

RESEARCH CENTRE

Nancy - Grand Est

IN PARTNERSHIP WITH:

CNRS, Université de Lorraine

2020

ACTIVITY REPORT

Project-Team

ORPAILLEUR

Knowledge representation, reasoning

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

DOMAIN

Perception, Cognition and Interaction

THEME

Data and Knowledge Representation and Processing

Contents

Project-Team ORPAILLEUR	1
1 Team members, visitors, external collaborators	2
2 Overall objectives	3
2.1 Introduction	3
3 Research program	4
3.1 Fundamentals of Knowledge Discovery in Databases	4
3.2 KDD in Practice	4
3.3 Explanations and Fairness in KDD	5
4 Application domains	5
4.1 Life Sciences: Agronomy, Biology, Chemistry, and Medicine	5
4.2 Text Mining	5
4.3 Knowledge Systems and Web of Data	6
5 Highlights of the year	6
6 New software and platforms	7
6.1 New software	7
6.1.1 ARPEnTAge	7
6.1.2 CarottAge	7
6.1.3 CORON	7
6.2 New platforms	8
6.2.1 FixOut: an ensemble approach to build fairer models	8
6.2.2 LatViz: Visualization of Concept Lattices	8
6.2.3 OrphaMine: Data Mining Platform for Orphan Diseases	8
6.2.4 Siren: Interactive and Visual Redescription Mining	9
7 New results	9
7.1 Fundamentals of KDD	9
7.1.1 FCA, Pattern Mining, MDL, and Hybrid Mining	9
7.2 KDD in Practice	10
7.3 Explanations and Fairness in KDD	11
8 Partnerships and cooperations	12
8.1 European initiatives	12
8.1.1 FP7 & H2020 Projects	12
8.2 National initiatives	12
8.2.1 ANR	12
8.2.2 Inria Project Labs, Exploratory Research Actions, and Technological Development Actions	13
9 Dissemination	14
9.1 Promoting scientific activities	14
9.1.1 Scientific Events: Organization	14
9.1.2 General chair, scientific chair	14
9.1.3 Member of Editorial Boards	14
9.1.4 Reviewer - reviewing activities	15
9.1.5 Invited talks	15
9.1.6 Teaching	15
9.1.7 Tutorials	15
9.1.8 Supervision	15

10 Scientific production	16
10.1 Major publications	16
10.2 Publications of the year	16
10.3 Cited publications	22

Project-Team ORPAILLEUR

Creation of the Project-Team: 2008 January 01, end of the Project-Team: 2020 July 01

Keywords

Computer sciences and digital sciences

- A3.1.1. – Modeling, representation
- A3.1.7. – Open data
- A3.2.1. – Knowledge bases
- A3.2.2. – Knowledge extraction, cleaning
- A3.2.4. – Semantic Web
- A3.2.5. – Ontologies
- A3.2.6. – Linked data
- A3.3.2. – Data mining
- A3.4.1. – Supervised learning
- A3.4.2. – Unsupervised learning
- A3.4.5. – Bayesian methods
- A3.4.8. – Deep learning
- A3.5.2. – Recommendation systems
- A4.8. – Privacy-enhancing technologies
- A8.1. – Discrete mathematics, combinatorics
- A9.1. – Knowledge
- A9.2. – Machine learning
- A9.6. – Decision support
- A9.8. – Reasoning
- A9.10. – Hybrid approaches for AI

Other research topics and application domains

- B1.1. – Biology
- B2.3. – Epidemiology
- B2.4.1. – Pharmaco kinetics and dynamics
- B2.4.2. – Drug resistance
- B3.5. – Agronomy
- B3.6. – Ecology
- B3.6.1. – Biodiversity
- B5.2.1. – Road vehicles

1 Team members, visitors, external collaborators

Research Scientist

- Amedeo Napoli [CNRS, Emeritus, HDR]

Faculty Members

- Miguel Couceiro [Team leader, Univ de Lorraine, Professor, HDR]
- Alexandre Blansch  [Univ de Lorraine, Associate Professor]
- Baya Lydia Boudjeloud-Assala [Univ de Lorraine, Associate Professor, HDR]
- Brieuc Conan-Guez [Univ de Lorraine, Associate Professor]
- Adrien Coulet [Univ de Lorraine, Associate Professor, HDR]
- S bastien Da Silva [Univ de Lorraine, Associate Professor]
- Alain Gely [Univ de Lorraine, Associate Professor]
- Jean-Fran ois Mari [Univ de Lorraine, Emeritus, HDR]
- Justine Reynaud [Univ de Lorraine, ATER, until Aug 2020]
- Yannick Toussaint [Univ de Lorraine, Professor, HDR]

Post-Doctoral Fellows

- Alexandre Bazin [Univ de Lorraine]
- Nyoman Juniarta [Inria]

PhD Students

- Nacira Abbas [Inria]
- Guilherme Alves Da Silva [Inria]
- Laurine Huber [Univ de Lorraine]
- Tatiana Makhalova [Inria]
- Pierre Monnin [Univ de Lorraine, until Dec 2020]
- Claire Theobald [CNRS]
- Laura Alejandra Zanella Calzada [Univ de Lorraine]
- Georgios Zervakis [Inria]

Technical Staff

- J r mie Nevin [Inria, Engineer, until Jun 2020]

Interns and Apprentices

- Manar Amezrine [Univ de Lorraine, from Jun 2020 until Sep 2020]
- Srilakshmi Balard [Univ de Lorraine, from Mar 2020 until Aug 2020]
- Clement Bellanger [Univ de Lorraine, from Mar 2020 until Aug 2020]
- Vaishnavi Bhargava [Univ de Lorraine, from Jan until Jun 2020]
- Walid Hafiane [Univ de Lorraine, from Mar 2020 until Aug 2020]
- Emilie Leroy [Univ de Lorraine, from Jun 2020 until Aug 2020]
- Esteban Marquer [Inria, from Feb 2020 until Jun 2020]
- Yi Ting Tsai [Univ de Lorraine, from Apr 2020 until Sep 2020]
- Charles Vernerey [Univ de Lorraine, from Apr 2020 until Aug 2020]

Administrative Assistant

- Emmanuelle Deschamps [Inria, until Jun 2020]

External Collaborator

- Florence Le Ber [Ecole Nationale du Génie de l'Eau et de l'Environnement de Strasbourg, until Jun 2020]

2 Overall objectives

2.1 Introduction

The three main scientific objectives of the Orpailleur Team are: (1) Fundamentals of Knowledge Discovery in Databases (KDD), (2) KDD in practice, and (3) Explanations and Fairness in KDD. KDD aims at discovering intelligible and reusable patterns in possibly large and complex databases. 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. Moreover, the KDD process is iterative, interactive, and “exploratory”, i.e. controlled by an expert of the data domain, called the analyst.

The KDD process –as implemented in the Orpailleur team– is based on data mining methods which are either symbolic or subsymbolic (numerical). Symbolic methods are based on pattern mining, Formal Concept Analysis (FCA), and extensions. Subsymbolic methods are based on classifiers such as Random Forests, SVM, and Neural Networks. Then a main objective is to mine complex and possibly large data in combining symbolic and subsymbolic data mining methods for improving applicability and explainability of KDD and making KDD more “reliable”. In this way, domain knowledge may improve and guide the KDD process, leading to “Knowledge Discovery guided by Domain Knowledge” (KDDK). The discovered patterns can be interpreted as knowledge units and reused for problem-solving activities in knowledge systems. Thus knowledge discovery can be considered as a key task in knowledge engineering, having an impact in various semantic activities, e.g. information retrieval, recommendation, and ontology engineering.

Recent advances in KDD and in Machine Learning (ML) are mostly due to the success of deep learning methods in recognition tasks. However, deep learning and in general subsymbolic data mining methods are based on complex models, whose outputs and proposed decisions, as accurate as they are, cannot be easily explained to the layman. Accordingly, we are studying how to design hybrid and more interpretable KDD methods, combining complex subsymbolic models and explainable symbolic models. In addition, explanations can be used for assessing algorithmic fairness, which is a critical point especially in decision support systems with societal and scientific impacts, and that became another main topic in the team.

Regarding applications, the team mainly works in life sciences, i.e. agronomy, biology, chemistry, medicine, pharmacogenomics, and as well in astronomy and in the web of data. Text mining is of major interest for the team and appears to be a ground task in many applications. Indeed it has gained a lot of attention in the last years, especially with the progress due to deep learning systems. The Orpailleur team has a rich experience in KDD and intends to improve and to extend this experience, paying also more attention to the impact of knowledge discovery in the real world. This should lead to the design of “green”, sustainable, explainable, and fair data mining systems.

3 Research program

3.1 Fundamentals of Knowledge Discovery in Databases

Keywords:

knowledge discovery in databases, knowledge discovery in databases guided by domain knowledge, pattern mining, data exploration, formal concept analysis, hybrid mining.

The Orpailleur team has an original and productive research activity in knowledge discovery and especially in developing hybrid –symbolic and subsymbolic– data mining methods. Actually, the team has an important activity in pattern mining, in Formal Concept Analysis and its extensions, in text mining, and as well in the search for explanations and studying the fairness of KDD algorithms. Moreover, the team combines in an original way knowledge discovery and knowledge engineering, in mining the web of data for ontology engineering, knowledge base completion, and link key discovery. KDD is a main task in many research projects as shown by the large range of applications in which the team is involved.

