currently investigating consensus theories [95] and their applicability to a bio-medical context, and as well aggregation operators as defined in various contexts, e.g. databases, data-warehouses, web of data, and graph theory. The idea is to consider relational and temporal data aggregation as a first class citizen in the data preparation phase of the knowledge discovery. This allows to assess the contribution of aggregation for such a task and in this context.

Another question is related to the construction of a prediction model for each bioprofile/subgroup –once validated by the clinicians– to be used in a decision support system. This will likely require the combination of symbolic and numerical methods for the classification task.

7.2.4. 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. 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. One main challenge is to mine these annotations.

As a first attempt, we adapted pattern structures to analyze the annotations of biomedical databases [85]. We considered annotated biomedical documents as objects and the corresponding annotations were classified according to various dimensions, i.e. a particular aspect of domain knowledge. The resulting classification of annotations allowed not only to discover correlations between annotations but also incomplete annotations that could be fixed afterward. This adaptation of pattern structures opens many perspectives in term of ontology reengineering and knowledge discovery.

7.3. Knowledge Engineering and Web of Data

Participants: Mehwish Alam, Aleksey Buzmakov, Victor Codocedo, Emmanuelle Gaillard, Florence Le Ber, Jean Lieber, Amedeo Napoli, Emmanuel Nauer.

Keywords: knowledge engineering, web of data, classification-based reasoning, case-based reasoning, belief revision, semantic web

7.3.1. Around the Taaable Research Project

The Taaable project was originally created as a challenger of the Computer Cooking Contest (ICCBR Conference) [84] (http://intoweb.loria.fr/taaable3ccc/). 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.

[2] "Recherche Hospitalo-Universitaire"

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 notions such as belief, trust, reputation and quality, as well as their relationships and rules to manage the 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 [61].

Another study shows how the case retrieval of a CBR system can be improved using typicality. Typicality discriminates subclasses of a class in the domain ontology depending of how a subclass is a good example for its class. An approach has been proposed to partition the subclasses of some classes into atypical, normal and typical subclasses in order to refine the domain ontology. The refined ontology allows a finer-grained generalization of the query during the retrieval process, improving at the same time the final results of the CBR system [62].

The Taaable system also includes a module for adapting textual preparations (from a source recipe text to an adapted recipe text, through a formal representation in the qualitative algebra INDU). The evaluation of this module as a whole thanks to users has been carried out and has shown its efficiency (w.r.t. text quality and recipe quality), when compared with another approach to textual adaptation [4].

FCA allows to organize objects according to the properties they share into a concept lattice. A lattice has been built on a large set a cooking recipes according to the ingredients they use, producing a hierarchy of ingredient combinations. When a recipe R has to be adapted, this lattice can be used to search the best ingredient combinations in the concepts that are the closest to the concept representing R [63].

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 adjust the ingredient quantities using engines included in the Revisor library (see § 6.4.5). This year, a mixed linear optimization has implemented to produce human easy understandable quantities. For example, when the ingredient is a lemon, its quantity will take the form of a quarter, a half, etc., instead of *54 g* (which corresponds to a half lemon) [63].

7.3.2. Exploring and Classifying the Web of Data

A part of the research work in Knowledge Engineering is oriented towards knowledge discovery in the web of data, as, with the increased interest in machine processable data, more and more data is now published in RDF (Resource Description Framework) format. The popularization and quick growth of Linked Open Data (LOD) has led to challenging aspects regarding quality assessment and data exploration of the RDF triples that shape the LOD cloud. Particularly, we are interested in the completeness of the data and the their potential to provide concept definitions in terms of necessary and sufficient conditions [1]. We have proposed a novel technique based on Formal Concept Analysis which organizes subsets of RDF data into a concept lattice. This allows data exploration as well as the discovery of implication rules which are used to automatically detect missing information and then to complete RDF data and to provide definitions. Moreover, this is also a way of reconciling syntax and semantics in the LOD cloud. Experiments on the DBpedia knowledge base shows that this kind of approach is well-founded and effective.

Other important aspects are concerned with data access, data visualization w.r.t. the SPARQL query language [46], [49]. SPARQL queries over the web of data usually produce lists of tuples as answers that may be voluminous and hard to interpret. We introduced Lattice-Based View Access (LBVA), a framework based on FCA, which provides a classification of the answers of SPARQL queries based on a concept lattice. This concept lattice can be considered as a materialized view of the data resulting from a SPARQL query and can be navigated for retrieving or mining specific patterns. We associate a VIEW-BY clause to SPARQL for facilitating the interaction between analysts and LOD. The organization of answers is based on an original proposition on pattern structures for structured sets of attributes, which appears to be quite efficient and very well-adapted to the classification and analysis of RDF data. The visualization and the navigation of the concept lattice are guided by RV-Xplorer (i.e. RDF View eXplorer), an adapted interactive visualization system. Experiments show that the approach is well-founded and that it opens many new perspectives in the domain.

7.4. Advances in Graph Theory

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

Keywords: graph theory, extremal graph theory, chromatic number, triangle-free graph, planar graph, graph coloring

We announced in the last report that we started to work on a conjecture by Heckman and Thomas from 1999. We managed to confirm the conjecture and the demonstration was published in January 2014. A classical result by Staton, from 1979, states that every triangle-free graph G with maximum degree at most 3 contains an independent set of order at least $5n/14$, where n is the number of vertices of G. Heckman and Thomas conjectured a stronger fact: the fractional chromatic number of such a graph is at most $14/5$. We confirmed their conjecture by establishing the following stronger assertion: for any assignment of weights (i.e., real numbers) to the vertices of such a graph G, there exists an independent set I such that the weights of the vertices in I is at least $5/14$ times the total weight of the G.

