Activity Report 2021

Team SEMLIS

Semantics, Logics, Information Systems for Data-User Interaction

D7 – Data and Knowledge Management

1 Team composition

Researchers and faculty

Peggy Cellier, Associate Professor (HDR), INSA Rennes (moved to LACODAM mid-year)
Mireille Ducassé, Professor, INSA Rennes
Sébastien Ferré, Professor, Univ. Rennes 1, *head of the team*
Annie Foret, Associate Professor (HDR), Univ. Rennes 1
Olivier Ridoux, Professor, Univ. Rennes 1

Associate members

Shridhar B. Dandin, Sarala Birla University, Ranchi, India
Archil Elizbarashvili, Ivane Javakhishvili Tbilisi State University, Georgia

PhD students

Hugo Ayats
Francesco Bariatti (as ATER since 01/09)
Julie Boudebs (since September)
Aurélien Lamercerie, co-supervised with team HYCOMES

Administrative assistant

Gaëlle Tworkowski

2 Overall objectives

2.1 Overview

In a context of ever-increasing volumes of data and knowledge, both in quantity and in diversity (Big Data), **the main objective of SemLIS is to bring back to users the power on their data**. By users we mean any individual or group who has a strong interest over some data, and the need to exploit them in order to derive new knowledge and to take decisions. That includes tasks such as search, authoring, data mining, and business intelligence. Those data can range from the personal data of an individual to the information systems of large companies, through project management inside a team. We take a subjective view on "Big Data" where the complexity does not lie in efficiently performing a given task on a large volume of data (e.g., query evaluation), but in enabling users to perform tasks that could not be anticipated (e.g, query formulation). In that subjective view, "Big" only means an amount of data that is too large or too complex for users to grasp and analyze by hand or by simple tools (e.g., spreadsheets).

Our objectives fit in the scope of axis 26 (human-machine collaboration) of challenge 7 (society of information and communication) of the national strategy for research. We particularly agree with the notion of man-machine collaboration, where the machine is not supposed, in our view, to *replace* humans by full automation, but rather to *support* them in information-intensive tasks. In this view, both the human and the machine should learn one from the other.

> *One will review the human-computer interaction in the light of natural human behavior and progress in the decisional and operational autonomy of machines. To develop a real collaboration between man and machine, research on self-learning process between man and machine must be amplified. The machine should adapt to unpredictable aspects of user behavior, and develop a greater wealth of interactions for "intelligent" automation.*

That main objective of **bringing back to users the power on their data** can be decomposed into five high-level objectives:

AUTO (O1): to make users **autonomous and agile** in the process of exploiting data and knowledge by avoiding intermediates (e.g., database administrators);

SEM (O2): to facilitate the **semantic** representation and alignment of heterogeneous and multi-source data;

FLEX (O3): to provide **flexibility** by enabling out-of-schema data acquisition, and continuous evolution of the data schema;

CON (O4): to provide **control and confidence** in the information system by promoting transparency and predictability of system actions;

COLL (O5): to support the **collaborative** acquisition and verification of data and knowledge.

Those objectives are the different facets of a unique approach that targets user guidance as a trade-off between full automation (aka. artificial intelligence) and no automation (aka. adhoc programming). We are conscious that this set of objectives is ambitious but we think we can address them because we do not target the hard problems of full automation, and because we now have an effective design pattern, ACN (Abstract Conceptual Navigation) [Fer14a], to encapsulate an expressive formal language into data-user interaction and natural language.

2.2 Scientific foundations

A distinctive aspect of our team is the application of formal methods coming from software engineering and theoretical computer science (formal languages and grammars, logics, type theory, declarative programming languages, theorem proving) to artificial intelligence tasks (knowledge representation and reasoning, data mining, user-data interaction). This is explained by the combination of a theoretical background shared by permanent members and a real interest for data and their users. Some members, Olivier Ridoux and Mireille Ducassé, have had a long research experience in software engineering in general, and in logic programming in particular. Annie Foret studies different variants of substructural logics for the analysis of natural languages. Peggy Cellier did her PhD thesis on the application of data mining to the localization of faults in programs [CDFR18]. Sébastien Ferré relies on formal languages to formalize user-data interaction models, and to prove usability properties such as the safeness and completeness of user guidance.

We briefly describe the scientific foundations of the team, organized by high-level research topics, along with references to a few former contributions in each topic.

2.2.1 Knowledge Representation and Querying

The team uses symbolic approaches, and in particular the Semantic Web technologies [AvH04,HKR09]. Indeed, those are an active research domain, and provide W3C standards for concepts introduced by widely recognized formalisms for knowledge representation: e.g., Datalog [CGT89], description logics [BCM+03], or conceptual graphs [CM08]. The Semantic Web defines languages for the representation of facts and rules (RDF, RDFS, OWL, SWRL), and for their querying (SPARQL). Moreover, the Semantic Web has an active community, both in academy and in industry. That research domain solicits competencies in formal languages (syntax and semantics), in logics, and in automated

[AvH04] G. Antoniou, F. van Harmelen, *A Semantic Web Primer*, MIT Press, 2004.

[HKR09] P. Hitzler, M. Krötzsch, S. Rudolph, *Foundations of Semantic Web Technologies*, Chapman & Hall/CRC, 2009.

[CGT89] S. Ceri, G. Gottlob, L. Tanca, "What you Always Wanted to Know About Datalog (And Never Dared to Ask)", *IEEE Trans. Knowl. Data Eng. 1*, 1, 1989, p. 146–166.

[BCM+03] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, P. F. Patel-Schneider (editors), *The Description Logic Handbook: Theory, Implementation, and Applications*, Cambridge University Press, 2003.

[CM08] M. Chein, M.-L. Mugnier, *Graph-based knowledge representation: computational foundations of conceptual graphs*, *Advanced Information and Knowledge Processing*, Springer, 2008.

reasoning.

2.2.2 Natural Language Processing

Here again, the team uses symbolic approaches. One task is to extract structured and semantic information from texts. The employed techniques are: a) categorial grammars [MR12] associating syntactic/semantic types to words, b) Montague grammars [DWP81] associating grammars, lambda calcul, and logic, and c) sequential patterns [AS95]. Those techniques can be used for syntactic/semantic analysis of sentences, for Information Extraction (IE), for Semantic Elevation (SE), and for defining Controlled Natural Languages (CNL) [Kuh13]. In those topics, we have for instance contributed to the learnability of pregroup grammars [BFT07], and their extension with option and iteration [BDF12], to a CNL (SQUALL) for querying and updating RDF graphs [Fer14b], and to the discovery of linguistic patterns from texts [BCCC12].

2.2.3 Symbolic Data Mining

The team has competencies in the conception and application of symbolic data mining algorithms, in particular for sequential patterns, and their application to texts. It also has competencies in learning the grammar of natural languages from a structured corpus [BFT07, FB19]. Moreover, the LIS team was scientifically founded on Formal Concept Analysis (FCA) [GW99]. It produced FCA-based contributions for data mining [CFRD08] and machine learning [FR02], as well as for data exploration [FH12].