Advances in data and knowledge engineering have emphasized the needs for hybrid mining tools working on complex data. Accordingly, the Orpailleur team is developing knowledge discovery methods based on pattern mining, FCA and its extensions, subgroup discovery and redescription mining as well, to be used in real-sized applications on complex data including multi-valued attributes, n-ary relations, sequences, trees, and graphs... One main aspect is to make pattern mining more useful and to take into account domain knowledge, user-preferences, and constraints to guide the discovery process. In addition, the combination of symbolic and subsymbolic data mining methods should lead to more applicable, explainable, and reliable methods, as shown in [12].

3.2 KDD in Practice

Keywords:

knowledge discovery in life sciences, text mining, exploration of web of data, knowledge engineering, ontology engineering from texts and web resources.

Life Sciences constitute a challenging domain for KDD. Biological data are complex from many points of views, e.g. huge size, high-dimensionality, deep inter-connections, and quick evolution. Analyzing such data for discovering interesting patterns is an important issue in domains such as health, environment, and agronomy, for solving problems such as disease understanding, drug discovery, pharmacovigilance, and pharmacogenomics. Moreover, many ontologies are available which can be used to improve KDD in biology, and, in turn, KDD can help improving the quality and completion of ontologies especially thanks to text mining. In particular, we work on the mining of real-world texts in combining subsymbolic and symbolic data mining methods.

From a knowledge discovery perspective, text mining is a particular process aimed at extracting interesting units from texts which can be used further for ontology engineering, e.g. design or completion of a knowledge base. In parallel, knowledge discovery in the web of data, i.e. mining set of RDF (Resource Description Framework) triples, is a major task for discovering concept definitions in terms of necessary and sufficient conditions. Both operations, text mining and mining the web of data can be considered as complementary mining operations.

3.3 Explanations and Fairness in KDD

Keywords:

explainable artificial intelligence, production of explanations, fairness of data mining algorithms, fairness and explanations in decision making.

The output of KDD may be of different types, e.g. rules, classes, and concepts, which can be reused for solving problems and decision making. Then, elements supporting an algorithmic decision should be available as well [76, 78, 79]. In practice, subsymbolic KDD methods show that they are more flexible and better suited to deal with the complexity of tasks such as recognition. However, symbolic KDD methods are more often used in pattern mining and in domains where the learning model should be understandable by human experts, or be related to domain ontologies for knowledge representation, reasoning, or decision making needs. Thus we investigate different ways of providing explanations and facilitating interpretation of the outputs of subsymbolic KDD methods. Besides explainability, one main limit that retains KDD systems from being more widely operated is their tendency to be biased. Indeed, since these systems are based on a KDD loop or generalize from a training set, they are reproducing biases that are initially present in the data, and may lead to inaccurate or unfair results. At present, this new direction of investigation is of major interest in the Orpailleur Team and takes place among the main topics studied in the team.

4 Application domains

4.1 Life Sciences: Agronomy, Biology, Chemistry, and Medicine

Keywords:

knowledge discovery in life sciences, biology, chemistry, medicine, pharmacogenomics and precision medicine.

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. 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 the environment, or between genetic and phenotypic data, especially for public health and precision medicine (pharmacogenomics). Pharmacogenomics is a main challenge in the Orpailleur team as it considers a large panel of complex data ranging from biological to medical data, and various kinds of encoded domain knowledge ranging from texts to formal ontologies.

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 including 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 labeled graphs. Graph mining and Formal Concept Analysis methods play an important role in this application domain and can be used in an efficient and well-founded way [73].

4.2 Text Mining

Keywords:

text mining, knowledge discovery from texts, text annotation, ontology engineering from texts.

The objective of a text mining process is to extract useful knowledge units from large collections of texts [64]. 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 difficult task. 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 knowledge discovery perspective, text mining aims at extracting “interesting units” (nouns and relations) from texts with the help of domain knowledge encoded within a knowledge base (it should be noticed that process is roughly similar for text annotation). Text mining is especially useful in the context of semantic web for ontology engineering. In the Orpailleur team, we work on the mining of real-world texts in application domains such as biology and medicine, using subsymbolic and symbolic data mining methods. Accordingly, the text mining process may be used to enrich and to extend linguistic resources, while, in turn, linguistic and ontological resources can be exploited to guide a text mining process”.

4.3 Knowledge Systems and Web of Data

Keywords:

combining knowledge engineering and knowledge discovery, web of data, semantic web, ontology design, link keys.

The web of data constitutes a good platform for experimenting ideas on the combination of knowledge discovery and knowledge engineering (KE). A software agent may be able to understand and manipulate information on the web if and only if domain knowledge and ontologies for achieving those tasks are available. In particular, OWL (“Web Ontology Language” <https://www.w3.org/OWL/>), which is based on description logics (DLs [68]) is the representation language commonly used for designing ontologies and knowledge engineering. In OWL, knowledge units are represented by classes having properties and instances. Concepts are organized within a partially ordered set based on a subsumption relation and the inference services are based on subsumption and classification.

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 provides a natural supports to a concept definition in an ontology [65, 66]. In this way, we study how the web of data, considered as a set of knowledge sources, e.g. DBpedia, Wikidata, Yago, Freebase, can be mined for guiding the design of ontological definitions, and further, how knowledge discovery techniques can be applied for allowing a better usage of RDF triples in the web of data. In particular, in the Orpailleur team, we are interested in (i) providing thanks to FCA and pattern structures a support to concept definitions, (ii) using in addition redescription mining to support again concept definitions, (iii) and combining FCA and redescription mining for discovering link keys, which can be used for identifying individuals which share the same set of properties [6].

5 Highlights of the year

- The publication [31] was selected for the best paper award at the International Conference on Concept Lattices and Applications (CLA 2020 <https://cs.ttu.ee/events/cla2020/>).
- The new research direction of Orpailleur about explanations and fairness of knowledge discovery algorithms AI algorithms was encouraged by the invitation of Miguel Couceiro to give an invited talk (online) on the subject “Making ML models fairer through explanations: the case of LimeOut”, Miguel Couceiro, at the 9th International Conference on Analysis of Images, Social Networks and Texts, Russia [18].
- Adrien Coulet with Anita Burgun and Sarah Zohar of UMR1138 Inserm and Université de Paris proposed a project to the “2020 Call for Expression of Interest” for the creation of an Inserm Inria joint team in digital health (<https://inserm-inria.sciencescall.org/>). This project was selected and one output is the creation of the joint team “HeKA” (Inria, Inserm, Université de Paris) to be officially created in Spring 2021 at the CRI of Paris. Accordingly, Adrien Coulet obtained a 5-year secondement (“détachement”) at Inria Paris.

6 New software and platforms

6.1 New software

6.1.1 ARPEntAge

Name: Analyse de Régularités dans les Paysages : Environnement, Territoires, Agronomie

Keywords: Stochastic process, Hidden Markov Models

Functional Description: ARPEntAge is a software based on stochastic models (HMM2 and Markov Field) for analyzing spatio-temporal data-bases. 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.

URL: http://carottage.loria.fr/index_in_english.html

Contact: Jean-François Mari

Partner: INRA

6.1.2 CarottAge

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. CarottAge is currently used by INRA researchers interested in mining the changes in territories related to the loss of biodiversity (projects ANR BiodivAgrim and ACI Ecoger) and/or water contamination. CarottAge is also used for mining hydromorphological data. Actually a comparison was performed with three other algorithms classically used for the delineation of river continuum and CarottAge proved to give very interesting results for that purpose.

URL: http://carottage.loria.fr/index_in_english.html

Contacts: Florence Le Ber, Jean-François Mari

Participants: Florence Le Ber, Jean-François Mari

Partner: INRA

6.1.3 CORON

Keywords: Data mining, Closed itemset, Frequent itemset, Generator, Association rule, Rare itemset

Functional Description: The Coron platform is a KDD toolkit organized around three main components: (1) Coron-base, (2) AssRuleX, and (3) pre- and post-processing modules.

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. 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.

URL: <http://coron.loria.fr/site/index.php>

Authors: Laszlo Szathmary, Florent Marcuola, Mehdi Kaytoue, Amedeo Napoli

Contact: Amedeo Napoli

Participants: Adrien Coulet, Aleksey Buzmakov, Amedeo Napoli, Florent Marcuola, Jérémie Bourseau, Laszlo Szathmary, Mehdi Kaytoue, Victor Codocedo, Yannick Toussaint

6.2 New platforms

6.2.1 FixOut: an ensemble approach to build fairer models

Name: FixOut: an ensemble approach to build fairer models

Keywords: explanations in artificial intelligence, XAI, Lime explanations, fairness metrics, Shapley values, feature importance, feature dropout, ensemble classifier.

Functional Description: The FixOut system is aimed at evaluating fairness of processes and models in artificial intelligence. FixOut is a human-centered and model-agnostic framework, using any explanation method (based on feature importance) to assess the reliance of a classifier on sensitive features. Given a pre-trained classifier, FixOut first checks whether it relies on user defined sensitive features. If it does, then FixOut employs feature dropout to produce a pool of simplified classifiers that are then aggregated into an ensemble classifier. Empirical results using different classifiers on several real-world datasets show a consistent improvement in terms of widely used fairness metrics, decreased reliance on sensitive features, and without compromising the accuracy of classifiers.