Exploring further the methods we introduced to solve this conjecture, we obtained new results concerning the fractional chromatic number of planar triangle-free graphs. While the fractional chromatic number of such graphs is at most 3 (because their chromatic number is), a construction of Jones proved the existence of triangle-free planar graphs with fractional chromatic number arbitrarily close to 3. Thus one wonders whether there could be such graphs with fractional chromatic number exactly 3. We demonstrated this not to be the case, by proving a general upper bound of $\frac{9n}{3n+1} = 3(1 - \frac{1}{3n+1})$ for every triangle-free planar graph G with n vertices. This bound is qualitatively the best possible: Jones's construction yields graphs with fractional chromatic number $3 - \frac{c}{n}$ for some constant c. In addition, a tight bound was obtained if the graphs considered are furthermore required to have maximum degree at most 4. In this case, the bound becomes $\frac{3n}{3n+1}$.

Motivated by frequency assignment in office blocks, we study the chromatic number of the adjacency graph of a 3-dimensional parallelepiped arrangement. In the case each parallelepiped is within one floor, a direct application of the Four-Colour Theorem yields that the adjacency graph has chromatic number at most 8. We provide an example of such an arrangement needing exactly 8 colors. We also discuss bounds on the chromatic number of the adjacency graph of general arrangements of 3-dimensional parallelepipeds according to geometrical measures of the parallelepipeds (side length, total surface area or volume).

8. Partnerships and Cooperations

8.1. International Initiatives

8.1.1. Inria International Labs:SNOWFLAKE

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

Inria@SiliconValley

Associate Team involved in the International Lab: SNOWFLAKE

Title: Knowledge Discovery from Linked Data and Clinical Notes

International Partner (Institution - Laboratory - Researcher):

Stanford (United States) - Department of Medicine, Stanford Center for Biomedical Informatics Research (BMIR) - Nigam Shah

Start year: 2014

See also: http://snowflake.loria.fr/

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 would help 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. (adverse) reactions to drugs. Identified 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 showing a specific reaction profile to a given group of drugs.

8.1.2. Participation In other International Programs: 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.

8.1.3. STIC AmSud: Autonomic Knowledge Discovery (AKD)

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

This research project involves researchers with different specialties, from Brazil (Universidade Federal Rio Grande do Sul), from Chile (UFSM Santiago and Valparaiso), from Uruguay (Universidad de la Repùblica), and the Orpailleur Team. The projects targets the design of solutions able to proactively understand the behavior of systems and networks in order to prevent vulnerable states. Accordingly, we aim at integrating knowledge discovery techniques within autonomic systems in order to provide intelligent self-configuration and self-protection mechanisms. The results of this project may not only benefit to end-users but also highly contribute to the scientific community by providing solid foundations for the development of more secure, scalable, and reliable management approaches.

8.1.4. Miscellaneous

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

- 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 fifth edition should take place this year in August at ECAI 2016 (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.

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 the Orpailleur team is involved, 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. 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.

8.2.2. FUI PoQemon

Participants: Matthieu Osmuk, Chedy Raïssi [Contact Person], Mickaël Zehren.

The PoQemon project aims at developing new pattern mining methods and tools for supporting privacy preserving knowledge discovery from monitoring purposes on mobile phone networks. The main idea is to develop sound approaches that handle the trade-off between privacy of data and the power of analysis. Original approaches to this problem were based on value perturbation, damaging data integrity. Recently, value generalization has been proposed as an alternative; still, approaches based on it have assumed either that all items are equally sensitive, or that some are sensitive and can be known to an adversary only by association, while others are non-sensitive and can be known directly. Yet in reality there is a distinction between sensitive and non-sensitive items, but an adversary may possess information on any of them. Most critically, no antecedent method aims at a clear inference-proof privacy guarantee. In this project, we integrated the ρ-uncertainty privacy concept that inherently safeguards against sensitive associations without constraining the nature of an adversary's knowledge and without falsifying data. The project integrates the ρ-uncertainty pattern mining approach with novel data visualization techniques.

The PoQemon research project involves the following partners: Altran, DataPublica, GenyMobile, HEC, IP-Label, Next Interactive Media, Orange and Université Paris-Est Créteil, along with Inria Nancy Grand Est.

8.2.3. PEPS

8.2.3.1. PEPS Approppre

Participants: Mehwish Alam, Quentin Brabant, Aleksey Buzmakov, Victor Codocedo, Miguel Couceiro [Contact Person], Adrien Coulet, Esther Galbrun, Amedeo Napoli, Chedy Raïssi, Yannick Toussaint.

This PEPS Approppre research project (see http://www.cnrs.fr/ins2i/spip.php?article1183) is aimed at setting a framework for characterizing the mining of preferences in massive data. Such a unified framework for the mining of qualitative preferences is not yet existing and can be related to recent studies in decision theory (aggregation models and consensus), machine learning and data mining. A particular focus will be done on the aggregation model of Sugeno integral which can be applied on a symbolic representation of preferences for two main operations, reduction of dimensionality (feature selection) and prediction.

8.2.3.2. PEPS Confocal

Participants: Adrien Coulet, Amedeo Napoli, Chedy Raïssi, Malika Smaïl-Tabbone.

The Confocal Project (see http://www.cnrs.fr/ins2i/spip.php?article1183) is interested in the design of new methods in bioinformatics for analyzing and classifying heterogeneous omics data w.r.t. biological domain knowledge. We are planning to adapt FCA and pattern structures for discovering patterns and associations in gene data with the help of domain ontologies. One important objective of the project is to check whether such a line of research could be reused on so-called discrete models in molecular biology.