2.2.4 User-Data Interaction

Because of the importance that we give to user-data interaction, the team invested into techniques that enable to structure and reason on those interactions. We can refer, in particular, to faceted search [ST09] (often used in e-commerce platforms), On-Line Analytical Processing (OLAP, often used in business intelligence) [CCS93], and Geographical

---

[MR12] R. Moot, C. Retoré, *The Logic of Categorial Grammars: A Deductive Account of Natural Language Syntax and Semantics*, *FoLLI-LNCS*, Springer, 2012, `https://hal.archives-ouvertes.fr/hal-00829051`.

[DWP81] D. R. Dowty, R. E. Wall, S. Peters, *Introduction to Montague Semantics*, D. Reidel Publishing Company, 1981.

[AS95] R. Agrawal, R. Srikant, "Mining Sequential Patterns", *in : Proceedings of the Eleventh International Conference on Data Engineering, ICDE '95*, IEEE Computer Society, p. 3–14, 1995.

[Kuh13] T. Kuhn, "A Survey and Classification of Controlled Natural Languages", *Computational Linguistics*, 2013.

[GW99] B. Ganter, R. Wille, *Formal Concept Analysis — Mathematical Foundations*, Springer, 1999.

[ST09] G. M. Sacco, Y. Tzitzikas (editors), *Dynamic taxonomies and faceted search, The information retrieval series*, Springer, 2009.

[CCS93] E. Codd, S. Codd, C. Salley, *Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate*, Codd & Date, Inc, San Jose, 1993.

Information Systems (GIS) [LT92]. In those topics, we have for instance contributed to the exploration of geographical data [BFRQ08], to the discovery of functional dependencies and association rules with OLAP cubes [AFR10], and to the extension of faceted search to RDF graphs [FH12].

2.3 Application Domains

The application field of SemLIS is widely open as it covers the field of the Semantic Web. According to a study done in September 2011, the Semantic Web that is available as Linked Open Data (LOD) counts 30 billions triples covering many domains: e.g., life sciences, media, governmental organizations, publications, geography. In addition to those public data, we can count the numerous internal data of companies and other organizations, as well as personal data. Social networks and wikis are yet another source of semantic data: e.g., photo annotations, relationships between people, restaurant ratings.

The approach to applications of the team is to first design generic information systems, then to evaluate the generic design on different use cases or domains, and finally to specialize and adapt it to a particular application if need be. This follows software engineering of reusability and orthogonality.

Our past and current experiences and collaborations have led us to target in priority the large domains below. In particular, we target users in the middle of the spectrum going from pure IT people to the general public, i.e., individuals and groups who are experts in a domain that implies data and knowledge management. Our objective is to enable those users to perform tasks that normally require IT technical competencies.

Social Sciences. Here, users are often other researchers in domains that have been strongly impacted by the increasing availability of digital data: e.g., geography, linguistics, law, group decision and negotiation. Our objective is not to solve their own scientific problems, but to make those users more autonomous and more efficient in the management and exploration of their data, and to guide them in the knowledge extraction process.

Business Intelligence. Here, users are groups of various sizes (e.g., teams, committees, companies, organizations) collaborating around one or several projects (e.g., strategic orientation, recruitment process). Our priority will go to small- to medium-sized groups because our emphasis is on expressivity rather than scalability. The objective is to enable a group to capitalize facts and knowledge continuously, to analyze data for self-evaluation or diagnostic, and help in decision making. To be effective, those functions should be coupled with information systems and private social networks.

[LT92] R. Laurini, D. Thompson, *Fundamentals of Spatial Information Systems*, Elsevier, Academic Press Limited, 1992.

3 Scientific achievements

3.1 Extracting Relations in Text with Concepts of Neighbours

Participants: Hugo Ayats, Sébastien Ferré, Peggy Cellier.

During the last decade, the need for reliable and massive Knowledge Graphs (KG) increased. KGs can be created in several ways: manually with forms or automatically with Information Extraction (IE), a natural language processing task for extracting knowledge from text. Relation Extraction is the part of IE that focuses on identifying relations between named entities in texts, which amounts to find new edges in a KG. Most recent approaches rely on deep learning, achieving state-of- the-art performances. However, those performances are still too low to fully automatize the construction of reliable KGs, and human interaction remains necessary. This is made difficult by the statistical nature of deep learning methods that makes their predictions hardly interpretable. In this work [8], we presented a new symbolic and interpretable approach for Relation Extraction in texts. It is based on a modeling of the lexical and syntactic structure of text as a knowledge graph, and it exploits Concepts of Neighbours, a method based on Graph-FCA for computing similarities in knowledge graphs. An evaluation has been performed on a subset of TACRED (a relation extraction benchmark), showing promising results.

3.2 Mining Tractable Sets of Graph Patterns with the Minimum Description Length Principle

Participants: Francesco Bariatti, Peggy Cellier, Sébastien Ferré.

Many graph pattern mining algorithms have been designed to identify recurring structures in graphs. The main drawback of these approaches is that they often extract too many patterns for human analysis. Recently, pattern mining methods using the Minimum Description Length (MDL) principle have been proposed to select a characteristic subset of patterns from transactional, sequential and relational data. We have proposed MDL-based approaches for selecting a characteristic subset of patterns on labeled graphs [9, 1]. A key notion in this work is the introduction of ports to encode connections between pattern occurrences without any loss of information. Experiments performed on real-life datasets from different domains show that the number of patterns is drastically reduced and that the extracted patterns help human analysis of the data. The selected patterns have complex shapes and are representative of the data.

3.3 Visualization of Databases

Participants: Shridhar B. Dandin, Mireille Ducassé.

Interpreting data with many attributes is a difficult issue. A simple 2D display, projecting two attributes onto two dimensions, is relatively easy to interpret but provides limited help to see multidimensional correlations. We propose a tool, ComVisMD, which

displays, from a dataset, five dimensions in compact 2D maps. A map contains cells; each one represents an object from the dataset. In addition to the usual horizontal and vertical projections and the use of colors, we offer holes and shapes. In order to compact the display, we partition objects according to two dimensions, grouping values of each dimension into up to seven categories. This year's work focused on two case studies covering two different domains, a cricket player dataset and a heart disease dataset. The cricket dataset has 15 attributes and 2170 objects. We showed how, using ComVisMD, correlations between variables can be found in an intuitive way. The heart disease dataset has 14 attributes and 297 objects. Blokh and Stambler, in the June 2015 issue of "Aging and Disease," state that individual attributes show little correlation with heart disease. Yet in combination the correlation improves dramatically. We showed how ComVisMD helps visualize those multidimensional correlations between four attributes and heart disease diagnosis. This activity, done in collaboration with an international partner, has led to a publication of a chapter in a book [4].

3.4 Application of Concepts of Neighbours to Knowledge Graph Completion

Participants: Sébastien Ferré.

The open nature of Knowledge Graphs (KG) often implies that they are incomplete. Knowledge graph completion (aka. link prediction) consists in inferring new relationships between the entities of a KG based on existing relationships. Most existing approaches rely on the learning of latent feature vectors for the encoding of entities and relations. In general however, latent features cannot be easily interpreted. Rule-based approaches offer interpretability but a distinct ruleset must be learned for each relation. In both latent- and rule-based approaches, the training phase has to be run again when the KG is updated. We proposed a new approach [6] that does not need a training phase, and that can provide interpretable explanations for each inference. It relies on the computation of Concepts of Nearest Neighbours (C-NN) to identify clusters of similar entities based on common graph patterns. Different rules are then derived from those graph patterns, and combined to predict new relationships. We evaluate our approach on standard benchmarks for link prediction, where it gets competitive performance compared to existing approaches.

