



Activity Report 2019

Team SemLIS

Semantics, Logics, Information Systems
for Data-User Interaction

D7 – Data and Knowledge Management



Abstract. The main objective of the SemLIS team is **to bring back to users the power on their data**. It aims at facilitating data-user interaction by making users more autonomous and agile, by providing flexibility and expressivity, and yet control and confidence in the information system. It should support users in the semantic representation of heterogeneous data, and in the collaborative acquisition of domain knowledge. Its scientific foundations are logics and formal languages for knowledge representation and reasoning, the Semantic Web, information systems, natural language processing, symbolic data mining, and user-data interaction. A key idea is to reconcile the power of formal languages and the usability of natural language and interaction. On the application side, the focus will be put on social sciences and on business intelligence.

Keywords: information systems, knowledge representation, logics, formal languages, natural language processing, data mining, user-data interaction, business intelligence, group decision and negotiation.

1 Team composition

Researchers and faculty

Peggy Cellier, Associate Professor, INSA Rennes

Mireille Ducassé, Professor, INSA Rennes

Sébastien Ferré, Associate Professor (HDR), 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, Birla K.B. Institute of Engineering & Technology, Pilani, India

Research engineers, technical staff

Pierre Maillot, research engineer, SEMELEV project (until May)

PhD students

Francesco Bariatti

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. GOTTLÖB, 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), 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 Categorical 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 Data-driven Assessment of Structural Evolution of RDF Graphs

Participants: Carlos Bobed, Pierre Maillot, Peggy Cellier, Sébastien Ferré.

Since the birth of the Semantic Web, numerous knowledge bases have appeared. The applications that exploit them rely on the quality of their data through time. In this regard, one of the main dimensions of data quality is conformance to the expected usage of the vocabulary. However, the vocabulary usage (i.e., how classes and properties are actually populated) can vary from one base to another. Moreover, through time, such usage can evolve within a base and diverges from the previous practices. Methods have been proposed to follow the evolution of a knowledge base by the observation of the changes of their intentional schema (or ontology); however, they do not capture the evolution of their actual data, which can vary greatly in practice.

We have proposed a data-driven approach to assess the global evolution of vocabulary usage in large RDF graphs [2]. Our proposal relies on two structural measures defined at different granularities (dataset vs update), which are based on pattern mining techniques. We have performed a thorough experimentation which shows that our approach is scalable, and can capture structural evolution through time of both synthetic (LUBM) and real knowledge bases (different snapshots and updates of DBpedia).

3.2 Graph-FCA: An Extension of Formal Concept Analysis to Knowledge Graphs

Participants: Sébastien Ferré, Peggy Cellier.

Knowledge graphs offer a versatile knowledge representation, and have been studied under different forms, such as conceptual graphs or RDF graphs in the Semantic Web. A challenge is to discover conceptual structures in those graphs, in the same way as Formal Concept Analysis (FCA) discovers conceptual structures in tables. FCA has been successful for analysing, mining, learning, and exploring tabular data, and our aim is to help transpose those results to graph-based data. Previous several FCA approaches have already addressed relational data, hence graphs, but with various limits.

We propose Graph-FCA [3] as an extension of FCA where a dataset is a hypergraph instead of a binary table. We show that it can be formalized simply by replacing objects by tuples of objects. This leads to the notion of "n-ary concept", whose extent is an n-ary relation of objects, and whose intent is a "projected graph pattern". We formally reconstruct the fundamental results of FCA for knowledge graphs. We describe in detail the representation of hypergraphs, and the operations on them, as they are much more complex than the sets of attributes that they extend. We also propose an algorithm based on a notion of "pattern basis" to generate and display n-ary concepts in a more efficient and more compact way. We explore a few use cases, in order to study the feasibility and usefulness of Graph-FCA. We consider two use cases: workflow patterns in cooking recipes and linguistic structures from parse trees. In addition, we report on experiments about quantitative aspects of the approach.