8.2.3.3. PEPS Prefute

Participants: Mehwish Alam, Quentin Brabant, Aleksey Buzmakov, Victor Codocedo, Adrien Coulet, Miguel Couceiro [Contact Person], Esther Galbrun, Amedeo Napoli, Chedy Raïssi, Mohsen Sayed, Malika Smaïl-Tabbone, My Thao Tang, Yannick Toussaint.

The PEPS Prefute project is mainly interested in interaction and iteration in the knowledge discovery (KD) process. Usually the KD process is organized around three main steps which are (i) selection and preparation of the data, (ii) data mining, and (iii) interpretation of (selected) resulting patterns. For leading such a process, which actually is a loop, an analyst who is most of the time an expert of the data domain, is present. This materializes the fact that the KD process requires interaction and iteration. However, it appears that until recently the most important progress were made on the second step of the KD process, i.e. data mining, and especially form the algorithmic point of view. This gave birth to a variety of efficient and fast algorithms. This second step is in between the two other steps whose importance is now becoming very clear as the analyst is facing very large amounts of data and even larger amounts of resulting patterns. Actually, KDDK is one possible way of tackling such a problem as the principle is to push domain knowledge for improving the KD process.

Accordingly, the PEPS Prefute project is interested in the study of interactions between the analyst and the KD process, i.e. pushing constraints, preferences and domain knowledge, for guiding and improving the KD process. One possible way is to discover some original and generic pattern which can be considered as a reference for going farther and to search the pattern space w.r.t. this original pattern linked to some preferences of the analyst. In this way, the interesting pattern space is much more concise and of much lower size. Moreover, the PEPS Prefute project contributes also to consolidate the place of the analyst in the KD process. In particular this means that more studies have to be carried out on the possible interactions with the analyst and on the importance of preferences and domain knowledge in this interaction. In addition, visualization tools associated to KD systems have to be improved for being able to work with the actual large amounts of data and patterns as well (see https://www.greyc.fr/fr/node/2207).

8.3. Regional Initiatives

8.3.1. PEPS Mirabelle EXPLOD-Biomed

Participants: Adrien Coulet [contact person], 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.

8.3.2. Hydreos

Participant: Jean-François Mari [contact person].

Hydreos is a state organization –actually a so-called “Pôle de compétitivité”– aimed at evaluating the delivering and the quality of water (http://www.hydreos.fr/fr). Actually, data about water resources rely on many agronomic variables, including land use successions. 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. Accordingly, one main challenge in our participation in Hydreos is to process and analyze space-time data for reaching a better understanding of the changes in the organization of a territory.

The systems ARPEnTAge (see § 6.2.2) and CarottAge (see § 6.2.1) are used in this context, especially by agronomists of INRA (ASTER Mirecourt http://www6.nancy.inra.fr/sad-aster. Currently, various display tools are under study and implementation for providing the agrnomy expert an easier interpretation of the clustering outputs http://www.loria.fr/~jfmari/App/Arpentage/Yar.avi.

8.3.3. PEPS Truffinet

Participant: Chedy Raïssi [contact person].

The Truffinet PEPS project aims at developing new graph mining methods and tools to support knowledge discovery from the truffle’s complex network of interactions happening in the soil between different bacterias and the subterranean Ascomycete fungus. This work uses Log-Linear Analysis (LogLA) which is a well established statistical technique for finding associations between discrete variables in data. The general objective of LogLA is to select a model that satisfactorily explains the observed frequencies of a given categorical dataset. General approaches to LogLA are exponential with respect to the number of variables. Recently, new approaches based on multiplicative log-linear models and using notions from graph theory have been developed. We applied successfully these methods in the case of the truffle bacterial environment to discover new associations in our data.

The Truffinet PEPS project involves several partners among which Intitut Elie Cartan de Lorraine (IECL), Intitut National de Recherche en Agronomie (INRA) and Centre de Recherche en Automatique de Nancy (CRAN) along with Inria Nancy Grand Est.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organization, General Chairs, Scientific Chairs

- “CCC”. Emmanuel Nauer co-organized the “Computer Cooking Contest” at ICCBR 2015 which held this year in Bad Homburg (Germany) (see http://ccc2015.loria.fr/).
- “FCA4AI 2015”. Amedeo Napoli organized with Sergei O. Kuznetsov (HSE Moscow) and Sebastian Rudolph (TU Dresden) the fourth workshop FCA4AI (“What can do FCA for Artificial Intelligence”) which was associated with the IJCAI Conference in Buenos Aires (Argentina, July 2015, see http://www.fca4ai.hse.ru/2015 and http://ceur-ws.org/Vol-1430).