3.5 Conceptual Navigation in Large Knowledge Graphs

Participants: Sébastien Ferré.

A growing part of Big Data is made of knowledge graphs. Major knowledge graphs such as Wikidata, DBpedia or the Google Knowledge Graph count millions of entities and billions of semantic links. A major challenge is to enable their exploration and querying by end-users. The SPARQL query language is powerful but provides no support for exploration by end-users. Question answering is user-friendly but is limited in expressivity and reliability. Navigation in concept lattices supports exploration but is limited in expressivity and scalability.

In this work [7], we introduce a new exploration and querying paradigm, Abstract Conceptual Navigation (ACN), that merges querying and navigation in order to reconcile expressivity, usability, and scalability. ACN is founded on Formal Concept Analysis (FCA) by defining the navigation space as a concept lattice. We then instantiate the ACN paradigm to knowledge graphs (Graph-ACN) by relying on Graph-FCA, an extension of FCA to knowledge graphs. We continue by detailing how Graph-ACN can be efficiently implemented on top of SPARQL endpoints, and how its expressivity can be increased in a modular way. Finally, we present a concrete implementation available online, Sparklis, and a few application cases on large knowledge graphs.

3.6 Adding Structure and Removing Duplicates in SPARQL Results with Nested Tables

Participants: Sébastien Ferré.

The results of a SPARQL query are generally presented as a table with one row per result, and one column per projected variable. This is an immediate consequence of the formal definition of SPARQL results as a sequence of mappings from variables to RDF terms. However, because of the flat structure of tables, some of the RDF graph structure is lost. This often leads to duplicates in the contents of the table, and difficulties to read and interpret results.

We introduced nested tables to improve the presentation of SPARQL results [11]. A nested table is a table where cells may contain embedded tables instead of RDF terms, and so recursively. We introduce an automated procedure that lifts flat tables into nested tables, based on an analysis of the query. We have implemented the procedure on top of Sparklis, a guided query builder in natural language, in order to further improve the readability of its UI. It can as well be implemented on any SPARQL querying interface as it only depends on the query and its flat results. We illustrated our proposal in the domain of pharmacovigilance, and evaluated it on complex queries over Wikidata.

3.7 Analytical Queries on Vanilla RDF Graphs with a Guided Query Builder Approach

Participants: Sébastien Ferré.

As more and more data are available as RDF graphs, the availability of tools for data analytics beyond semantic search becomes a key issue of the Semantic Web. Previous work require the modelling of data cubes on top of RDF graphs. We have proposed an approach that directly answers analytical queries on unmodified (vanilla) RDF graphs by exploiting the computation features of SPARQL 1.1 [12]. We rely on the NAF design pattern [Fer16a] to design a query builder that completely hides SPARQL behind a verbalization in natural language; and that gives intermediate results and suggestions at each step. Our evaluations have shown that our approach covers a large range of use cases, scales well on large datasets, and is easier to use than writing SPARQL queries.

3.8 Categorial Grammars and NLP

Participants: Annie Foret, Aurélien Lamercerie.

A part of our approach is to consider several classes of categorial grammars and discuss their learnability. We consider learning as a symbolic issue in an unsupervised setting, from raw or from structured data, for some variants of Lambek grammars and of categorial dependency grammars. In that perspective, we discuss for these frameworks different type constructors and structures, some limitations (negative results) but also some algorithms (positive results) under some hypothesis. On the experimental side, we also consider the Logical Information Systems approach, that allows for navigation, querying, updating, and analysis of heterogeneous data collections where data are given (logical) descriptors. Categorial grammars can be seen as a particular case of Logical Information System.

This general approach had been discussed by A. Foret at the 2018 LACompling conference (invited talk), and the post-conference paper [FB19]. The CDG (categorial dependency grammar) case is revisited and presented in details in our recent paper in the Journal of Machine Learning [3]. This is also under experiment on recent linguistic data in the universal dependency format.

The approach has also been studied for the construction of formal representations of natural language texts. The mapping from a natural language to a logical representation is realized with a grammatical formalism, linking the syntactic analysis of the text to a semantic representation.

3.9 Meaning Representation and Semantic Transduction

Participants: Aurélien Lamercerie, Annie Foret.

Semantic representations provide an interesting intermediary between the natural expression of statement and their processing by automatic methods. They allow to formally capture the meaning of texts, making them more accessible for automatic processing. We propose to exploit this kind of representations for documents analysis. Specifically, we provide a methodology to link natural language statements to formal models that can be exploited in a given context. Thus, the use of semantic structures allows the construction of pivot representations. From these representations, we define an analysis process named semantic transduction. The central idea is reflected by a series of transformations on the interpretation of a semantic graph, whose execution is guided using transduction patterns. This technique, which applies to any structure that can be reduced to a labeled graph, thus opening the way for the composition of simple, readable and adaptable processes for interpreting language statements natural.

In particular, we use Abstract Meaning Representation (AMR), supplemented by a transformation process to achieve the expected formal definitions [2]. We target the behavioral aspect of the specifications for cyber-physical systems, i.e. any type of system in which software components interact closely with a physical environment. In this way, the challenge would be to provide assistance to the designer. So, we could simulate and verify, by automatic or assisted methods, "systems" specifications

expressed in natural language. We have proposed a new construction to meet this need, namely Deterministic Propositional Acceptance Automata [2], a formalism with good properties and adapted to integrate into a complete processing chain starting from statement in natural language.

4 Software development

4.1 Software development

4.1.1 GraphMDL Visualizer: Interactive Visualization of Graph Patterns

Participants: Francesco Bariatti, Peggy Cellier, Sébastien Ferré.

Pattern mining algorithms allow to extract structures from data to highlight interesting and useful knowledge. However, those approaches can only be truly helpful if the users can actually understand their outputs. Thus, visualization techniques play a great role in pattern mining, bridging the gap between the algorithms and the users. We have developed GraphMDL Visualizer [BCF20], a tool for the interactive visualization of the graph patterns extracted with GraphMDL, a graph mining approach based on the MDL principle (see 3.2).

GraphMDL Visualizer is structured according to the behavior and needs of users when they analyze GraphMDL results. The tool has different views, ranging from more general (distribution of pattern characteristics), to more specific (visualization of specific patterns). It is also highly interactive, allowing the users to customize the different views, and navigate between them, through simple mouse clicks. GraphMDL Visualizer is freely available online.

4.1.2 Sparklis

Participants: Sébastien Ferré, Pierre Maillot.

Sparklis [Fer17] is a Web user interface that works on top of SPARQL endpoints, i.e. semantic data repositories. It is not tied to a particular endpoint, and works with any endpoint provided that it grants public access. The principle of Sparklis is to let users see and explore data and build expressive queries in natural language at the same time. A SPARQL query is built at the same time but it is only visible at the bottom of the page, for curious expert users. Users don't need to know the data schema, and discover it on the fly. They don't need to write anything, apart from filter values (e.g., matching keywords), which ensures that none of lexical, syntactic, and schema errors are introduced. Sparklis covers a large fragment of SPARQL: graph patterns, optional, union, negation, ordering, aggregation, main filters (string matching, inequalities and intervals, language or datatype). By default, Sparklis connects to DBpedia, a semantic version of the Wikipedia encyclopedia, and several other datasets are available: e.g., Mondial (geographical data), Bretagne tourism (touristic information in Brittany), Wikidata, Nobel prizes.