URL: <https://fixout.loria.fr/>

Contact: Miguel Couceiro

6.2.2 LatViz: Visualization of Concept Lattices

Name: LatViz: Visualization of Concept Lattices.

Keywords: Formal Concept Analysis, pattern structures, concept lattice, implications, visualization.

Functional Description: LatViz is a tool allowing the construction, the display and the exploration of concept lattices. LatViz proposes some noticeable improvements over existing tools and introduces various functionalities focusing on interaction with experts, such as visualization of pattern structures for dealing with complex non-binary data, AOC-poset which is composed of the core elements of the lattice, concept annotations, filtering based on various criteria and a visualization of implications [67]. This way the user can effectively perform interactive exploratory knowledge discovery as often needed in knowledge engineering. The LatViz platform can be associated with the Coron platform and extends its visualization capabilities (see <http://coron.loria.fr>). Recall that the Coron platform includes a complete collection of data mining algorithms for extracting itemsets and association rules.

URL: <http://latviz.loria.fr/>

Contact: Amedeo Napoli

6.2.3 OrphaMine: Data Mining Platform for Orphan Diseases

Name: OrphaMine: A Data Mining Platform for Orphan Diseases

Keywords: bioinformatics, data mining, biology, health, data visualization, drug development.

Functional Description: The OrphaMine platform enables visualization, data integration and in-depth analytics in the domain of “orphan diseases”, where data is extracted from the OrphaData ontology (<https://www.orpha.net/consor/cgi-bin/index.php>). At present, we aim at building a true collaborative portal that will serve different actors: (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.

URL: <http://orphamine.inria.fr/>

Contact: Miguel Couceiro, Laureline Nevin, Amedeo Napoli.

6.2.4 Siren: Interactive and Visual Redescription Mining

Name: Siren: Interactive and Visual Redescription Mining

Keywords: redescription mining, interactivity, visualization.

Functional Description: Siren is a tool for interactive mining and visualization of redescrptions. Redescription mining aims to find distinct common characterizations of the same objects and, vice versa, to identify sets of objects that admit multiple shared descriptions. The goal is to provide domain experts with a tool allowing them to tackle their research questions using redescription mining. Merely being able to find redescrptions is not enough. The expert must also be able to understand the redescrptions found, adjust them to better match his domain knowledge and test alternative hypotheses with them, for instance. Thus, Siren allows mining redescrptions in an anytime fashion through efficient, distributed mining, to examine the results in various linked visualizations, to interact with the results either directly or via the visualizations, and to guide the mining algorithm toward specific redescrptions. New features, such as a visualization of the contribution of individual literals in the queries and the simplification of queries as a post-processing, have been added to the tool.

URL: <http://siren.gforge.inria.fr/main/>

Contact: Esther Galbrun

7 New results

7.1 Fundamentals of KDD

Participants Nacira Abbas, Guilherme Alves Da Silva, Alexandre Bazin, Alexandre Blansché, Lydia Boudjeloud-Assala, Brieuc Conan-Guez, Miguel Couceiro, Adrien Coulet, Sébastien Da Silva, Alain Gély, Laurine Huber, Nyoman Juniarta, Florence Le Ber, Tatiana Makhalova, Jean-François Mari, Pierre Monnin, Amedeo Napoli, Laureline Nevin, Abdelkader Ouali, François Pirot, Frédéric Pennerath, Justine Reynaud, Claire Theobald, Yannick Toussaint, Laura Alejandra Zanella Calzada, Georgios Zervakis.

7.1.1 FCA, Pattern Mining, MDL, and Hybrid Mining

Advances in data and knowledge engineering have emphasized the needs for pattern mining tools working on complex data. FCA, which usually applies to binary data-tables, can be adapted to work on more complex data. In this way, we have contributed to some main extensions of FCA, namely Pattern Structures, Relational Concept Analysis and application of the “Minimum Description Length” (MDL) within FCA. Pattern Structures (PS [71, 72]) allow building a concept lattice from complex data, e.g. numbers, sequences, trees and graphs. Relational Concept Analysis (RCA [77]) is able to analyze objects described both by binary and relational attributes and can play an important role in text classification and text mining. 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. Reusing ideas from subgroup discovery,

we have initiated a whole line of research on the covering of the pattern spaces based on the “Minimum Description Length” (MDL) principle and we are working on the adaptation of MDL within the FCA framework [33, 35]. Moreover, we also investigated the discovery of functional dependencies [20], the mining of RDF data, and the production of explanations in an FCA framework [21].

In addition, we are working on the design of hybrid mining methods, based on mining methods able to deal with symbolic and numerical data in parallel. In the context of the GEENAGE project, we are interested in the identification, in biomedical data, of biomarkers that are predictive of the development of diseases in the elderly population [21]. Actually, the data are issued from a preceding study on metabolomic data for the detection of diabetes of type 2 [12]. The problem can be viewed as a classification problem where features which are predictive of a class should be identified. This leads us to study the notions of prediction and discrimination in classification problems, and as well of hybrid mining. Combining subsymbolic machine learning methods such as random forests, neural networks, and SVM, then multicriteria decision making methods (Pareto fronts), and pattern mining methods (including FCA), we developed a hybrid mining approach for selecting the features which are the most predictive and/or discriminant. Then the selected features are organized within a concept lattice to be presented to the analyst together with the reasons for their selection. The concept lattice facilitates the understanding of the feature selection. As such, this approach can also be seen as an explicable mining method, where the output includes the reasons for which features are selected in terms of prediction and discrimination.

Besides symbolic knowledge discovery approaches, we also experimented “Bayesian deep learning” approaches in image recognition to take into account the uncertainty attached to such images and to the process classifying these images. We are dealing with astronomical images where galaxies should be recognized. First experiments trying to understand the action of feature dropout were carried out in the context of the AstroDeep project and will be consolidated the present year [42].

Redescription mining is another pattern mining method developed in the team, which aims at finding distinct common characterizations of the same objects and, reciprocally, at identifying sets of objects having multiple shared descriptions [70]. This is motivated by the idea that in scientific investigations data often have different nature, and may originate from distinct sources or terminologies. Following this way, we applied redescription mining for analyzing and mining RDF data in the web of data with the objective of discovering definitions of concepts [74, 75]. Redescription mining is well adapted to the task as a definition is naturally based on two sides of an equation, a left-hand side and a right-hand side. Moreover, we are studying how redescription mining can be used in the discovery of link keys over two RDF datasets.

7.2 KDD in Practice

Participants Alexandre Bazin, Miguel Couceiro, Adrien Coulet, Sébastien Da Silva, Florence Le Ber, Jean-François Mari, Pierre Monnin, Amedeo Napoli, Yannick Toussaint.

KDD in Life Sciences. In the ANR project PractiKPharma¹ three main research activities were developed, namely the validation of pharmacogenomic (PGX) knowledge, the mining of Electronic Health Records (EHRs), and biological data aggregation for knowledge discovery. The validation of PGX knowledge is based on the PGxO ontology² which is instantiated by data from various provenance, e.g. biomedical databases, literature, and EHRs [51]. Moreover, we defined and applied a set of so-called “reconciliation rules” that compare and align whenever possible knowledge units of different provenance [38].

We also studied how cross-corpus training may guide relation extraction from texts, and “transfer learning” techniques, i.e. how large annotated corpora developed for alternative tasks may improve the performance of text mining in biomedical texts, for which only a few annotated resources are available. Then for improving the quality of knowledge units in PGxLOD originating from biomedical literature, we manually built an annotated corpus named PGxCorpus [13]. Thanks to this corpus, the first one attached

¹<http://praktikpharma.loria.fr/>

²<https://pgxlod.loria.fr/>

to PGx, we were able to train supervised ML approaches to discover PGx knowledge units from texts with a high quality, in particular using the transformer architecture BERT [61].

Furthermore, the “Fighting Heart Failure” project aims at identifying and describing relevant bio-profiles of patients suffering from heart failure. We worked on the clustering of patients, by defining multidimensional similarity measures based on graph aggregation, namely “Unsupervised Extremely Randomized Trees” [30, 10].

Text Mining. The research work in text mining is based on two main lines. The first research subject is related to the study of discourse and argumentation structures in a text based on tree mining and redescription mining [32], while the second research work is related to the mining of Pubmed abstracts about rare diseases. In the first research line, we investigate the similarities existing between discourse and argumentation structures by aligning subtrees in a corpus where texts are annotated. Based on tree mining and redescription mining, we are able to show that the structures underlying discourse and argumentation can be (partially) aligned.

In the second research line, the objective is to discover features related to rare diseases, e.g. symptoms, related diseases, treatments, and possible disease evolution or variations. The texts to be analyzed are from Pubmed, i.e. a platform collecting millions of publications in the medical domain. This research project aims at developing new methods and tools for supporting knowledge discovery in textual data by combining methods from Natural Language Processing (NLP) and Knowledge Discovery in Databases (KDD).

Knowledge Systems and Web of Data. The increase of data published in RDF format leads to challenging aspects regarding quality assessment and data exploration [66]. For example, the discovery of implication rules is used to automatically detect possible completions of RDF data and to provide definitions. In this same way, we are investigating the characterization of functional dependencies and of similarity dependencies within FCA and pattern structures [20]. This is fully related to the ANR Elker project where we study “link keys” [6]. Link keys are more complex than functional dependencies in databases and they raise new problems to solve, e.g. they may be used to identify the same objects in different RDF datasets.