9.1.1.1. Scientific Animation

- The scientific animation in the Orpailleur team is based on the Team Seminar which is called the “Malotec” seminar (http://malotec.loria.fr/?p=1). The Malotec 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. members of the team are also aware of the BINGO seminar which is organized by the Capsid Team (composed of former members of the Orpailleur Team), whose topics are related to biology, chemistry, and medicine. 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 - Juries

- 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] M. ALAM, A. BUZMAKOV, V. CODOCEDO, A. NAPOLI. *Mining Definitions from RDF Annotations Using Formal Concept Analysis*, in "International Joint Conference in Artificial Intelligence", Buenos Aires, Argentina, Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, July 2015, https://hal.archives-ouvertes.fr/hal-01186204

[2] M. COUCEIRO, E. LEHTONEN, K. SCHÖLZEL. *Hypomorphic Sperner Systems and Non-Reconstructible Functions*, in "Order", July 2015, vol. 32, n^{o} 2, pp. 255-292 [*DOI :* 10.1007/S11083-014-9330-Z], https://hal.archives-ouvertes.fr/hal-01090540

[3] K. DALLEAU, N. COUMBA NDIAYE, A. COULET. *Suggesting valid pharmacogenes by mining linked data*, in "Semantic Web Applications and Tools for Life Sciences (SWAT4LS) 2015", Cambridge, United Kingdom, Proceedings of the Semantic Web Applications and Tools for Life Sciences (SWAT4LS) 2015, December 2015, https://hal.inria.fr/hal-01239568

[4] V. DUFOUR-LUSSIER, J. LIEBER. *Evaluating a textual adaptation system*, in "International Conference on Case-Based Reasoning", Frankfurt, Germany, September 2015, https://hal.inria.fr/hal-01178331

[5] Z. DVOŘÁK, J.-S. SERENI, J. VOLEC. *Fractional coloring of triangle-free planar graphs*, in "Electronic Journal of Combinatorics", 2015, vol. 22, n^{o} 4, #P4.11 p. , https://hal.archives-ouvertes.fr/hal-00950493

[6] A. GHOORAH, M.-D. DEVIGNES, S. Z. ALBORZI, M. SMAÏL-TABBONE, D. RITCHIE. *A Structure-Based Classification and Analysis of Protein Domain Family Binding Sites and Their Interactions*, in "Biology", April 2015, vol. 4, n^{o} 2, pp. 327-343 [*DOI :* 10.3390/BIOLOGY4020327], https://hal.inria.fr/hal-01216748

[7] L. F. MELO MORA, Y. TOUSSAINT. *Automatic Validation of Terminology by Means of Formal Concept Analysis*, in "International Conference in Formal Concept Analysis - ICFCA 2015", Nerja , Spain, J. BAIXERIES, C. SACAREA, M. OJEDA-ACIEGO (editors), Formal Concept Analysis - Lecture Notes in Computer Science, Springer, June 2015, vol. 9113, pp. 236-251, https://hal.inria.fr/hal-01176422

[8] W. UGARTE ROJAS, P. BOIZUMAULT, B. CRÉMILLEUX, A. LEPAILLEUR, S. LOUDNI, M. PLANTEVIT, C. RAÏSSI, A. SOULET. *Skypattern mining: From pattern condensed representations to dynamic constraint satisfaction problems*, in "Artificial Intelligence", April 2015, 22 p. [*DOI :* 10.1016/J.ARTINT.2015.04.003], https://hal.inria.fr/hal-01188928

[9] M. VAN LEEUWEN, E. GALBRUN. *Association Discovery in Two-View Data*, in "IEEE Transactions on Knowledge and Data Engineering", December 2015, vol. 27, n^{o} 12, pp. 3190 - 3202 [*DOI :* 10.1109/TKDE.2015.2453159], https://hal.archives-ouvertes.fr/hal-01242988

[10] Y. XIAO, C. MIGNOLET, M. BENOÎT, J.-F. MARI. *Characterizing historical (1992–2010) transitions between grassland and cropland in mainland France through mining land-cover survey data* , in "Journal of Integrative Agriculture ", August 2015, vol. 14, n^{o} 8, pp. ”1511 - 1523”, https://hal.inria.fr/hal-01202510

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] M. ALAM. *Interactive Knowledge Discovery over Web of Data*, Loria & Inria Grand Est, December 2015, https://hal.inria.fr/tel-01245458

[12] A. BUZMAKOV. *Formal Concept Analysis and Pattern Structures for mining Structured Data*, Universite de Lorraine, October 2015, https://hal.inria.fr/tel-01229062

[13] V. CODOCEDO-HENRIQUEZ. *Contributions to indexing and retrieval using Formal Concept Analysis*, Université de Lorraine, September 2015, https://hal.inria.fr/tel-01241474

Articles in International Peer-Reviewed Journals

[14] J. Almeida, M. Couceiro, T. Waldhauser. *On the topological semigroup of equational classes of finite functions under composition*, in "Journal of Multiple-Valued Logic and Soft Computing", 2016, 24 p. , https://hal.archives-ouvertes.fr/hal-01090645

[15] S. Bessy, D. Gonçalves, J.-S. Sereni. *Two floor building needing eight colors*, in "Journal of Graph Algorithms and Applications", January 2015, vol. 19, n^o 1, pp. 1–9, https://hal.archives-ouvertes.fr/hal-00996709

[16] 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 P4 -sparse graphs*, in "Information Processing Letters (IPL)", 2015, vol. 115, n^o 6-8, pp. 600-603, Paper submitted to a journal, 11/05/2014, https://hal.archives-ouvertes.fr/hal-01102515

[17] F. Bonomo, G. Duran, M. Valencia-Pabon. *Complexity of the cluster deletion problem on some subclasses of chordal graphs*, in "Journal of Theoretical Computer Science (TCS)", 2015, vol. 600, pp. 59-69, Paper submitted to a Journal, 28/10/2014, https://hal.archives-ouvertes.fr/hal-01102512

[18] G. Bosc, P. Tan, J.-F. Boulicaut, C. Raïssi, M. Kaytoue. *A Pattern Mining Approach to Study Strategy Balance in RTS Games*, in "IEEE Transactions on Computational Intelligence and AI in Games (T-CIAIG)", December 2015 [*DOI :* 10.1109/TCIAIG.2015.2511819], https://hal.archives-ouvertes.fr/hal-01252728