In 2021, Sparklis has been further improved for the purpose of the PEGASE project: more customization of the displayed elements, and higher efficiency with hierarchies of terms by materializing transitive relations. Two new features were introduced: a navigation history to easily retrieve previously built queries, and the possibility to download the table of results as a CSV file. Beyond PEGASE, to favor the reuse and adaptation of Sparklis, it has been released as a GitHub repository under Apache Licence 2.0. Finally, a JavaScript API is under development to allow for programmatic access and customization of Sparklis' workflow.

4.1.3 Kartu-Verbs

Participants: Mireille Ducassé, Archil Elizbarashvili.

The Georgian language has a complex verbal system, both agglutinative and inflectional, with many exceptions. It is still a controversial issue to determine which lemmas should represent a verb in dictionaries. Verb tables help neophytes to track lemmas starting from inflected forms but if in paper documents they are tedious and error-prone to browse. We propose Kartu-Verbs, a Semantic Web base of inflected Georgian verb forms. For a given verb, all inflected forms are present. Knowledge can easily be traversed in all directions: from Georgian to French and English; from an inflected form to a verbal noun that represent a verb ("masdar"), and conversely from a masdar to any inflected form; from component(s) to forms and from a form to its components. Users can easily retrieve the lemmas that are relevant to access their preferred dictionary. Kartu-Verbs can be seen as a front-end to any Georgian dictionary, thus bypassing the lemmatization issues. An article illustrates in detail how to use Kartu-Verbs [10, 5]. Our base, in its current state, is already a successful proof of concept. It has proven helpful to learn about Georgian verbs. It can be accessed at `https://www-semlis.irisa.fr/software/georgian-verb-inflected-forms-base/`. Collaboration with a researcher from Ivane Javarishvili Tbilisi State University has started last Autumn to enlarge the scope of the tool.

4.1.4 Graph-FCA: Computation and Graphical Display of Concepts

Participants: Sébastien Ferré, Peggy Cellier.

Graph-FCA is a command-line tool for the computation and graphical display of Graph-FCA concepts from knowledge graphs, also known as multi-relational data. Graph-FCA [FC20] is an extension of Formal Concept Analysis (FCA) [GW99] to knowledge graphs where objects are nodes, and attributes are the labels of hyperedges. The intension of a Graph-FCA concept can be seen as a conjunctive query, combining a graph pattern and a projection tuple. The extension of a Graph-FCA concept equals the set of answers of the intension, and the intension is the most specific query for those answers.

The repository of the tool contains the source code, executable programs, a user manual, and examples of inputs and outputs. The inputs are textual files whose syntax

[GW99] B. Ganter, R. Wille, *Formal Concept Analysis — Mathematical Foundations*, Springer, 1999.

is inspired by λProlog [BBR99] and RDF/Turtle, and the outputs are both textual and graphical (DOT and SVG files). Graph-FCA has been applied to genealogical data, descriptions of cooking recipes, environmental and health data, and linguistic data (parse trees). As an alternative to downloading the tool, a web service is available on A||GO.

4.1.5 SQUALL: a Semantic Query and Update High-Level Language

Participants: Sébastien Ferré.

SQUALL (Semantic Query and Update High-Level Language) is a controlled natural language (CNL) for querying and updating RDF graphs [Fer14b]. The main advantage of CNLs is to reconcile the high-level and natural syntax of natural languages, and the precision and lack of ambiguity of formal languages. SQUALL has a strong adequacy with RDF, and covers all constructs of SPARQL, and most constructs of SPARQL 1.1. Its syntax completely abstracts from low-level notions such as bindings and relational algebra. It features disjunction, negation, quantifiers, built-in predicates, aggregations with grouping, and n-ary relations through reification.

SQUALL is available as a Web application at `http://servolis.irisa.fr/squall/` under two forms: one that translates SQUALL sentences to SPARQL, and another one that directly return query answers from a SPARQL endpoint.

4.1.6 NaturaLIS

Participants: Olivier Ridoux, and students.

La Nature is a weekly journal that published vulgarization articles on sciences and techniques from 1873 to 1972, when it was absorbed in La Recherche publishing group. The archives have been poorly numerized in the 1990's and since then made available on the internet. They cover almost a century of scientific and technical development, and in particular developments that shaped the modern world [Smi05,Smi06], like energy and communication networks, modern physics with relativity and quantics, modern cosmology with the big bang, modern geology with plate tectonics, or modern biology with evolution and genetics. Moreover, it is also an indirect witness of geopolitical questions like (de)colonialism, man and its environment, etc.

The NaturaLIS project aims at providing a high-level access to these archives for people who are interested with subjects covered by La Nature, and more specifically with the history of these subjects. This includes post-OCR cleansing, recognition of

[BBR99] C. Belleannee, P. Brisset, O. Ridoux, "A Pragmatic Reconstruction of λProlog", *The Journal of Logic Programming 41*, 1999, p. 67–102.

[Smi05] V. Smil, *Creating the Twentieth Century: Technical Innovations of 1867-1914 and Their Lasting Impact*, Oxford University Press, 2005, `https://books.google.fr/books?id=w3Mh7qQRM-IC`.

[Smi06] V. Smil, *Transforming the Twentieth Century: Technical Innovations and Their Consequences*, Oxford University Press, 2006, `https://books.google.fr/books?id=s\_kRDAAAQBAJ`.

named entities, attribution of keywords to articles, and shaping the whole in an RDF graph. Last step is to present the RDF graph and help navigating into it using the conceptual style developped in the team.

The two most important difficulties are, first the bad quality of the numerization that pollutes every dowstream operations, and second, the diachronic nature of these archives. How to guess that an ancestor of optic fiber was once called a photophone and another a Colladon's fountain? How to guess that ancient Sudan was almost 5000 km West to today Sudan? How to guess that oignon, or *Allium cepa*, was once *Lilium cepa*? Conceptual navigation should help to overcome these difficulties by permitting navigation through association of ideas.

4.1.7 TermLis

Participants: Annie Foret.

TermLis (2015-) is a collection of Logical information contexts for terminological resources (possibly with workflows) as an application of the Logical Information System approach to this field. The current version is to be used with Camelis.

4.1.8 Ares: Abstract Requirement Extraction for Systems

Participants: Aurélien Lamercerie.

Using formal methods to assist system design relies on expected behaviors modeling. The construction of these representations requires extracting behavior rules, called requirements, generally defined in a specification document. ARES (Abstract Requirement Extraction for Systems) meets this need starting from a statement of requirements in natural language. This tool operates an intermediate semantic representation (AMR), and converts it into exploitable formal requirements to model the behaviors of systems.

5 Contracts and collaborations

5.1 National Initiatives

5.1.1 MiKroloG: The Microdata Knowledge Graph

Participants: Sébastien Ferré, Peggy Cellier, Julie Boudebs.