For example, link keys may identify the same book or article in different bibliographical data sources, where a link key is a statement of the form: $\{\langle \text{auteur}, \text{creator} \rangle, \langle \text{titre}, \text{title} \rangle\}$ *linkkey* $\langle \text{Livre}, \text{Book} \rangle$ stating that whenever an instance of the class *Livre* has the same values for properties *auteur* and *titre* as an instance of class *Book* has for properties *creator* and *title*, then they denote the same entity.

Currently we are investigating the connections with FCA and pattern structures [16]. Moreover, we are also interested in the possible connections with Relational Concept Analysis and redescription mining. We would like to study the formulation of the discovery of link keys in reusing and extending some construction developed in redescription mining. In practice, redescription mining aims at constructing pairs of descriptions, i.e., pairs of logical statements, one for each of two datasets, such that their support sets, i.e., the sets of objects that satisfy each statements of a pair, respectively, are most similar, as measured for example by their Jaccard index. This is in full correspondence with the discovery of link keys in two RDF datasets.

7.3 Explanations and Fairness in KDD

Participants Guilherme Alves Da Silva, Alexandre Bazin, Miguel Couceiro, Adrien Coulet, Alain Gély, Tatiana Makhalova, Pierre Monnin, Amedeo Napoli, Georgios Zervakis.

Inspired by the FCA machinery, we are seeking for a global approach to support hybridization and intelligibility of KDD systems. Indeed, the description of KDD algorithms through dimensions such as interpretability, fairness, accuracy, and tractability, leads to a classification into algorithmic concepts revealing synergies –and thus model-based interpretations and explanations– as well as suitability and compatibility w.r.t. context and fairness [69].

In [18, 22], we addressed fairness issues of KDD algorithms based on the idea of “feature dropout” followed by an “ensemble approach” to improve model fairness. All these ideas have been implemented in FIXOUT³, a human-centered and model-agnostic framework [53]. Given a classifier, a dataset and a set of sensitive features, FIXOUT first assesses whether the classifier is fair by checking its reliance on sensitive features using explanations. If deemed unfair, FIXOUT then applies feature dropout to obtain a pool of classifiers which are combined into an ensemble classifier that will be less dependent on sensitive features without compromising the whole classifier accuracy.

Future investigations are in concern with the use of different explanation methods (e.g. Shapley values) and the application to different families of classifiers, ranging from random forests to neural networks. Moreover, we also would like to combine explanation and clustering methods to understand how the classification process is working and providing possible descriptions of the resulting clusters. Possible experiments will be carried out on complex chemical datasets.

8 Partnerships and cooperations

8.1 European initiatives

8.1.1 FP7 & H2020 Projects

H2020 Tailor (2020-2023)

Participants Guilherme Alves Da Silva, Miguel Couceiro, Alain Gély, Amedeo Napoli, Georgios Zervakis.

H2020 Tailor (2020-2023): Tailor stands for “Foundations of Trustworthy AI - Integrating Reasoning, Learning and Optimization” (<https://cordis.europa.eu/project/id/952215/fr>). The purpose of TAILOR is to build a strong academic-public-industrial research network with the capacity of providing the scientific basis for Trustworthy AI. The project aims at combining learning, optimization, and reasoning, for realizing AI systems that incorporate the safeguards that make them in the reliable, safe, transparent and respectful of human agency and expectations. At Inria NGE, Orpailleur is involved with the Multispeech team in the Tailor research project.

8.2 National initiatives

8.2.1 ANR

ANR ELKER (2017–2022)

Participants Nacira Abbas, Alexandre Bazin, Miguel Couceiro, Amedeo Napoli.

The objectives of the ELKER ANR Research Project (<https://project.inria.fr/elker/>) are to study, formalize, and implement the search for link keys in RDF data [6]. Link keys generalize database keys in two independent directions, as firstly they deal with RDF data and secondly they apply across two relation datasets. In this project, we study the discovery of link keys and reasoning with link keys, being based on the FCA formalism. The ELKER project relies on the competencies of the Orpailleur Team in FCA and pattern structure algorithms, and also in partition pattern structures which are related to the discovery of functional dependencies [16]. This project involves the EPI Orpailleur at Inria Nancy Grand Est, the EPI MOEX at Inria Grenoble Rhône Alpes, and LIASD at Université Paris 8.

³<https://fixout.loria.fr/>

ANR PractiKPharma (2016–2020)

Participants Miguel Couceiro, Adrien Coulet, Pierre Monnin, Amedeo Napoli, Yannick Toussaint.

PractiKPharma for “Practice-based evidences for actioning Knowledge in Pharmacogenomics” is an ANR research project (<http://praktikpharma.loria.fr/>) about the validation of domain knowledge in pharmacogenomics. Pharmacogenomics is interested in understanding how genomic variations related to patients have an impact on drug responses. While most of the available knowledge in pharmacogenomics –state of the art knowledge– lies in the biomedical literature, with various levels of validation, an originality of PractiKPharma is to use Electronic Health Records (EHRs) to constitute cohorts of patients. These cohorts are then mined for discovering potential pharmacogenomics patterns to be validated w.r.t. literature knowledge for becoming actionable knowledge units. The PractiKPharma involves three other laboratories, namely LIRMM in Montpellier, SSPIM in St-Etienne, and CRC in Paris.

ANR AstroDeep (2019–2023)

Participants Brieuc Conan-Guez, Miguel Couceiro, Amedeo Napoli, Frédéric Pennerath, Claire Theobald.

Astronomical surveys planned for the coming years will produce data that present analysis challenges not only because of their scale (hundreds of petabytes), but also by the complexity of the measurement challenges on very deep images (for instance subpercent-level measurement of colors and shapes in blended objects). New machine learning techniques appear very promising: once trained, they are very efficient and excel at extracting features from complex images. In the AstroDeep project, we aim at developing such machine learning techniques that can be applied directly on complex images without going through the traditional steps of astronomical image processing, that lose information at each stage [42]. The developed techniques will help to leverage the observation capabilities of future surveys (LSST, Euclid, and WFIRST), and will allow a joint analysis of data.

The AstroDeep ANR Project involves three labs, namely APC Paris (“Astroparticules et Cosmologie Paris”), the Orpailleur Team at Inria Nancy Grand Est/LORIA, and “Département d’Astrophysique CEA Saclay”.

8.2.2 Inria Project Labs, Exploratory Research Actions, and Technological Development Actions

Participants Guilherme Alves Da Silva, Miguel Couceiro, Nyoman Juniarta, Amedeo Napoli, Laureline Nevin, Georgios Zervakis.

HyAIAI (IPL 2019-2022) Recent progress in Machine Learning (ML) and especially in Deep Learning has made ML present and prominent in a wide range of applications. However, current and efficient ML approaches rely on complex numerical models. Then, the decisions which are proposed may be accurate but cannot be easily explained to the layman, especially in some cases where complex and human-oriented decisions should be made, e.g. to get a loan or not, to obtain a chosen enrollment at university. The objectives of the HyAIAI IPL are to study the problem of making ML methods more interpretable. For that, we will design hybrid ML approaches that combine state of the art subsymbolic models (e.g. neural networks) with explainable symbolic models (e.g. pattern mining). More precisely, one goal is to integrate high level domain constraints into ML models, to provide model designers information on ill-performing parts of the model, and to give the layman/practitioner understandable explanations on the results of the ML model.

The HyAIAI IPL project involves seven Inria Teams, namely Lacodam in Rennes (project leader), Magnet and SequeL in Lille, Multispeech and Orpailleur in Nancy, and TAU in Saclay.

Ordem (ADT 2019-2020) One of the outputs of the former Hybride ANR project was the Orphamine system which aims at information retrieval and diagnosis aid in the domain of “rare diseases” (<http://orphamine.loria.fr/>). The Orphamine system is based on domain knowledge, and in particular on medical ontologies such as ORDO (“Orphanet Rare Diseases Ontology”) and HPO (“Human Phenotype Ontology”). The objective of the “Ordem” ADT is to update Orphamine, in making the system more accessible and more open. This mainly requires software development for building the connections with domain knowledge, graph mining methods for retrieving relevant units in knowledge graphs, actual visualization tools, pattern mining, statistical decision tools for decision making (in particular log-linear models), and as well text mining tools for analyzing expert queries and medical literature about rare diseases. A large part of this development was carried out until during the last year, for making the Orphamine system more robust. The new version of the Orphamine system will be publicly accessible through a web interface in the next future.

HyGraMi (PRE Inria 2018-2020) The so called “projet de recherche exploratoire” (PRE) HyGraMi for “Hybrid Graph Mining for the Design of New Antibacterials” is about the fight against resistance of bacteria to antibiotics. The objective of HyGraMi is to design a hybrid data mining system for discovering new antibacterial agents. This system should rely on a combination of numeric and symbolic classifiers, that will be guided by expert domain knowledge. The analysis and classification of the chemical structures is based on an interaction between symbolic methods e.g. graph mining techniques, and numerical supervised classifiers based on exact and approximate matching [54].