[19] A. Buzmakov, E. Egho, N. Jay, S. O. Kuznetsov, A. Napoli, C. Raïssi. *On Mining Complex Sequential Data by Means of FCA and Pattern Structures*, in "Int. J. Gen. Syst.", 2015, pp. 1-25 [*DOI :* 10.1080/03081079.2015.1072925], https://hal.archives-ouvertes.fr/hal-01186715

[20] M. Couceiro, D. Dubois, H. Prade, T. Waldhauser. *Decision-making with sugeno integrals*, in "Order", December 2015 [*DOI :* 10.1007/s11083-015-9382-8], https://hal.archives-ouvertes.fr/hal-01184340

[21] M. Couceiro, L. Haddad, I. G. Rosenberg. *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

[22] M. Couceiro, L. Haddad, K. Schölzel, T. Waldhauser. *A Solution to a Problem of D. Lau: Complete Classification of Intervals in the Lattice of Partial Boolean Clones*, in "Journal of Multiple-Valued Logic and Soft Computing", 2016, 8 p. , https://hal.inria.fr/hal-01183004

[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", 2016, vol. 27, n^o 1-2, 15 p. , https://hal.inria.fr/hal-01175695

[25] 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 [*DOI :* 10.1051/ro/2014034], https://hal.archives-ouvertes.fr/hal-01090621

[26] M. COUCEIRO, E. LEHTONEN, K. SCHÖLZEL. *Hypomorphic Sperner Systems and Non-Reconstructible Functions*, in "Order", July 2015, vol. 32, n^{o} 2, pp. 255-292 [*DOI :* 10.1007/S11083-014-9330-Z], https://hal.archives-ouvertes.fr/hal-01090540

[27] M. COUCEIRO, E. LEHTONEN, K. SCHÖLZEL. *Set-reconstructibility of Post classes*, in "Discrete Applied Mathematics", May 2015, vol. 187, pp. 12-18, 8 pages. arXiv admin note: text overlap with arXiv:1306.5578, https://hal.archives-ouvertes.fr/hal-01090618

[28] 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

[29] M. COUCEIRO, J.-L. MARICHAL, B. TEHEUX. *Conservative Median Algebras and Semilattices*, in "Order", May 2015, 11 p. [*DOI :* 10.1007/S11083-015-9356-X], https://hal.inria.fr/hal-01175696

[30] M. COUCEIRO, J.-L. MARICHAL, B. TEHEUX. *Relaxations of associativity and preassociativity for variadic functions*, in "Fuzzy Sets and Systems", November 2015, https://hal.archives-ouvertes.fr/hal-01184334

[31] Z. DVOŘÁK, J.-S. SERENI, J. VOLEC. *Fractional coloring of triangle-free planar graphs*, in "Electronic Journal of Combinatorics", 2015, vol. 22, n^{o} 4, #P4.11 p. , https://hal.archives-ouvertes.fr/hal-00950493

[32] E. EGHO, C. RAÏSSI, T. CALDERS, N. JAY, A. NAPOLI. *On measuring similarity for sequences of itemsets*, in "Data Mining and Knowledge Discovery", May 2015, vol. 29, n^{o} 3, 33 p. [*DOI :* 10.1007/S10618-014-0362-1], https://hal.inria.fr/hal-01094383

[33] A. GHOORAH, M.-D. DEVIGNES, S. Z. ALBORZI, M. SMAÏL-TABBONE, D. RITCHIE. *A Structure-Based Classification and Analysis of Protein Domain Family Binding Sites and Their Interactions*, in "Biology", April 2015, vol. 4, n^{o} 2, pp. 327-343 [*DOI :* 10.3390/BIOLOGY4020327], https://hal.inria.fr/hal-01216748

[34] J.-P. METIVIER, A. LEPAILLEUR, A. BUZMAKOV, G. POEZEVARA, B. CRÉMILLEUX, S. O. KUZNETSOV, J. LE GOFF, A. NAPOLI, R. BUREAU, B. CUISSART. *Discovering structural alerts for mutagenicity using stable emerging molecular patterns*, in "Journal of Chemical Information and Modeling", 2015, vol. 55, n^{o} 5, pp. 925–940 [*DOI :* 10.1021/CI500611V], https://hal.archives-ouvertes.fr/hal-01186716

[35] W. UGARTE ROJAS, P. BOIZUMAULT, B. CRÉMILLEUX, A. LEPAILLEUR, S. LOUDNI, M. PLANTEVIT, C. RAÏSSI, A. SOULET. *Skypattern mining: From pattern condensed representations to dynamic constraint satisfaction problems*, in "Artificial Intelligence", April 2015, 22 p. [*DOI :* 10.1016/J.ARTINT.2015.04.003], https://hal.inria.fr/hal-01188928

[36] M. VAN LEEUWEN, E. GALBRUN. *Association Discovery in Two-View Data*, in "IEEE Transactions on Knowledge and Data Engineering", December 2015, vol. 27, n^{o} 12, pp. 3190 - 3202 [*DOI :* 10.1109/TKDE.2015.2453159], https://hal.archives-ouvertes.fr/hal-01242988

[37] Y. XIAO, C. MIGNOLET, M. BENOÎT, J.-F. MARI. *Characterizing historical (1992–2010) transitions between grassland and cropland in mainland France through mining land-cover survey data*, in "Journal of Integrative Agriculture ", August 2015, vol. 14, n^{o} 8, pp. "1511 - 1523", https://hal.inria.fr/hal-01202510