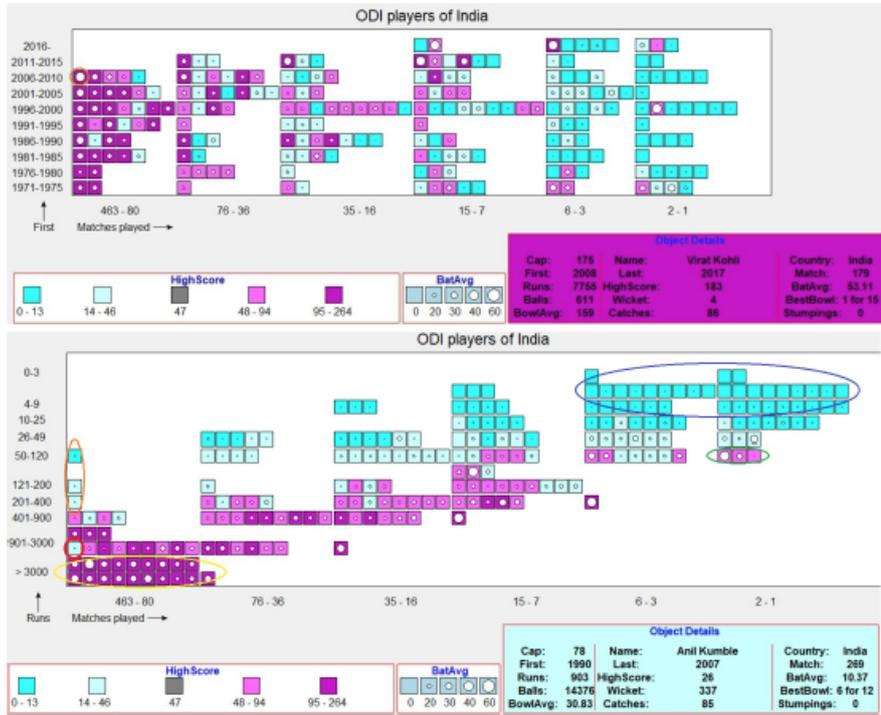


Figure 1: Two Views of ComVisMD : changing displayed dimensions to reveal correlations and exceptions

3.3 Link Prediction in Knowledge Graphs with Concepts of Nearest Neighbours

Participants: Sébastien Ferré.

The open nature of Knowledge Graphs (KG) often implies that they are incomplete. Link prediction consists in inferring new links between the entities of a KG based on existing links. 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, and computation time is difficult to control.

We propose a new approach [6, 7] 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 (CNN) to identify similar entities based on common graph patterns. Dempster-Shafer theory is then used to draw inferences from CNNs. We evaluate our approach on FB15k-237, a challenging benchmark for link prediction, where it gets competitive performance compared to existing approaches.

3.4 Visualization of Databases

Participants: Shridhar B. Dandin, Mireille Ducassé.

Database information is multidimensional and often displayed in tabular format (row/column display). A Choropleth map is a thematic map in which areas are colored according to a variable of interest. They are used mostly for compact graphical representation of geographical information. We propose a system, ComVisMD, inspired by choropleth map, to visualize multidimensional data taking sets of 4 dimensions and projecting them on a compact 2D-display (see Figure 1). The first dimension uses the attribute of main interest to color areas according to a 5-color scale. The next 2 dimensions define the displayed areas as square cells and give the horizontal and vertical axes. The fourth dimension is displayed in the form of varying-size holes in the cells. This activity, done in collaboration with an international partner, has been moving at a slow pace in 2019 due to internal restructuring at the partner institution. We are currently adding a new dimension related to the shape of the cells and investigating the integration of non numerical values.

3.5 An Experimental Approach For Information Extraction in Multi-Party Dialogue Discourse

Participants: Ludovic Jean-Baptiste¹, Peggy Cellier, Thierry Charnois², Bruno Crémilleux³, Albrecht Zimmermann⁴.

Meetings are a large part of our work life. Summarizing a meeting by highlighting its most important points is highly relevant to produce effective meetings. We have addressed the task of identifying which decisions have been taken. This is a challenging task because dialogue in meetings is not well-structured, participants do not always use proper grammar, interrupt each other, etc. Even more challenging, the definition of decisions itself is sometimes unclear in a given part of the dialogue. We propose a pattern mining approach based on sequences to identify text segments related to decisions, which needs far fewer descriptors than the state of the art. We evaluate our method on a benchmark corpus. Based on the results, we discuss the details of decision identification in the used meetings. We tentatively conclude that decision identification, as formulated in the literature, is not a realistic task today.

3.6 Categorical Grammars and NLP

Participants: Annie Foret, Aurélien Lamercerie.

A part of our approach is to consider several classes of categorical 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 categorical 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,

¹Normandie Univ. - GREYC

²Université de Paris 13 - LIPN

³Normandie Univ. - GREYC

⁴Normandie Univ. - GREYC

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. Categorical grammars can be seen as a particular case of Logical Information System.

This general approach has been discussed by A. Foret at the LACompling conference (invited talk), and detailed after in our [4] paper. This is 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. Another approach is to explore the use of an intermediate semantic representation (AMR), supplemented by a transformation process to achieve the expected formal definitions. 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 [10], a formalism with good properties and adapted to integrate into a complete processing chain starting from statement in natural language.

3.7 Legal Domain

Participants: Guillaume Aucher, Annie Foret, Olivier Ridoux.

In 2017 we contributed to two related events on "convergences du droit et du numérique", where we consider information systems in the domain of law and possible applications of the LIS approach in this field. This study is presented in our paper [5] which concerns information systems in law and proposes a new prototype. We discuss the information systems currently used by legal researchers, including their limitations. We present the principles of a new prototype for a better system. This work is accompanied by a first concrete realization, a system with semantic facets, result of our processing chain on a set of decisions of the Constitutional Council.