9 Dissemination

9.1 Promoting scientific activities

9.1.1 Scientific Events: Organization

- Laurine Huber was co-chair in the scientific organization of the “6ième conférence conjointe Journées d’Études sur la Parole (JEP), Traitement Automatique des Langues Naturelles (TALN), et Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL)”, Nancy, June 2020 (<https://jep-taln2020.loria.fr>).
- Miguel Couceiro, Amedeo Napoli, and Pierre Monnin, were the General Chairs and the scientific organizers of the online workshop ALGOS 2020 dedicated to Maurice Pouzet for his 75th birthday, August 2020 [49] (<https://algos2020.loria.fr/>).
- Amedeo Napoli was the co-Program Chair and scientific co-organizer of the 8th workshop FCA4AI co-located with ECAI 2020 in August 2020 [50] (<http://www.fca4ai.hse.ru/>).

9.1.2 General chair, scientific chair

- Amedeo Napoli was co-Program chair of the special track “General Topics of Data Analysis” at AIST in 2020 (<http://aistconf.org/>).
- See also above items 2 and 3.

9.1.3 Member of Editorial Boards

- Editor in Journal of Multiple-Valued Logic and Soft Computing and International Journal on Information Technologies and Security (Miguel Couceiro).
- Miguel Couceiro was elected as a member of the IEEE Technical Committee on Multiple-Valued Logic for the 2018-2020 period.

9.1.4 Reviewer - reviewing activities

- The members of the Orpailleur team are involved in the program and steering committees of main Conferences (AAAI, ECAI, ECML-PKDD, ICDM, IJCAI, KDD, VLDB...), as members of editorial boards, and in the organization of journal special issues.

9.1.5 Invited talks

- Invited Talk (Online): “Knowledge Discovery and Knowledge Representation: A Formal Concept Analysis Perspective”, Amedeo Napoli, Center for Strategic Research, Saint Petersburg, Russia, November 2020.
- Invited Talk (Online): “Making ML models fairer through explanations: the case of LimeOut”, Miguel Couceiro, The 9th International Conference on Analysis of Images, Social Networks and Texts, Russia [18].

9.1.6 Teaching

- All the permanent members of the Orpailleur team are involved in teaching at all levels and mainly at Université de Lorraine. Actually, most of the members of the Orpailleur team are employed on “Université de Lorraine” positions.
- Responsibility of the 2nd year of the NLP Master’s program in the IDMC, Université de Lorraine.
- Local coordination of the European Erasmus Mundus Master’s program LCT (Language and Communication Technologies). The LCT Master’s program “Language and communication Technologies” (LCT) is designed to provide students with practice-oriented knowledge in computational and theoretical linguistics, natural language processing, and computer science, to meet the demands of industry and research in these rapidly growing areas. The LCT consortium includes 7 European Universities, i.e. Saarland, Lorraine, Trento, Malta, Groningen, Charles in Prague, Basque Country, and includes several partners, e.g., DFKI, IBM (Czech Rep.), VICOMTECH, Sony (Europe), IBM (Ireland), and Inria (France).
- Responsibility in teaching courses about Artificial Intelligence and Knowledge-Based Systems at TELECOM Nancy, a engineer school for graduation in computer science at Université de Lorraine.

9.1.7 Tutorials

- Online Tutorial: “A Smooth Introduction to FCA” (AN), Troisième édition des JRI-2020 UFR-SEA, Université Joseph Ki-Zerbo, Ouagadougou, Burkina Faso, September 2020 (AN).
- Online Tutorial: “FCA and Knowledge Discovery” (AN), Tutorial at the 25th International Conference on Conceptual Structures (ICCS), Bolzano, Italy, September 18, 2020 (https://iccs-conference.org/?page_id=538).

9.1.8 Supervision

- The members of the Orpailleur team are involved in student supervision, at all university levels, from under-graduate until post-graduate students, engineers, PhD, postdoc students.
- The permanent members of the Orpailleur team are also involved in HDR and thesis defenses, being thesis referees or thesis committee members.
- PhD in progress: Nacira Abbas, discovery of link keys in the web of data, started in 2018, Jérôme David (Inria Rhône-Alpes) and Amedeo Napoli.
- PhD in progress: Guilherme Alves, Fixout: a system for checking the fairness of knowledge discovery algorithms, started in 2018, Miguel Couceiro and Amedeo Napoli.

- PhD in progress: Laurine Huber, Analysis of Discourse and Text mining, started in 2018, Yannick Toussaint.
- PhD in progress: Tatiana Mahkalova, a study of pattern mining from an FCA perspective (MDL, numerical pattern mining), started in 2017, Sergei O. Kuznetsov (HSE Moscow) and Amedeo Napoli.
- PhD in progress: Esteban Marquer, Reasoning over Data: analogy based and transfer learning to improve machine learning, started in 2020, Miguel Couceiro and Alain Gély.
- PhD defended: Pierre Monnin, Data mining and ontology design in pharmacogenomics, started in 2016, Adrien Coulet and Amedeo Napoli.
- PhD in progress: Claire Theobald, Discovering and understanding images of galaxies in astronomical data, started in 2020, Miguel Couceiro and Frédéric Pennarh.
- PhD in progress: Laura Zanella, Text mining and Deep Learning, started in 2019, Yannick Toussaint.
- PhD in progress: Georgios Zervakis, Hybrid data mining systems, started in 2019, Miguel Couceiro and Emmanuel Vincent (Multispeech).

10 Scientific production

10.1 Major publications

- [1] Q. Brabant, M. Couceiro, D. Dubois, H. Prade and A. Rico. ‘Learning rule sets and Sugeno integrals for monotonic classification problems’. In: *Fuzzy Sets and Systems* 401 (Dec. 2020), pp. 4–37. DOI: [10.1016/j.fss.2020.01.006](https://doi.org/10.1016/j.fss.2020.01.006). URL: <https://hal.inria.fr/hal-02427608>.
- [2] L. Crochepierre, L. Boudjeloud-Assala and V. Barbesant. ‘Interpretable Dimensionally-Consistent Feature Extraction from Electrical Network Sensors’. In: *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases ECML/PKDD’20*. Gand (Virtual Conference), Belgium, Sept. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02928936>.
- [3] D. Grissa, B. Comte, M. Petera, E. Pujos-Guillot and A. Napoli. ‘A hybrid and exploratory approach to knowledge discovery in metabolomic data’. In: *Discrete Applied Mathematics* 273.SI (2020), pp. 103–116. DOI: [10.1016/j.dam.2018.11.025](https://doi.org/10.1016/j.dam.2018.11.025). URL: <https://hal.inria.fr/hal-02195463>.
- [4] J. Legrand, R. Gogdemir, C. Bousquet, K. Dalleau, M.-D. Devignes, W. Digan, C.-J. Lee, N.-C. Ndiaye, N. Petitpain, P. Ringot, M. Smail-Tabbone, Y. Toussaint and A. Coulet. ‘PGxCorpus, a manually annotated corpus for pharmacogenomics’. In: *Scientific Data* 7.3 (Jan. 2020). DOI: [10.1038/s41597-019-0342-9](https://doi.org/10.1038/s41597-019-0342-9). URL: <https://hal.inria.fr/hal-02547607>.
- [5] F. Pennerath, P. Mandros and J. Vreeken. ‘Discovering Approximate Functional Dependencies using Smoothed Mutual Information’. In: *KDD 2020 - 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. San Diego / Virtual, United States, Aug. 2020. URL: <https://hal-centralesupelec.archives-ouvertes.fr/hal-02880553>.

10.2 Publications of the year

International journals

- [6] M. Atencia, J. David, J. Euzenat, A. Napoli and J. Vizzini. ‘Link key candidate extraction with relational concept analysis’. In: *Discrete Applied Mathematics* 273 (2020), pp. 2–20. DOI: [10.1016/j.dam.2019.02.012](https://doi.org/10.1016/j.dam.2019.02.012). URL: <https://hal.archives-ouvertes.fr/hal-02196757>.
- [7] Q. Brabant, M. Couceiro, D. Dubois, H. Prade and A. Rico. ‘Learning rule sets and Sugeno integrals for monotonic classification problems’. In: *Fuzzy Sets and Systems* 401 (15th Dec. 2020), pp. 4–37. DOI: [10.1016/j.fss.2020.01.006](https://doi.org/10.1016/j.fss.2020.01.006). URL: <https://hal.inria.fr/hal-02427608>.
- [8] M. Couceiro, L. Haddad and V. Lagerkvist. ‘A Survey on the Fine-grained Complexity of Constraint Satisfaction Problems Based on Partial Polymorphisms’. In: *Journal of Multiple-Valued Logic and Soft Computing* (2020). URL: <https://hal.archives-ouvertes.fr/hal-02943018>.