Articles in Non Peer-Reviewed Journals

[38] M.-D. Devignes, M. Smaïl-Tabbone, D. Ritchie. *Kbdock - Searching and organising the structural space of protein-protein interactions*, in "ERCIM News", January 2016, n^o 104, pp. 24-25, https://hal.inria.fr/hal-01258117

Invited Conferences

[39] M. Couceiro, T. Waldhauser. *Lattice-Theoretic Approach to Version Spaces in Qualitative Decision Making*, in "Statistical Learning and Data Sciences", London, United Kingdom, Lecture Notes in Computer Science, April 2015, vol. 9047, pp. 234-238 [*DOI :* 10.1007/978-3-319-17091-6_18], https://hal.inria.fr/hal-01175691

[40] X. Goaoc, A. Hubard, R. De Joannis De Verclos, J.-S. Séréni, J. Volec. *Limits of order types*, in "Symposium on Computational Geometry 2015", Eindhoven, Netherlands, L. A. Janos Pach (editor), June 2015, vol. 34, 876 p. [*DOI :* 10.4230/LIPIcs.SOCG.2015.300], https://hal.inria.fr/hal-01172466

[41] A. Napoli. *Concept Lattices for Knowledge Discovery and Knowledge Engineering*, in "Brazilian Conference on Artificial Intelligence, Natal, Brazil (BRACIS 2015)", Natal, Brazil, G. L. Pappa, K. C. Revoredo (editors), Anne Magaly de Paula Canuto (UFRN), November 2015, https://hal.inria.fr/hal-01254131

[42] A. Napoli. *Exploratory Knowledge Discovery with Formal Concept Analysis*, in "Advanced Information Technology, Services and Systems (AIT2S-15)", Settat, Morocco, M. Bahaj (editor), December 2015, https://hal.inria.fr/hal-01254141

[43] A. Napoli. *Exploring Complex and Large Data with Formal Concept Analysis*, in "Colloque sur l'Optimisation et les Systèmes d'Information (COSI 2015)", Oran, Algeria, J.-M. Petit (editor), Rachid Nourine, June 2015, https://hal.inria.fr/hal-01254129

International Conferences with Proceedings

[44] M. Alam, A. Buzmakov, V. Codocedo, A. Napoli. *Bridging DBpedia Categories and DL-Concept Definitions using Formal Concept Analysis*, in "Proceedings of the 4th International Workshop "What can FCA do for Artificial Intelligence?", FCA4AI 2015, co-located with the International Joint Conference on Artificial Intelligence (IJCAI 2015)", Buenos Aires, Argentina, CEUR Workshop Proceedings, July 2015, vol. 1430, https://hal.inria.fr/hal-01186330

[45] M. Alam, A. Buzmakov, V. Codocedo, A. Napoli. *Mining Definitions from RDF Annotations Using Formal Concept Analysis*, in "International Joint Conference in Artificial Intelligence", Buenos Aires, Argentina, Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, July 2015, https://hal.archives-ouvertes.fr/hal-01186204

[46] M. Alam, A. Buzmakov, A. Napoli, A. Sailanbayev. *Revisiting Pattern Structures for Structured Attribute Sets*, in "Proceedings of the Twelfth International Conference on Concept Lattices and Their Applications, October 13-16, 2015", Clermont-Ferrand, France, CEUR Workshop Proceedings, August 2015, vol. 1466, pp. 241-252, https://hal.inria.fr/hal-01186339

[47] M. Alam, A. Napoli. *An Approach Towards Classifying and Navigating RDF data based on Pattern Structures*, in "Proceedings of the International Workshop on Formal Concept Analysis and Applications 2015 co-located with 13th International Conference on Formal Concept Analysis", Nerja, Spain, Formal Concept Analysis and Applications, June 2015, vol. 1434, pp. 33-48, https://hal.inria.fr/hal-01186327

[48] M. Alam, A. Napoli. *Interactive Exploration over RDF Data using Formal Concept Analysis*, in "Proceedings of IEEE International Conference on Data Science and Advanced Analytics", Paris, France, August 2015 [*DOI :* 10.1109/DSAA.2015.7344838], https://hal.inria.fr/hal-01186335

[49] M. Alam, A. Napoli, M. Osmuk. *RV-Xplorer: A Way to Navigate Lattice-Based Views over RDF Graphs*, in "Proceedings of the Twelfth International Conference on Concept Lattices and Their Applications", Clermont-Ferrand, France, October 2015, vol. 1466, https://hal.inria.fr/hal-01186344

[50] M. Barrère, G. Betarte, V. Codocedo, M. Rodríguez, H. Astudillo, M. Aliquintuy, J. Baliosian, R. Badonnel, O. Festor, C. Raniery Paula dos Santos, J. Campos Nobre, L. Z. Granville, A. Napoli. *Machine-assisted Cyber Threat Analysis using Conceptual Knowledge Discovery*, in "FCA4AI 2015 - Workshop What can FCA do for Artificial Intelligence?", Buenos Aires, Argentina, July 2015, pp. 75 - 85, https://hal.archives-ouvertes.fr/hal-01186213

[51] O. Bruneau, S. Garlatti, M. Guedj, S. Laubé, J. Lieber. *SemanticHPST: Applying Semantic Web Principles and Technologies to the History and Philosophy of Science and Technology*, in "The Semantic Web: ESWC 2015 Satellite Events", Portoroz, Slovenia, F. Gandon, A. Zimmermann, C. Faron-Zucker, J. Breslin, S. Villata, C. Guéret (editors), Lecture Notes in Computer Science, Springer International Publishing, May 2015, vol. 9341, pp. 416-427 [*DOI :* 10.1007/978-3-319-25639-9_53], https://hal.archives-ouvertes.fr/hal-01214698