- Project type: CominLabs – Data, AI, Robotics
- Dates: 2021–2024
- PI institution: Univ. Rennes 1
- Other partners: Univ. Nantes (LS2N, team GDD), Institut du Thorax (Nantes)

Searching the web has changed our daily lives, documents on the web containing a list of keywords can be found in a snap. Then, users wanted to find Things, not strings.

Thanks to knowledge graphs (KG), users who request movies of James Cameron receive a list of movies where James Cameron and his movies are Things, i.e. entities defined in the KG. However, searching the web offers diversity with noise, while searching KG just returns exact knowledge. The main issue to bring back diversity from knowledge, i.e. search the web with Things. For instance, we may search the websites selling "James Cameron movies" ordered by price and ratings. This query first retrieves a collection of things, i.e. "James Cameron movies", then asks for "commercial web sites" that refer to these things, i.e. we searched the web with Things. Finding web pages with Things requires a close connection between the web of documents and Knowledge Graphs. Currently, this connection is partially powered by the embedding of microdata in web pages. In MiKroloG, we aim to search the web with Things. To achieve this, we have three main scientific challenges: entity matching to connect web pages and Things, query processing and ranking, and end-user query interface.

At IRISA, we aim to extend Sparklis to work with the entity links of the microdata knowledge graph, while hiding them to end-users. We also aim to add a user interface layer allowing end-users to build queries in a more spontaneous way by entering keywords and phrases, while retaining the interactivity of query construction. The latter objective is the subject of the PhD of Julie Boudebs, who was recruited in September to work on the project.

5.1.2 PEGASE: Improved Pharmacovigilance and Signal Detection with Groupings

Participants: Sébastien Ferré, Annie Foret, Peggy Cellier.

- Project type: ANR
- Dates: 2016–2021
- PI institution: Univ. Rennes 1
- Other partners: LIMICS (INSERM U1142), Regional Centers for Pharmacovigilance in 4 University Hospitals (Besançon, Lille, Paris HEGP, Toulouse), CIC-IT Evalab

The SemLIS team was invited to join the PEGASE project for its Sparklis software, as a way to reconcile the formal aspect of Semantic Web languages, and the need for usability for the end-users, here pharmacovigilance experts.

The mission of those experts is to collect, annotate, store, analyze, and prevent the undesirable effects of drugs. They rely on the MedDRA terminology (Medical Dictionary for Regulatory Activities) to annotate new cases, and to retrieve former cases. An important issue is the large size of MedDRA (about 20,000 terms), and the fact that several terms must generally be used to retrieve all relevant cases from the base. A Semantic Web version of that terminology, the OntoADR ontology, already exists. It allows the precise querying of MedDRA with formal languages like SPARQL. The objective of the project is to develop and compare several user interfaces enabling pharmacovigilance experts to navigate and query the terminology in order to identify the relevant terms.

The leader of the project is Cédric Bousquet from SSPIM (“Service de santé publique et de l’information médicale”) and CHU St Etienne. The project gathers computer scientists from LIMICS (INSERM U1142) and IRISA, pharmacovigilance experts from 4 regional centers (Besançon, Lille, Paris HEGP, Toulouse), and ergonomists in the medical domain from CIC-IT Evalab.

In 2021, several additional improvements were brought to Sparklis (see 4.1.2), and the final user study was prepared (results expected in 2022).

5.1.3 LangNum-br-fr: a DGLF-LF "Langue et numérique" Project

Participants: Annie Foret (coordinator), Karen Kechis (2018-2019), Pierre Morvan (2019-2020), Pierre Martinet (2021).

- Project type: Ministère de la culture, DGLF
- Dates: 2018, 2020
- PI institution: Univ. Rennes 1
- Other partners: Univ. Rennes 2, LIG (Grenoble)

This project (led by Annie Foret) is funded by the "Delegation générale à la langue française et aux langues de france" (DGLF-LF, French culture minister) in the theme "languages and digital" and concerns the French-Breton language pair. The general approach of the scientific project is multidisciplinary, involving computer scientists specialized in natural language processing [Partner A: IRISA and Rennes 1 University, Partner B: LIG Grenoble, Partner C: IT Laboratory in Tours), linguists specialized in Celtic languages [Partner D: CRBC and Rennes2] and specialists in ICT usage [Partner E: Loustic Laboratory]. This work includes technical design work (partners A, B, C in TAL), linguistic work (CRBC) and work on usages (Loustic).

The current challenge is to improve and develop resources and tools for Breton, in coordination between different disciplines, and with a pedagogical concern. A state of the art on tools and resources, and new proposals can be found in our previous contributions. Before defining a software development (a processing chain), an analysis of usages and needs is undertaken with support from a specific Loustic project involving one month engineer.

5.1.4 GTnum : Artificial Intelligence and Education

The proposal described in `https://chaireunescorel.ls2n.fr/2020/10/23/gt-numerique-cest-parti/` for a thematic group “L’impact de l’Intelligence Artificielle à travers l’Éducation Ouverte”, has been accepted by the French "Ministère de l’Éducation Nationale". The project is co-animated by Fahima Djelil (Brest) and Colin de la Higuera.

The SemLIS Team is a participant of this working group.

5.2 Collaborations

- Since the end of 2019, Peggy Cellier is involved in the ADT project SKM in collaboration with Alexandre Termier, Laurent Guillo and Rémi Adon (Engineer on the project since December 2019) about the development of a library of pattern mining tools compatible with the Scikit-learn python library.
- Mireille Ducassé collaborates with Ivane Javakhishvili Tbilisi State University, in Georgia (Caucasus). She is co-supervising the PhD thesis of Archil Elizbarashvili in relation with the Kartu-verbs project (see 4.1.3). She also collaborates with Sarala Birla University, India, on visualization of databases (see 3.3).
- Annie Foret collaborates with LS2N (research lab. Nantes), TALN team (Natural Language Processing), she is a member of "Agence Universitaire de la Francophonie" (AUF), LTT network on "Lexicologie, terminologie et traduction". Annie Foret is member of ATALA (Association pour le Traitement automatique des Langues), and of SIF (Société Informatique de France).

6 Dissemination

6.1 Promoting scientific activities

6.1.1 Scientific Events Organisation

General Chair, Scientific Chair

- Annie Foret acted as co-chair of LACL 2021 Logical Aspects of Computational Linguistics. LACL 2021 was part of the MALIN 2021 (Mathematical Linguistics) event, conjointly with Mathematics of Language, Logic and Algorithms in Computational Linguistics, and Natural Language and Computer Science. The conference was held virtually in december 2021.
- Peggy Cellier is co-Journal track chair of the European Conference in Machine Learning and Knowledge Discovery ECML PKDD 2022 in Grenoble.

6.1.2 Scientific Events Selection

Member of Conference Program Committees

- Sébastien Ferré and Peggy Cellier are members of the Editorial Board of the International Conference on Formal Concept Analysis (ICFCA).
- Peggy Cellier was in the program committee of ECMLPKDD 2021 ("European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases"); EGC 2021 ("Extraction et Gestion de la Connaissances") as senior PC; ICCS 2021 ("International Conferences on Conceptual Structures"); ICFCA 2021("International Conference on Formal Concept Analysis"); RealDataFCA 2021 (workshop at ICFCA); TALN 2021 ("Conférence sur le Traitement Automatique des Langues Naturelles").