- [9] M. Couceiro, E. Lehtonen, P. Mercuriali and R. Péchoux. ‘On the efficiency of normal form systems for representing Boolean functions’. In: *Theoretical Computer Science* 813 (12th Apr. 2020), pp. 341–361. DOI: [10.1016/j.tcs.2020.01.009](https://doi.org/10.1016/j.tcs.2020.01.009). URL: <https://hal.inria.fr/hal-02153506>.
- [10] K. Dalleau, M. Couceiro and M. Smaïl-Tabbone. ‘Unsupervised Extra Trees: a stochastic approach to compute similarities in heterogeneous data.’ In: *International Journal of Data Science and Analytics. Advances in Intelligent Data Analysis [XVIII] - 18th International Symposium on Intelligent Data Analysis, IDA 2020 Lecture Notes in Computer Science.12080* (31st Mar. 2020), pp. 132–144. DOI: [10.1007/s41060-020-00214-4](https://doi.org/10.1007/s41060-020-00214-4). URL: <https://hal.inria.fr/hal-01982232>.
- [11] E. Fimmel, C. J. Michel, F. Pirot, J.-S. Sereni, M. Starman and L. Strümgmann. ‘The relation between k -circularity and circularity of codes’. In: *Bulletin of Mathematical Biology* 82.8 (4th Aug. 2020). DOI: [10.1007/s11538-020-00770-7](https://doi.org/10.1007/s11538-020-00770-7). URL: <https://hal.archives-ouvertes.fr/hal-02466859>.
- [12] D. Grissa, B. Comte, M. Petera, E. Pujos-Guillot and A. Napoli. ‘A hybrid and exploratory approach to knowledge discovery in metabolomic data’. In: *Discrete Applied Mathematics* 273.SI (2020), pp. 103–116. DOI: [10.1016/j.dam.2018.11.025](https://doi.org/10.1016/j.dam.2018.11.025). URL: <https://hal.inria.fr/hal-02195463>.
- [13] J. Legrand, R. Gogdemir, C. Bousquet, K. Dalleau, M.-D. Devignes, W. Digan, C.-J. Lee, N.-C. Ndiaye, N. Petitpain, P. Ringot, M. Smaïl-Tabbone, Y. Toussaint and A. Coulet. ‘PGxCorpus, a manually annotated corpus for pharmacogenomics’. In: *Scientific Data* 7.3 (2nd Jan. 2020). DOI: [10.1038/s41597-019-0342-9](https://doi.org/10.1038/s41597-019-0342-9). URL: <https://hal.inria.fr/hal-02547607>.
- [14] F. Pirot, J.-S. Sereni and R. Škrekovski. ‘Variations on the Petersen colouring conjecture’. In: *The Electronic Journal of Combinatorics* 27.1 (10th Jan. 2020). DOI: [10.37236/8515](https://doi.org/10.37236/8515). URL: <https://hal.archives-ouvertes.fr/hal-02133840>.
- [15] L. Ricci, J. Epstein, A. Buisson, C. Devos, Y. Toussaint, L. Peyrin Biroulet and F. Guillemin. ‘Flare-IBD: development and validation of a questionnaire based on patients’ messages on an internet forum for early detection of flare in inflammatory bowel disease: study protocol’. In: *BMJ Open* 10.7 (1st Oct. 2020). DOI: [10.1136/bmjopen-2020-037211](https://doi.org/10.1136/bmjopen-2020-037211). URL: <https://hal.archives-ouvertes.fr/hal-03013918>.

International peer-reviewed conferences

- [16] N. Abbas, J. David and A. Napoli. ‘Discovery of Link Keys in RDF Data Based on Pattern Structures: Preliminary Steps’. In: *CLA 2020 - The 15th International Conference on Concept Lattices and Their Applications. Proceedings of the 15th International Conference on Concept Lattices and Their Applications. Tallinn / Virtual, Estonia, 29th June 2020*. URL: <https://hal.archives-ouvertes.fr/hal-02921643>.
- [17] A. Abboud, A. Lahmadi, M. Rusinowitch, M. Couceiro, A. Bouhoula and M. Ayadi. ‘Double Mask: An efficient rule encoding for Software Defined Networking’. In: *ICIN 2020 - 23rd Conference on Innovation in Clouds, Internet and Networks and Workshops. 23rd Conference on Innovation in Clouds, Internet and Networks and Workshops, {ICIN} 2020, Paris, France, February 24-27, 2020. Paris, France, 2020*, pp. 186–193. URL: <https://hal.archives-ouvertes.fr/hal-02547097>.
- [18] G. Alves, V. Bhargava, M. Couceiro and A. Napoli. ‘Making ML models fairer through explanations: the case of LimeOut’. In: *Lecture Notes in Computer Science, Analysis of Images, Social Networks, and Texts 2020*. 9th International Conference on Analysis of Images, Social Networks, and Texts 2020 (AIST 2020). Lecture Notes in Computer Science, Analysis of Images, Social Networks, and Texts 2020. Moscow, Russia, 15th Oct. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02864059>.
- [19] G. Alves, M. Couceiro and A. Napoli. ‘Sélection de mesures de similarité pour les données catégorielles’. In: *EGC 2020 - 20ème édition de la conférence Extraction et Gestion des Connaissances. Bruxelles, Belgium, 27th Jan. 2020*. URL: <https://hal.archives-ouvertes.fr/hal-02410221>.

- [20] J. Baixeries, V. Codocedo, M. Kaytoue and A. Napoli. 'Building a Representation Context Based on Attribute Exploration Algorithms'. In: FCA4AI 2020 - 8th International Workshop "What can {FCA} do for Artificial Intelligence?". Vol. CEUR Workshop Proceedings. Proceedings of the 8th International Workshop "What can {FCA} do for Artificial Intelligence?" (FCA4AI 2020) co-located with ECAI 2020, 2729. Santiago de Compostela/Virtual, Spain, 29th Aug. 2020, pp. 141–152. URL: <https://hal.inria.fr/hal-03122343>.
- [21] A. Bazin, M. Couceiro, M.-D. Devignes and A. Napoli. 'Explaining Multicriteria Decision Making with Formal Concept Analysis'. In: Concept Lattices and Applications 2020. Vol. 2668. CEUR Workshop Proceedings. Tallinn, Estonia, 29th June 2020. URL: <https://hal.archives-ouvertes.fr/hal-02909383>.
- [22] V. Bhargava, M. Couceiro and A. Napoli. 'LimeOut: An Ensemble Approach To Improve Process Fairness'. In: *Communications in Computer and Information Science*. ECML-PKDD International Workshop on eXplainable Knowledge Discovery in Data Mining (XKDD 2020). Vol. 1323. Communications in Computer and Information Science. Ghent, Belgium, 2020, pp. 475–491. URL: <https://hal.archives-ouvertes.fr/hal-02979233>.
- [23] A. Blansch e and L. Boudjeloud-Assala. 'Biclustering it eratif par une approche  volutionnaire'. In: EGC 2020 Extraction et Gestion des Connaissances. Vol. RNTI-E-36. Brussels, Belgium, 27th Jan. 2020. URL: <https://hal.archives-ouvertes.fr/hal-03122451>.
- [24] M. Boltenhagen, B. Chetioui and L. Huber. 'An Alignment Cost-Based Classification of Log Traces Using Machine-Learning'. In: ML4PM2020 - First International Workshop on Leveraging Machine Learning in Process Mining. Padua/ Virtual, Italy, 5th Oct. 2020. URL: <https://hal.inria.fr/hal-03134114>.
- [25] G. Bonfante and M. Couceiro. 'Termination of graph rewriting systems through language theory'. In: 1st international conference "Algebras, graphs and ordered sets" (Algos 2020). Proceedings of the 1st International Conference on Algebras, Graphs and Ordered Sets (ALGOS 2020). Nancy, France: <https://algos2020.loria.fr/>, 26th Aug. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02912877>.
- [26] M. Couceiro and M. Grabisch. 'Monotonic computation rules for nonassociative calculus'. In: 1st international conference on ALgebras, Graphs and Ordered Sets. Proceedings of the 1st International Conference on Algebras, Graphs and Ordered Sets (ALGOS 2020). Nancy, France: <https://algos2020.loria.fr/>, 15th May 2020. URL: <https://hal.archives-ouvertes.fr/hal-02912878>.
- [27] M. Couceiro and E. Lehtonen. 'Linearly definable classes of Boolean functions'. In: 1st international conference "Algebras, graphs and ordered sets" (Algos 2020). Proceedings of the 1st International Conference on Algebras, Graphs and Ordered Sets (ALGOS 2020). Nancy, France: <https://algos2020.loria.fr/>, 26th Aug. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02912876>.
- [28] M. Couceiro, E. Lehtonen, L. Miclet, H. Prade and G. Richard. 'When nominal analogical proportions do not fail'. In: *Springer Lecture Notes in Computer Science (LNCS)*. SUM 2020 - 14th International Conference on Scalable Uncertainty Management. Vol. 12322. Lecture Notes in Computer Science. Bolzano / Virtual, Italy, 23rd Sept. 2020, pp. 68–83. URL: <https://hal.archives-ouvertes.fr/hal-02864071>.
- [29] L. Crochepierre, L. Boudjeloud-Assala and V. Barbesant. 'Interpretable Dimensionally-Consistent Feature Extraction from Electrical Network Sensors'. In: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases ECML/PKDD'20. Gand (Virtual Conference), Belgium, 14th Sept. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02928936>.
- [30] K. Dalleau, M. Couceiro and M. Sma il-Tabbone. 'Computing Vertex-Vertex Dissimilarities Using Random Trees: Application to Clustering in Graphs'. In: IDA 2020 -18th International Symposium on Intelligent Data Analysis. Vol. 12080. Lecture Notes in Computer Science. Konstanz / Virtual, Germany: <https://ida2020.org/>, 27th June 2020, pp. 132–144. DOI: 10.1007/978-3-030-44584-3_11. URL: <https://hal.archives-ouvertes.fr/hal-02864678>.