[52] A. Buzmakov, S. O. Kuznetsov, A. Napoli. *Fast Generation of Best Interval Patterns for Nonmonotonic Constraints*, in "Machine Learning and Knowledge Discovery in Databases", Porto, Portugal, Lecture Notes in Computer Science, September 2015, vol. 9285, pp. 157-172 [*DOI :* 10.1007/978-3-319-23525-7_10], https://hal.archives-ouvertes.fr/hal-01186718

[53] *Best Paper*
A. Buzmakov, S. O. Kuznetsov, A. Napoli. *Revisiting Pattern Structure Projections*, in "International Conference in Formal Concept Analysis - ICFCA 2015", Nerja, Spain, J. Baixeries, C. Sacarea, M. Ojeda-Aciego (editors), Lecture Notes in Computer Science, Springer International Publishing, June 2015, vol. 9113, pp. 200–215 [*DOI :* 10.1007/978-3-319-19545-2_13], https://hal.archives-ouvertes.fr/hal-01186719.

[54] V. Codocedo, A. Napoli. *Formal Concept Analysis and Information Retrieval – A Survey*, in "International Conference in Formal Concept Analysis - ICFCA 2015", Nerja, Spain, J. Baixeries, C. Sacarea, M. Ojeda-Aciego (editors), Formal Concept Analysis - Lecture Notes in Computer Science, Springer, June 2015, vol. 9113, pp. 61-77 [*DOI :* 10.1007/978-3-319-19545-2_4], https://hal.archives-ouvertes.fr/hal-01186196

[55] M. Couceiro, Q. Brabant. *Axiomatisation des intégrales de Sugeno k-maxitives*, in "24ème Conférence Rencontres francophones sur la Logique Floue et ses Applications (LFA 2015)", Poitiers, France, November 2015, https://hal.inria.fr/hal-01247510

[56] M. Couceiro, L. Haddad, M. Pouzet, K. Schölzel. *Hereditary Rigid Relations*, in "45th IEEE International Symposium on Multiple-Valued Logic (ISMVL 2015)", Waterloo, Canada, May 2015, https://hal.inria.fr/hal-01175699

[57] M. COUCEIRO, J.-L. MARICHAL, B. TEHEUX. *Median preserving aggregation functions*, in "8th International Summer School on Aggregation Operators and their Applications (AGOP 2015)", Katowice, Poland, Proc. 8th Int. Summer School on Aggregation Operators and their Applications (AGOP 2015), Michał Baczyński, Bernard De Baets, Radko Mesiar, July 2015, pp. 85-89, https://hal.archives-ouvertes.fr/hal-01184116

[58] M. COUCEIRO, B. TEHEUX. *Clones of pivotally decomposable functions*, in "45th IEEE International Symposium on Multiple-Valued Logic (ISMVL 2015)", Waterloo, Canada, May 2015, https://hal.inria.fr/hal-01175698

[59] K. DALLEAU, N. COUMBA NDIAYE, A. COULET. *Suggesting valid pharmacogenes by mining linked data*, in "Semantic Web Applications and Tools for Life Sciences (SWAT4LS) 2015", Cambridge, United Kingdom, December 2015, https://hal.inria.fr/hal-01239568

[60] V. DUFOUR-LUSSIER, J. LIEBER. *Evaluating a textual adaptation system*, in "International Conference on Case-Based Reasoning", Frankfurt, Germany, September 2015, https://hal.inria.fr/hal-01178331

[61] E. GAILLARD, J. LIEBER, E. NAUER. *How Managing the Knowledge Reliability Improves the Results of a Reasoning Process*, in "16th European Conference on Knowledge Management - ECKM 2015", Udine, Italy, September 2015, 10 p. , https://hal.inria.fr/hal-01178315

[62] E. GAILLARD, J. LIEBER, E. NAUER. *Improving Case Retrieval Using Typicality*, in "23rd International Conference on Case-Based Reasoning (ICCBR 2015)", Frankfurt am Main, Germany, September 2015, https://hal.inria.fr/hal-01178317

[63] E. GAILLARD, J. LIEBER, E. NAUER. *Improving Ingredient Substitution using Formal Concept Analysis and Adaptation of Ingredient Quantities with Mixed Linear Optimization*, in "Computer Cooking Contest Workshop", Frankfort, Germany, September 2015, https://hal.inria.fr/hal-01240383

[64] M. HASSAN, O. MAKKAOUI, A. COULET, Y. TOUSSAINT. *Extracting Disease-Symptom Relationships by Learning Syntactic Patterns from Dependency Graphs*, in "BioNLP 15", Beijing, China, Proceedings of BioNLP 15, Association for Computational Linguistics, July 2015, 184 p. , https://hal.inria.fr/hal-01184655

[65] M. KAYTOUE, V. CODOCEDO, A. BUZMAKOV, J. BAIXERIES, S. O. KUZNETSOV, A. NAPOLI. *Pattern Structures and Concept Lattices for Data Mining and Knowledge Processing*, in "Machine Learning and Knowledge Discovery in Databases", Porto, Portugal, A. BIFET, M. MAY, B. ZADROZNY, R. GAVALDA, D. PEDRESCHI, F. BONCHI, J. CARDOSO, M. SPILIOPOULOU (editors), Lecture Notes in Computer Science, Springer International Publishing, 2015, vol. 9286, pp. 227-231 [*DOI :* 10.1007/978-3-319-23461-8_19], https://hal.archives-ouvertes.fr/hal-01188637