- Sébastien Ferré was a member of the program committee of several conferences and workshops:
  - WWW (The Web Conference),
  - ICFCA (Int. Conf. Formal Concept Analysis),
  - CNL (Controlled Natural Languages),
  - FQAS (Flexible Query Answering Systems),
- Annie Foret was a member of the following program committees :
  - ICGI 2020 [postponed, member of ICGI 2021 PC] The International Conference on Grammatical Inference (ICGI) is "the major forum for presentation and discussion of original research papers on all aspects of grammar learning."
  - LACL 2021 was the tenth international conference on Logical Aspect of Computational Linguistics. It was part of the MALIN 2021 (Mathematical Linguistics) event.
- Francesco Bariatti was a member of the review committee for SAC 2022 (37th ACM/SIGAPP Symposium On Applied Computing)

6.1.3 Journal

Reviewer - Reviewing Activities

- Sébastien Ferré made reviews for the following journals:
- SOCO (Soft Computing).
- In 2020-2021, Annie Foret is a reviewer for the Journal of Logic, Language and Information (https://www.springer.com/journal/10849).

6.1.4 Invited Talks

- In December 2021, Annie Foret gave an *invited talk* (see proceedings) with Denis Béchet at the Symposium Logic and Algorithms in Computational Linguistics 2021 (LACompLing2021). The talk was on "Categorial Dependency Grammars: Analysis and Learning". The talk gave an overview on the family of Categorial Dependency Grammars (CDG), as computational grammars for language processing, CDGs are a class of categorial grammars defining dependency structures. They can be viewed as a formal system, where types are attached to words, combining the classical categorial grammars' elimination rules with valency pairing rules defining non projective (discontinuous) dependencies. We discussed both formal aspects of CDG in terms of strength, derivability and complexity and practical issues and algorithms. We also pointed to some open problems. We also reviewed results on CDG learnability and discuss CDG as large scale grammars with respect to Mel'čuk principles and to various hypotheses on training corpora.

6.1.5 Research Administration

- Olivier Ridoux was head of the AI transversal axis of IRISA.
- Sébastien Ferré is a member of the committee of the DKM scientific department (Data and Knowledge Management) at IRISA. Since the end of 2018, Sébastien Ferré is a member of the scientific committee of ABES, the Agency of Libraries in Higher Education, as an expert in Semantic Web technologies.
- Since September 2018, Peggy Cellier is in charge of the Irisa Ph.D. students at IRISA, i.e. she is involved in the "commission du personnel" and organizes the selection of Ph.D. students for ministerial grants (contrats doctoraux). She is also an elected member of the "Conseil de Composante IRISA/INSA" at INSA and an elected member of the "Conseil de laboratoire" at IRISA.

  She served as an external member of the selection committee for Université d'Orléans.
- Hugo Ayats and Francesco Bariatti take part in the organisation of monthly scientific seminars for the DKM department at IRISA and the organisation of the yearly "DKM day".
- Annie Foret is the team correspondant for the (new) GDR TAL.

6.1.6 Other services

- Peggy Cellier is "secrétaire" of "Revue de Traitement automatique des langues"[1] since 2019.
- Mireille Ducassé takes part in the Mentoring program of IRISA as a mentor of a younger colleague.
- Olivier Ridoux is a member of the EcoInfo CNRS service group (GDS) on sustainable development and information technology (aka Green IT). He coauthored with other EcoInfo members an education curricula on Green IT [13].

6.2 Teaching, supervision

6.2.1 Administration

- Since October 2021, Peggy Cellier is responsible of the last year at Computer Science Department at INSA (master 2 level, about 70 students).
- Mireille Ducassé has been the dean of international affairs of INSA Rennes from December 2010 to February 2021. As such, she is was member of the direction of INSA Rennes. She was an active member of the international relations committee of Groupe INSA. She is still "chargée de mission" related to the internationalisation of the Campus and collaborations with Georgia.

[1] https://www.atala.org/revuetal

She supervises an Erasmus+ International credit mobility programs with *Tbilisi State University*, *Akaki Tsereteli State University* of Kutasi and *Georgian technical University* of Tbilisi in Georgia.

- Sébastien Ferré is vice-director of the MIAGE at ISTIC.

  Along with Simon Malinowski, he created a new track on Data Science in the EIT Digital Master School at Univ. Rennes 1. The first master year opened in September 2020, and this year the second master year opened too.

  As of November 2020, he is responsible of the first year of Master Miage, which includes four tracks: classic, alternance, EIT Data Science, and EIT Financial Technologies. They count 75 students in total this year.

- Annie Foret is an elected member of the scientic committee of ISTIC/Rennes 1. She is a member of the IRISA local committee on sustainable development. She was responsible of the internships of computer science students (Master 1 IL and SSR) until september 2018. In 2018-2021 she is responsible with Olivier Ridoux of the second year computer science studies at Rennes 1 university (the group has nearly 200 students).

  She participated in the recruitment committees of external candidates to the L2info level.

- Olivier Ridoux is an elected member of the administration board of ISTIC (CS and Electronic engineering departement of University of Rennes 1). He is co-head, with Annie Foret, for the second year CS studies (bachelor).

6.2.2 Teaching

- Hugo Ayats taught Database (License 2) at INSA Rennes.

- Francesco Bariatti intervened as ATER at ISTIC, teaching *Basics of algorithms in Java* (Licence 1), *Software engineering* (Licence 2), *Structured data and databases* (Licence 2), *Operating Systems* (Licence 3), *Syntax-driven programming* (Master 1), *Semantic Web* (Master 1), and *Symbolic data mining* (Master 2)

- At INSA, Peggy Cellier is responsible of four courses: *Databases and web development* (Licence 3 INFO), *Databases* (Licence 3 Math) *Data Mining* (Licence 3) and *Advanced Database and Semantic Web* (Master 2). She also teaches some other courses: *Database* (Licence 2), *Use and functionalities of an operating system* (Licence 3).

  At master 2 SIF, she teaches in English 4 hours in the data mining course (DMV). In addition she gives a lecture of 2 hours also in master 2 SIF about "Qu'est-ce qu'une thèse, un doctorat, un·e doctorant·e ?".

- Mireille Ducassé, at INSA Rennes, is responsible of two courses, taught in English if international students are present: *Constraint Programming* at Master 1 level, as well as *User-centered Design* at Master 2 level. The latter course, entirely given online, has been open to a group of 30 Indian students from SBU for **virtual exchange**, a première at INSA Rennes.

- Sébastien Ferré teaches Symbolic data mining, Semantic Web, Basics of data analysis with Python, and Compiler techniques at the master level. He also teaches functional programming at license level.
- Annie Foret teaches university courses including formal logic and formal methods for computer scientists, XML technology and related notions and databases at ISTIC and ESIR, Rennes.
- Aurélien Lamercerie teaches compiler techniques at the master level. He also teaches scientific programming and principles of information systems at license level.
- Olivier Ridoux teaches formal language theory and compiler design at ESIR and ISTIC, and logic and constraint programming, operating system, and epistemology at ISTIC. He also teaches Green IT at ESIR and ISTIC, and also to the university staff. He participated in the ISTIC program for high-school teachers on Turing machines and on Green IT.