- [31] A. Gély, M. Couceiro, L. Miclet and A. Napoli. ‘Steps in the Representation of Concept Lattices and Median Graphs’. In: CLA 2020 - 15th International Conference on Concept Lattices and Their Applications. Proceedings of The 15th International Conference on Concept Lattices and Their Applications. Tallinn, Estonia: <https://cs.ttu.ee/events/cla2020/>, 29th June 2020, pp. 1–11. URL: <https://hal.inria.fr/hal-02912312>.
- [32] L. Huber, J. Reynaud, M. Dargnat and Y. Toussaint. ‘AOC-Poset on discourse and argumentation subgraphs: what can we learn on their dependencies?’ In: Concept Lattice and their Applications. Vol. 2668. Proceedings of the Fifteenth International Conference on Concept Lattices and Their Applications. Tallinn, Estonia, 29th June 2020. URL: <https://hal.archives-ouvertes.fr/hal-02909216>.
- [33] T. Makhalova, S. O. Kuznetsov and A. Napoli. ‘Adaptive Multi-model Approaches to Pattern Set Mining’. In: European Starting AI Researchers’ Symposium 2020. Vol. CEUR 2655. Proceedings of the 9th European Starting AI Researchers’ Symposium 2020 co-located with 24th European Conference on Artificial Intelligence ECAI 2020. Santiago Compostela, Spain, 30th Aug. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02953084>.
- [34] T. Makhalova, S. O. Kuznetsov and A. Napoli. ‘Closure Structure: a Deeper Insight’. In: The 8th International Workshop "What can FCA do for Artificial Intelligence?" Vol. CEUR 2729. Proceedings of the 8th International Workshop "What can FCA do for Artificial Intelligence?" (FCA4AI 2020) co-located with 24th European Conference on Artificial Intelligence (ECAI 2020). Santiago de Compostela, Spain, 29th Aug. 2020. URL: <https://hal.archives-ouvertes.fr/hal-03121356>.
- [35] T. Makhalova, S. O. Kuznetsov and A. Napoli. ‘Gradual Discovery with Closure Structure of a Concept Lattice’. In: The 15th International Conference on Concept Lattices and Their Applications. Vol. CEUR 2668. Proceedings of the 15th International Conference on Concept Lattices and Their Applications. Tallinn, Estonia, 6th Oct. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02915309>.
- [36] E. Marquer, A. Kulkarni and M. Couceiro. ‘Embedding Formal Contexts Using Unordered Composition’. In: FCA4AI - 8th International Workshop "What can FCA do for Artificial Intelligence?" (colocated wit ECAI2020). Santiago de Compostela, Spain, 2020. URL: <https://hal.archives-ouvertes.fr/hal-02912874>.
- [37] P. Monnin, E. Bresso, M. Couceiro, M. Smaïl-Tabbone, A. Napoli and A. Coulet. ‘Tackling scalability issues in mining path patterns from knowledge graphs: a preliminary study’. In: 1st international conference "Algebras, graphs and ordered sets" (Algos 2020). Proceedings of the 1st International Conference on Algebras, Graphs and Ordered Sets (ALGOS 2020). Nancy, France: <https://algos2020.loria.fr/>, 26th Aug. 2020. URL: <https://hal.inria.fr/hal-02913224>.
- [38] P. Monnin, M. Couceiro, A. Napoli and A. Coulet. ‘Knowledge-Based Matching of n-ary Tuples’. In: *Ontologies and Concepts in Mind and Machine - 25th International Conference on Conceptual Structures, ICCS 2020, Bolzano, Italy, September 18–20, 2020, Proceedings*. 25th International Conference on Conceptual Structures, ICCS 2020. Vol. Lecture Notes in Computer Science. Ontologies and Concepts in Mind and Machine - 25th International Conference on Conceptual Structures, ICCS 2020, Bolzano, Italy, September 18–20, 2020, Proceedings 12277. Bolzano, Italy, 2020, pp. 48–56. DOI: [10.1007/978-3-030-57855-8_4](https://doi.org/10.1007/978-3-030-57855-8_4). URL: <https://hal.inria.fr/hal-02493067>.
- [39] A. Napoli. ‘FCA and Knowledge Discovery (Tutorial)’. In: ICCS 2020 - 25th International Conference on Conceptual Structures. Vol. Lecture Notes in Computer Science. Proceedings: Ontologies and Concepts in Mind and Machine - 25th International Conference on Conceptual Structures, ICCS 2020, Bolzano, Italy, September 18-20, 2020. 12277. Bolzano/ Virtual, Italy, 18th Sept. 2020. DOI: [10.1007/978-3-030-57855-8](https://doi.org/10.1007/978-3-030-57855-8). URL: <https://hal.inria.fr/hal-03122350>.
- [40] F. Pennerath, P. Mandros and J. Vreeken. ‘Discovering Approximate Functional Dependencies using Smoothed Mutual Information’. In: KDD 2020 - 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. KDD '20: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. San Diego / Virtual, United States, 22nd Aug. 2020,

pp. 1254–1264. URL: <https://hal-centralesupelec.archives-ouvertes.fr/hal-02880553>.

Conferences without proceedings

- [41] L. Huber, C. Memmadi, M. Dagnat and Y. Toussaint. ‘Do Sentence Embeddings Capture Discourse Properties of Sentences from Scientific Abstracts ?’ In: CODI 2020 - EMNLP 1st Workshop on Computational Approaches to Discourse. Punta Cana, Dominican Republic, 20th Nov. 2020. URL: <https://hal.archives-ouvertes.fr/hal-03000308>.
- [42] C. Theobald, F. Pennerath, B. Conan-Guez, M. Couceiro and A. Napoli. ‘A Bayesian Neural Network based on Dropout Regulation’. In: Workshop on Uncertainty in Machine Learning (WUML) at ECML-PKDD 2020 Conference. N.A. (online), France, 2020. URL: <https://hal.inria.fr/hal-03122764>.

Scientific book chapters

- [43] S. Ferré, M. Huchard, M. Kaytoue, S. Kuznetsov and A. Napoli. ‘Formal Concept Analysis: From Knowledge Discovery to Knowledge Processing’. In: *A Guided Tour of Artificial Intelligence Research*. Vol. 2. AI Algorithms. 8th May 2020, pp. 411–445. DOI: [10.1007/978-3-030-06167-8_13](https://doi.org/10.1007/978-3-030-06167-8_13). URL: <https://hal.archives-ouvertes.fr/hal-03085699>.
- [44] B. Fuchs, J. Lieber, L. Miclet, A. Mille, A. Napoli, H. Prade and G. Richard. ‘Case-Based Reasoning, Analogy, and Interpolation’. In: *A Guided Tour of Artificial Intelligence Research*. Vol. Volume I: Knowledge Representation, Reasoning and Learning. A Guided Tour of Artificial Intelligence Research. https://link.springer.com/chapter/10.1007%2F978-3-030-06164-7_10, 8th May 2020, pp. 307–339. DOI: [10.1007/978-3-030-06164-7_10](https://doi.org/10.1007/978-3-030-06164-7_10). URL: <https://hal.inria.fr/hal-03119305>.

Edition (books, proceedings, special issue of a journal)

- [45] C. Benzitoun, C. Braud, L. Huber, D. Langlois, S. Ouni, S. Pogodalla and S. Schneider, eds. *Actes de la 6e conférence conjointe Journées d’Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 2 : Traitement Automatique des Langues Naturelles*. JEP-TALN-RECITAL 2020. Vol. 2. Volume 2 : Traitement Automatique des Langues Naturelles. Nancy, France: <http://talnarchives.atala.org/TALN/TALN-2020/>, 2020. URL: <https://hal.archives-ouvertes.fr/hal-02784750>.
- [46] C. Benzitoun, C. Braud, L. Huber, D. Langlois, S. Ouni, S. Pogodalla and S. Schneider, eds. *Actes de la 6e conférence conjointe Journées d’Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 1 : Journées d’Études sur la Parole*. Nancy, France, 2020. URL: <https://hal.archives-ouvertes.fr/hal-02798507>.
- [47] C. Benzitoun, C. Braud, L. Huber, D. Langlois, S. Ouni, S. Pogodalla and S. Schneider, eds. *Actes de la 6e conférence conjointe Journées d’Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 3 : Rencontre des Étudiants Chercheurs en Informatique pour le TAL*. JEP-TALN-RECITAL 2020 : 6e conférence conjointe Journées d’Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Vol. 3. Volume 3 : Rencontre des Étudiants Chercheurs en Informatique pour le TAL. Nancy, France: <http://talnarchives.atala.org/RECITAL/RECITAL-2020/>, 2020. URL: <https://hal.archives-ouvertes.fr/hal-02786181>.