[66] A. LEEUWENBERG, A. BUZMAKOV, Y. TOUSSAINT, A. NAPOLI. *Exploring Pattern Structures of Syntactic Trees for Relation Extraction*, in "International Conference in Formal Concept Analysis - ICFCA 2015", Nerja, Spain, J. BAIXERIES, C. SACAREA, M. OJEDA-ACIEGO (editors), Lecture Notes in Computer Science, Springer, June 2015, vol. 9113, pp. 153–168 [*DOI :* 10.1007/978-3-319-19545-2_10], https://hal.archives-ouvertes.fr/hal-01186717

[67] L. F. MELO MORA, Y. TOUSSAINT. *Automatic Validation of Terminology by Means of Formal Concept Analysis*, in "International Conference in Formal Concept Analysis - ICFCA 2015", Nerja , Spain, J.

BAIXERIES, C. SACAREA, M. OJEDA-ACIEGO (editors), Lecture Notes in Computer Science, Springer, June 2015, vol. 9113, pp. 236-251, https://hal.inria.fr/hal-01176422

National Conferences with Proceedings

[68] M. COUCEIRO, B. TEHEUX, J.-L. MARICHAL. *Agrégation des valeurs médianes et fonctions compatibles pour la comparaison*, in "24ème Conférence Rencontres francophones sur la Logique Floue et ses Applications (LFA 2015)", Poitiers, France, November 2015, https://hal.inria.fr/hal-01247525

[69] E. GAILLARD, J. LIEBER, E. NAUER. *Taaable : un système de raisonnement à partir de cas qui adapte des recettes de cuisine*, in "1ère Conférence Nationale sur les Applications Pratiques de l'Intelligence Artificielle (APIA 2015)", Rennes, France, July 2015, https://hal.inria.fr/hal-01178316

[70] J. LIEBER. *Révision des croyances dans une clôture propositionnelle de contraintes linéaires*, in "Journées d'intelligence artificielle fondamentale, plate-forme intelligence artificielle", Rennes, France, N. MAUDET, B. ZANUTTINI (editors), June 2015, 10 p. , https://hal.inria.fr/hal-01178267

Books or Proceedings Editing

[71] S. O. KUZNETSOV, A. NAPOLI, S. RUDOLPH (editors). *Workshop NotesInternational Workshop "What can FCA do for Artificial Intelligence?" (FCA4AI 2015)*, CEUR Workshop Proceedings 1430, July 2015, vol. CEUR Workshop Proceedings 1430, https://hal.inria.fr/hal-01252624

Research Reports

[72] J. LIEBER. *Révision des croyances dans une clôture propositionnelle de contraintes linéaires (version étendue)*, LORIA, UMR 7503, Université de Lorraine, CNRS, Vandoeuvre-lès-Nancy ; Inria Nancy - Grand Est (Villers-lès-Nancy, France), May 2015, 17 p. , https://hal.inria.fr/hal-01155235

[73] A. MULLER-GUEUDIN, M. BUEE, A. DEVEAU, A. GÉGOUT-PETIT, S. MARTIN, I.-C. MORARESCU, C. RAÏSSI. *"Truffinet": Inférence de réseaux d'interactions microbiennes dans la truffe. Rapport scientique du PEPS Mirabelles 2014*, IECL ; Inria BIGS, September 2015, https://hal.archives-ouvertes.fr/hal-01236087

Other Publications

[74] Y. ABID, A. IMINE, A. NAPOLI, C. RAÏSSI, M. RIGOLOT, M. RUSINOWITCH. *Analyse d'activité et exposition de la vie privée sur les médias sociaux*, January 2016, 16ème conférence francophone sur l'Extraction et la Gestion des Connaissances (EGC 2016), Poster, https://hal.inria.fr/hal-01241619

[75] M. COUCEIRO, S. FOLDES, G. MELETIOU. *Arrow Type Impossibility Theorems over Median Algebras*, August 2015, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01185977

[76] M. COUCEIRO, L. HADDAD, K. SCHÖLZEL, T. WALDHAUSER. *On the interval of strong partial clones of Boolean functions containing Pol((0,0),(0,1),(1,0))*, August 2015, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01184404

[77] M. COUCEIRO, G. MELETIOU. *On a special class of median algebras*, August 2015, working paper or preprint, https://hal.inria.fr/hal-01248142

[78] P. TORRES, M. VALENCIA-PABON. *Stable Kneser Graphs are almost all not weakly Hom-Idempotent* , February 2015, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01119741

References in notes

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

[80] 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

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

[82] 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

[83] 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

[84] 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

[85] 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

[86] 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

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

[88] 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

[89] 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

[90] R. Feldman, J. Sanger. *The Text Mining Handbook (Advanced Approaches in Analyzing Unstructured Data)*, Cambridge University Press, 2007

[91] 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

[92] B. Ganter, S. O. Kuznetsov. *Pattern Structures and Their Projections*, in "Proceedings of ICCS 2001", LNCS 2120, Springer, 2001, pp. 129–142

[93] B. Ganter, R. Wille. *Formal Concept Analysis*, Springer, Berlin, 1999

[94] 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, 681 p. , https://hal.archives-ouvertes.fr/hal-01057112

[95] O. Hudry, B. Monjardet. *Consensus Theories. An oriented survey*, in "Mathématiques et Sciences Humaines", 2010, vol. 190, n^{o} 2, pp. 139–167

[96] 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

[97] 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

[98] 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

[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*, 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

[100] 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

[101] 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

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

[103] 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