6.2.3 Supervision

- PhD defended[2] [2] on 08-04-2021 in Rennes 1 **Aurélien Lamercerie**, From texts carrying deontic modalities to their formal representations, started November 2017, supervised by Annie Foret and Benoît Caillaud[3]
- PhD defended[4] [1] on 23-11-2021 in Rennes 1: **Francesco Bariatti**, Mining Tractable Sets of Graph Patterns with the Minimum Description Length Principle, started October 2018, supervised by Sébastien Ferré (50%) and Peggy Cellier (50%)
- PhD in progress: **Hugo Ayats**, From prediction to automation with an explainable and user-centric AI, application to the construction of knowledge graphs from texts, started October 2020, supervised by Sébastien Ferré (50%) and Peggy Cellier (50%)
- PhD in progress: **Julie Boudebs**, An intelligent assistant on top of Sparklis for querying the Semantic Web, started September 2021, supervised by Sébastien Ferré (50%) and Peggy Cellier (50%)
- PhD in progress: **Josie Signe**, Animal welfare : characterizing the diversity between and within livestock farming situations with data mining methods used on information from dairy herd sensors, started September 2020, supervised by Yannick Lecozler (25%), Alexandre Termier (25%), Peggy Cellier (25%) and Véronique Masson (25%)
- PhD in progress: **Archil Elizbarshvili**, Navigation conceptuelle dans les formes conjuguées des verbes géorgiens, started October 2020, supervised by Mireille

[2] https://www.theses.fr/en/2021REN1S029
[3] Team Hycomes - IRISA
[4] `https://www.theses.fr/s210847`

Ducassé (50%), Manana Khachidze - TSU (25%) and Magda Tsintsadze - TSU (25%)

- research internship (M2): **Antoine Cellier**, on "Early detection of customer projects in a retail setting", 5 months, supervised by Peggy Cellier, Tassadit Bouadi and Alexandre Termier

- research internship (M1): **Adrien Paillé**, on "Visualisation de motifs périodiques extraits de traces", 2 months, supervised by Peggy Cellier, Esther Galbrun and Alexandre Termier

- research internship: **Hugo Boulier** from ENS Rennes, on a formal semantics for musical notations, that could be used for the automatic rendition of music, supervised by Olivier Ridoux.

- research internship: **Clément Morand** from ENS Rennes, on the recognition of binominal denominations (aka latin names of genus and species) in the archives of the La Nature journal (see section 4.1.6), supervised by Olivier Ridoux.

- internship (L3 ENS): **Clément Morand**, on "Grammar-LIS: un outil interactif pour la conception de grammaires à partir d'un corpus de phrases", 2 months, supervised by Sébastien Ferré.

- internship (M2, Montpellier): **Pierre Martinet**, on "Contributions à l'enrichissement automatisé de langues peu dotées. Cas du breton et des grammaires formelles.", 6 months, supported by the LangNnum-br-fr project (DGLF-LF), co-supervised by Annie Foret, Denis Béchet (Nantes) and Valérie Bellynck (Grenoble).

- internship: **Ana Tvaliashvili** and **Zurab Kitiashvili** from Georgian Technical University, Tbilisi, Georgia, 5 months, supervised by Mireille Ducassé.

6.2.4 Juries

- Sébastien Ferré served as a referee for the PhD of Nicholas Collis on *"AQUA-COLD: Aggregated Query Understanding and Construction Over Linked Data"*, supervised by Ingo Frommholz, at Univ. Luton, UK, on 06/05/2021. He also served as an examiner for the PhD of Jacques Everwyn on *"Découverte, explicitation et gestion de relations sémantiques par apprentissage profond sur base d'informations variée, volumineuse et de véracité variable"*, supervised by Abdel-Illah MOUADDIB (director), Sylvain GATEPAILLE and Stephan BRUNESSAUX (supervisors), at Univ. Caen, on 14/06/2021.

  In addition, he served in the CSID committee (yearly pre-PhD committee) of Camille Guerry (Univ. Rennes 1).

- Peggy Cellier was a member of the following PhD juries: Cheikh Brahim EL VAIGH, 07/01/2021, Univ. Rennes I (committee member); Francesco Bariatti, 23/11 Rennes (co-supervisor).

She was a member of mid-term evaluation committee of the following PhD candidates: Cyrielle Mallart (Univ. Rennes I), Priscilla Keip (Université de Montpellier), Albeiro Espinal (IMT Atlantique), Grégory Martin (Univ. Rennes I).

- Olivier Ridoux served in the CSID committee of François Mentec (Univ. Rennes 1, IRISA team Druid; PhD defense to come in 2022) and Nicolas Guillaudeux (Univ. Rennes 1, IRISA team Dyliss; PhD defense passed in December 2021).
- Annie Foret is in the CSID committee of Mathilde Régnault, on "Annotation et analyse de corpus hétérogènes", in Paris, within the PROFITEROLE ANR project (PRocessing Old French Instrumented TExts for the Representation Of Language Evolution), supervised by Sophie Prévost at Lattice – ENS.

6.3 Popularization

- In September 2021, Olivier Ridoux participated on behalf of GDS ÉcoInfo in a meeting with URSSAF's regional directors on their digital sustainability policy (URSSAF is the French national organisation that callects and manages employer's social contributions). Similarly, he participated in November in a meeting with Rennes Metropole (Rennes's urban area) on the subject of digital sobriety and time policy.

7 Bibliography

P. Allard, S. Ferré, O. Ridoux, "Discovering Functional Dependencies and Association Rules by Navigating in a Lattice of OLAP Views", *in : Concept Lattices and Their Applications*, M. Kryszkiewicz, S. Obiedkov (editors), CEUR-WS, p. 199–210, 2010, `http://ceur-ws.org/Vol-672/paper18.pdf`.

N. Béchet, P. Cellier, T. Charnois, B. Crémilleux, "Discovering Linguistic Patterns Using Sequence Mining", *in : Int. Conf. on Computational Linguistics and Intelligent Text Processing (CICLing)*, A. F. Gelbukh (editor), *LNCS*, *7181*, Springer, p. 154–165, 2012.

F. Bariatti, P. Cellier, S. Ferré, "GraphMDL Visualizer: Interactive Visualization of Graph Patterns", September 2020, `https://hal.inria.fr/hal-03142207`.

D. Béchet, A. Dikovsky, A. Foret, "Categorial grammars with iterated types form a strict hierarchy of k-valued languages", *Theor. Comput. Sci. 450*, 2012, p. 22–30.

O. Bedel, S. Ferré, O. Ridoux, E. Quesseveur, "GEOLIS: A Logical Information System for Geographical Data", *Revue Internationale de Géomatique 17*, 3-4, 2008, p. 371–390.

D. Béchet, A. Foret, I. Tellier, "Learnability of Pregroup Grammars", *Studia Logica 87*, 2-3, 2007.

P. Cellier, M. Ducassé, S. Ferré, O. Ridoux, "Data Mining for Fault Localization: towards a Global Debugging Process", *Research report*, INSA RENNES

; Univ Rennes, CNRS, IRISA, France, 2018, https://hal.archives-ouvertes.fr/hal-02003069.

P. Cellier, S. Ferré, O. Ridoux, M. Ducassé, "A Parameterized Algorithm to Explore Formal Contexts with a Taxonomy", *Int. J. Foundations of Computer Science (IJFCS) 19*, 2, 2008, p. 319–343.