- [48] C. Benzitoun, C. Braud, L. Huber, D. Langlois, S. Ouni, S. Pogodalla and S. Schneider, eds. *Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 4 : Démonstrations et résumés d'articles internationaux*. JEP-TALN-RECITAL 2020. Vol. 4. Volume 4 : Démonstrations et résumés d'articles internationaux. Nancy, France: <http://talnarchives.atala.org/TALN/TALN-2020/>, 2020. URL: <https://hal.archives-ouvertes.fr/hal-02768750>.
- [49] M. Couceiro, P. Monnin and A. Napoli, eds. *Proceedings of the 1st International Conference on Algebras, Graphs and Ordered Sets (ALGOS 2020)*. 1st International Conference on Algebras, Graphs and Ordered Sets (ALGOS 2020). Proceedings of the 1st International Conference on Algebras, Graphs and Ordered Sets (ALGOS 2020). Nancy, France: <https://algos2020.loria.fr/>, Aug. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02918958>.
- [50] S. O. Kuznetsov, A. Napoli and S. Rudolph, eds. *Eighth International Workshop "What can FCA do for Artificial Intelligence?" (FCA4AI at ECAI 2020)*. 8th International Workshop "What can {FCA} do for Artificial Intelligence?" (FCA4AI 2020). Vol. CEUR Workshop Proceedings (CEUR-WS.org). Proceedings of the 8th International Workshop "What can FCA do for Artificial Intelligence?" (FCA4AI 2020) 2729. Santiago de Compostela, Spain: <http://ceur-ws.org/Vol-2729>, 6th Nov. 2020, p. 153. URL: <https://hal.inria.fr/hal-03119295>.

Doctoral dissertations and habilitation theses

- [51] P. Monnin. 'Matching and mining in knowledge graphs of the Web of data - Applications in pharmacogenomics'. Université de Lorraine, 16th Dec. 2020. URL: <https://hal.inria.fr/tel-03122326>.

Reports & preprints

- [52] N. Abbas, K. Alghamdi, M. Alinam, F. Alloatti, G. Amaral, C. d'Amato, L. Asprino, M. Beno, F. Bensmann, R. Biswas et al. *Knowledge Graphs Evolution and Preservation – A Technical Report from ISWS 2019*. 15th Jan. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03111661>.
- [53] G. Alves, V. Bhargava, F. Bernier, M. Couceiro and A. Napoli. *FixOut: an ensemble approach to fairer models*. 1st Dec. 2020. URL: <https://hal.archives-ouvertes.fr/hal-03033181>.
- [54] C. Bellanger, J. Hung, N. Juniarta, V. Leroux, B. Maigret and A. Napoli. *Chemoinformatics approaches to help antibacterial discovery*. Inria Nancy - Grand Est, 22nd May 2020. URL: <https://hal.inria.fr/hal-02615395>.
- [55] G. Bonfante and M. Couceiro. *On termination of Graph Rewriting Systems through language theory*. 12th Feb. 2020. URL: <https://hal.inria.fr/hal-02478696>.
- [56] M. Couceiro and E. Lehtonen. *Stability of Boolean function classes with respect to clones of linear functions*. 17th Jan. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03119362>.
- [57] T. Makhalova, S. O. Kuznetsov and A. Napoli. *Discovery data topology with the closure structure. Theoretical and practical aspects*. 17th Dec. 2020. URL: <https://hal.archives-ouvertes.fr/hal-03079892>.
- [58] T. Makhalova, S. O. Kuznetsov and A. Napoli. *Mint: MDL-based approach for Mining INTEResting Numerical Pattern*. 17th Dec. 2020. URL: <https://hal.archives-ouvertes.fr/hal-03079884>.
- [59] J.-F. Mari, O. Horn and M. Benoît. *Stochastic and spectral analysis of time series for identifying periodic land use*. 21st Sept. 2020. URL: <https://hal.inria.fr/hal-02944574>.
- [60] F. Pirot and J.-S. Sereni. *Fractional chromatic number, maximum degree and girth*. 23rd Nov. 2020. URL: <https://hal.archives-ouvertes.fr/hal-02096426>.

Other scientific publications

- [61] W. Hafiane. ‘Apprentissage par transfert pour l’extraction de relations pharmacogénomiques à partir de textes’. Université de Lorraine, 9th Sept. 2020. URL: <https://hal.inria.fr/hal-02939161>.
- [62] L. Magno, A. Mollier, M. Couceiro, J.-C. Avice and S. Brunel-Muguet. *Modelling Memory: do crop models need to become nostalgic?* Montpellier, France, 3rd Feb. 2020. URL: <https://hal.inria.fr/hal-02737323>.
- [63] E. Marquer. ‘LatticeNN – Deep Learning and Formal Concept Analysis’. Université de Lorraine, 2nd Sept. 2020, p. 74. URL: <https://hal.inria.fr/hal-02954843>.

10.3 Cited publications

- [64] C. C. Aggarwal and C. Zhai, eds. *Mining Text Data*. Springer, 2012.
- [65] M. Alam, A. Buzmakov, V. Codocedo and A. Napoli. ‘Mining Definitions from RDF Annotations Using Formal Concept Analysis’. In: *Proceedings of IJCAI*. AAAI Press, 2015, pp. 823–829.
- [66] M. Alam, A. Buzmakov and A. Napoli. ‘Exploratory Knowledge Discovery over Web of Data’. In: *Discrete Applied Mathematics* 249 (2018), pp. 2–17. DOI: [10.1016/j.dam.2018.03.041](https://doi.org/10.1016/j.dam.2018.03.041). URL: <https://hal.inria.fr/hal-01673439>.
- [67] M. Alam, T. N. N. Le and A. Napoli. ‘LatViz: A New Practical Tool for Performing Interactive Exploration over Concept Lattices’. In: *Proceedings of the 13th International Conference on Concept Lattices and Their Applications (CLA)*. CEUR Workshop Proceedings 1624. 2016, pp. 9–20.
- [68] F. Baader, D. Calvanese, D. McGuinness, D. Nardi and P. Patel-Schneider, eds. *The Description Logic Handbook*. Cambridge, UK: Cambridge University Press, 2003.
- [69] M. Couceiro and A. Napoli. ‘Elements About Exploratory, Knowledge-Based, Hybrid, and Explainable Knowledge Discovery’. In: *ICFCA 2019 - 15th International Conference on Formal Concept Analysis*. Ed. by D. Cristea, F. L. Ber and B. Sertkaya. Vol. 11511. Formal Concept Analysis. Frankfurt, Germany: Springer International Publishing, June 2019, pp. 3–16. DOI: [10.1007/978-3-030-21462-3_1](https://doi.org/10.1007/978-3-030-21462-3_1). URL: <https://hal.inria.fr/hal-02195480>.
- [70] E. Galbrun and P. Miettinen. ‘Mining redescrptions with Siren’. In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 12.1 (2018), 6:1–6:30. DOI: [10.1145/3007212](https://doi.org/10.1145/3007212). URL: <https://hal.archives-ouvertes.fr/hal-01399213>.
- [71] B. Ganter and S. O. Kuznetsov. ‘Pattern Structures and Their Projections’. In: *Proceedings of ICCS 2001*. LNCS 2120. Springer, 2001, pp. 129–142.
- [72] M. Kaytoue, S. O. Kuznetsov, A. Napoli and S. Duplessis. ‘Mining gene expression data with pattern structures in formal concept analysis’. In: *Information Sciences* 181.10 (2011), pp. 1989–2001. URL: <https://hal.archives-ouvertes.fr/hal-00541100>.
- [73] J.-P. Metivier, A. Lepailleur, A. Buzmakov, G. Poezevara, B. Crémilleux, S. O. Kuznetsov, J. Le Goff, A. Napoli, R. Bureau and B. Cuissart. ‘Discovering structural alerts for mutagenicity using stable emerging molecular patterns’. In: *Journal of Chemical Information and Modeling* 55.5 (2015), pp. 925–940. DOI: [10.1021/ci500611v](https://doi.org/10.1021/ci500611v). URL: <https://hal.archives-ouvertes.fr/hal-01186716>.
- [74] J. Reynaud, Y. Toussaint and A. Napoli. ‘Redescription mining for learning definitions and disjointness axioms in Linked Open Data’. In: *ICCS 2019 - 24th International Conference on Conceptual Structures*. Marburg, Germany, July 2019. URL: <https://hal.inria.fr/hal-02170763>.
- [75] J. Reynaud, Y. Toussaint and A. Napoli. ‘Using Redescrptions and Formal Concept Analysis for Mining Definitions Linked Data’. In: *ICFCA 2019 - 15th International Conference on Formal Concept Analysis*. Ed. by D. Cristea, F. L. Ber and B. Sertkaya. Vol. 11511. Formal Concept Analysis. Francfort, Germany: Springer International Publishing, June 2019, pp. 241–256. URL: <https://hal.inria.fr/hal-02170760>.

- [76] M. T. Ribeiro, S. Singh and C. Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier'. In: *Proceedings of KDD*. ACM, 2016, pp. 1135–1144.
- [77] M. Rouane-Hacene, M. Huchard, A. Napoli and P. Valtchev. 'Relational Concept Analysis: Mining Concept Lattices From Multi-Relational Data'. In: *Annals of Mathematics and Artificial Intelligence* 67.1 (Jan. 2013), pp. 81–108. DOI: [10.1007/s10472-012-9329-3](https://doi.org/10.1007/s10472-012-9329-3). URL: <http://hal.inria.fr/lirmm-00816300>.
- [78] T. Speicher, H. Heidari, N. Grgic-Hlaca, K. P. Gummadi, A. Singla, A. Weller and M. B. Zafar. 'A Unified Approach to Quantifying Algorithmic Unfairness: Measuring Individual Unfairness via Inequality Indices'. In: *Proceedings of KDD*. ACM, 2018, pp. 2239–2248.
- [79] M. B. Zafar, I. Valera, M. Gomez-Rodriguez and K. P. Gummadi. 'Fairness Constraints: A Flexible Approach for Fair Classification'. In: *Journal of Machine Learning Research* 20 (2019), 75:1–75:42.