M. Ducassé, P. Cellier, "Using Bids, Arguments and Preferences in Sensitive Multi-unit Assignments: A p-Equitable Process and a Course Allocation Case Study", *Journal of Group Decision and Negotiation 25*, 6, 2016, p. 1211–1235.

A. Foret, D. Béchet, "On Categorial Grammatical Inference and Logical Information Systems", *in : Logic and Algorithms in Computational Linguistics 2018, Series: Advances in Intelligent Systems and Computing, Studies in Computational Intelligence*, 2019, https://hal.inria.fr/hal-02462675v1/bibtex.

S. Ferré, P. Cellier, "Graph-FCA in Practice", *in : Int. Conf. Conceptual Structures (ICCS) - Graph-Based Representation and Reasoning*, O. Haemmerlé, G. Stapleton, C. Faron-Zucker (editors), *LNCS 9717*, Springer, p. 107–121, 2016, https://hal.inria.fr/hal-01405491.

S. Ferré, P. Cellier, "Graph-FCA: An extension of formal concept analysis to knowledge graphs", *Discrete Applied Mathematics 273*, 2020, p. 81–102, http://www.sciencedirect.com/science/article/pii/S0166218X19301532.

S. Ferré, *Reconciling Expressivity and Usability in Information Access - From Filesystems to the Semantic Web*, Habilitation thesis, Matisse, Univ. Rennes 1, 2014, Habilitation à Diriger des Recherches (HDR), defended on November 6th.

S. Ferré, "SQUALL: The expressiveness of SPARQL 1.1 made available as a controlled natural language", *Data & Knowledge Engineering 94*, 2014, p. 163–188.

S. Ferré, "Bridging the Gap Between Formal Languages and Natural Languages with Zippers", *in : The Semantic Web (ESWC). Latest Advances and New Domains*, H. Sack, E. Blomqvist, M. d'Aquin, C. Ghidini, S. P. Ponzetto, C. Lange (editors), *LNCS 9678*, Springer, p. 269–284, 2016, https://hal.inria.fr/hal-01405488.

S. Ferré, "Semantic Authoring of Ontologies by Exploration and Elimination of Possible Worlds", *in : Int. Conf. Knowledge Engineering and Knowledge Management, LNAI 10024*, Springer, 2016, https://hal.inria.fr/hal-01405502.

S. Ferré, "Sparklis: An Expressive Query Builder for SPARQL Endpoints with Guidance in Natural Language", *Semantic Web: Interoperability, Usability, Applicability 8*, 3, 2017, p. 405–418.

S. Ferré, A. Hermann, "Reconciling faceted search and query languages for the Semantic Web", *Int. J. Metadata, Semantics and Ontologies 7*, 1, 2012, p. 37–54.

S. Ferré, O. Ridoux, "The Use of Associative Concepts in the Incremental Building of a Logical Context", *in : Int. Conf. Conceptual Structures*, G. A. U. Priss, D. Corbett (editor), *LNCS 2393*, Springer, p. 299–313, 2002.

Doctoral dissertations and "Habilitation" theses

[1] F. Bariatti, *Mining Tractable Sets of Graph Patterns with the Minimum Description Length Principle*, PhD Thesis, Université de Rennes 1, November 2021, https://hal.inria.fr/tel-03523742.

[2] A. Lamercerie, *Principe de transduction sémantique pour l'application de théories d'interfaces sur des documents de spécification*, PhD Thesis, Université de Rennes 1, April 2021, Thèse de doctorat dirigée par Caillaud, Benoit et Foret, Annie, http://www.theses.fr/2021REN1S029.

Articles in referred journals and book chapters

[3] D. Béchet, A. Foret, "Incremental learning of iterated dependencies", *Journal of Machine Learning*, 2021, https://doi.org/10.1007/s10994-021-05947-2.

[4] S. B. Dandin, M. Ducassé, "ComVisMD-Compact 2D Visualization of Multidimensional Data: Experimenting with Two Different Datasets", *in : Intelligent Learning for Computer Vision*, H. S. et al. (editor), *Lecture Notes on Data Engineering and Communications Technologies*, *61*, Springer Nature Singapore Pte Ltd, 2021, https://hal.archives-ouvertes.fr/hal-03131685.

[5] M. Ducassé, "Kartu-Verbs : un système d?informations logiques de formes verbales fléchies pour contourner les problèmes de lemmatisation des verbes géorgiens", *Revue des Nouvelles Technologies de l'Information Extraction et Gestion des Connaissances, RNTI-E-38*, 2022, p. 421–428, Démonstration, https://editions-rnti.fr/?inprocid=1002755.

[6] S. Ferré, "Application of Concepts of Neighbours to Knowledge Graph Completion", *Data Science: Methods, Infrastructure, and Applications 4*, 2021, p. 1–28.

[7] S. Ferré, "Conceptual Navigation in Large Knowledge Graphs", *in : Complex Data Analysis with Formal Concept Analysis*, R. Missaoui, L. Kwuida, and T. Abdessalem (editors), Springer, 2021, To appear.

Publications in Conferences and Workshops

[8] H. Ayats, P. Cellier, S. Ferré, "Extracting Relations in Texts with Concepts of Neighbours", *in : Formal Concept Analysis*, A. Braud, A. Buzmakov, T. Hanika, F. L. Ber (editors), *LNCS 12733*, Springer, p. 155–171, 2021, https://doi.org/10.1007/978-3-030-77867-5_10.

[9] F. Bariatti, P. Cellier, S. Ferré, "GraphMDL+: interleaving the generation and MDL-based selection of graph patterns", *in : ACM/SIGAPP Symp. Applied Computing (SAC)*, C. Hung, J. Hong, A. Bechini, E. Song (editors), ACM, p. 355–363, 2021, https://doi.org/10.1145/3412841.3441917.

[10] M. Ducassé, "Kartu-Verbs: A Semantic Web Base of Inflected Georgian Verb Forms to Bypass Georgian Verb Lemmatization Issues", *in : Proceedings of XIX EURALEX Conference*, Z. Gavriilidou (editor), Euralex association, 2020-2021, https://euralex.org/publications/.

[11] S. Ferré, "Adding Structure and Removing Duplicates in SPARQL Results with Nested Tables", *in : Further with Knowledge Graphs*, IOS Press, p. 227–240, 2021.

[12] S. Ferré, "Analytical Queries on Vanilla RDF Graphs with a Guided Query Builder Approach", *in : Flexible Query Answering Systems*, T. Andreasen, G. D. Tré, J. Kacprzyk, H. L. Larsen, G. Bordogna, S. Zadrozny (editors), *LNCS 12871*, Springer, p. 41–53, 2021, https://doi.org/10.1007/978-3-030-86967-0_4.

[13] A.-L. Ligozat, K. Marquet, A. Bugeau, J. Lefevre, P. Boulet, S. Bouveret, P. Marquet, O. Ridoux, O. Michel, "How to Integrate Environmental Challenges in Computing Curricula?", *in : Proceedings of the 53rd ACM Technical Symposium on Computer Science Education (SIGCSE'22)*, ACM, Providence, RI, USA., March 2022, http://recherche.noiraudes.net/resources/papers/SIGCSE22.pdf.