3.8 AI for Education

Participants: Annie Foret, Mireille Ducassé, Aurélien Lamercerie, Olivier Ridoux.

The ISNLP project outcome was presented at the Cominlab days by A. Foret and A. Lamercerie, within a poster session [9]. See section 5.1.2 for details.

4 Software development

4.1 Software development

4.1.1 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.

From 2018 to 2019, Sparklis has gone through a maturation process (project SEMELEV) with funding from SATT Ouest Valorisation and FEDER. The objective was to make Sparklis an easily reusable JavaScript library. Pierre Maillot was recruited as a research engineer for conducting the development and documentation tasks. The library and its documentation were delivered at the end of May. A new startup, Askelys, has been founded mid 2018 by Romain Thomas and Marc Eluard in order to commercialize services and products on top of Sparklis. This innovation was rewarded in 2018 by the Corporate Foundation of BPO (Banque Populaire de l'Ouest) by a "Project of the Future" prize.

The UI of Sparklis has also been further improved for the purpose of the PEGASE project, following recommendations made by project partners who are human factor specialists (Laura Douze and Romaric Marcilly from CIC-IT / Evalab Lille). Sparklis was also made more efficient on Wikidata, a knowledge graph quickly growing in importance, by relying on an external web service for keyword search.

4.1.2 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 [3] is an extension of Formal Concept Analysis (FCA) [GW99] to knowledge graphs

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

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 is inspired by λ Prolog [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.

On June 25, this tool was presented during the ICFCA'19 conference in the **tutorial** on "*How to deal with relational data in FCA? Two complementary approaches: RCA and Graph-FCA.*"

4.1.3 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.

4.1.4 PEW: Possible World Explorer

Participants: Sébastien Ferré, Sebastian Rudolph.

The **Possible World Explorer** (PEW) [Fer16b] targets ontology designers, and aims to help them correct and complete their ontologies. It reuses the query-based faceted search principles of Sewelis for exploring the “possible worlds” (i.e., models) of an OWL ontology. Users are guided in the incremental construction of class expressions, such that only satisfiable classes are reachable. All classes made of qualified existential restrictions, nominals, intersections, unions, and atomic negations are reachable.

PEW not only supports the exploration of an ontology’s possible worlds, but also supports its completion by the addition of axioms. When a class is found satisfiable, and

[BBR99] C. BELLEANNEE, P. BRISSET, O. RIDOUX, “A Pragmatic Reconstruction of λ Prolog”, *The Journal of Logic Programming* 41, 1999, p. 67–102.

this contradicts domain knowledge (e.g., a man that is not a person), the undesirable possible worlds can be excluded ("pew pew!") by asserting an axiom saying that this class is unsatisfiable (e.g., every man is a person). This could be made a game, where the player would strive to exclude as many undesirable worlds as possible. The benefits are to complete the ontology with more knowledge, and therefore to improve its deduction power.

In addition to completing existing ontologies, PEW also allows the edition of ontologies *de novo*. It allows for the extension of the signature by creating new classes, properties, and individuals; and it covers a wide range of OWL axioms.

4.1.5 Information extraction system based on sequential pattern mining and machine learning

Participants: Ludovic Jean-Baptiste⁵, Peggy Cellier, Thierry Charnois⁶, Bruno Crémilleux⁷, Albrecht Zimmermann⁸.

The tool is a web application developed in Javascript technology in the context of the FUI project REUs. A server, developed with Python and Flask, allows to do the user interface and the implementation of the method. The interface is divided into two parts: the corpus reader, and the model manager. About the corpus reader part, the first step is to load meeting transcripts into the application. No processing of the files has to be done beforehand since the system will clean and prepare the text: removal of disfluencies, removal of contracted forms, extraction of morpho-syntactic categories (pos tags) and identification of named entities. The second step is the application of a model to a new corpus. The other part of the application (model manager) allows you to create (train), read, delete, duplicate, import or export models. The main functionality is to create a new model. A demonstration of the tool is available online⁹.

4.1.6 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.

4.1.7 Kartu-Verbs

Participants: Mireille Ducassé.

The Georgian language has a complex verbal system, both agglutinative and inflectional,

⁵Normandie Univ. - GREYC

⁶Université de Paris 13 - LIPN

⁷Normandie Univ. - GREYC

⁸Normandie Univ. - GREYC

⁹<https://www.youtube.com/watch?v=pfukZtqMUTY>

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.

A report illustrates in detail how to use Kartu-Verbs. It also gives indications about how the base is built in Prolog and integrated in Sparklis [8]. 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-semliis.irisa.fr/software/georgian-verb-inflected-forms-base/>.

5 Contracts and collaborations

5.1 National Initiatives

5.1.1 PEGASE: Improved Pharmacovigilance and Signal Detection with Groupings

Participants: Sébastien Ferré, Annie Foret, Peggy Cellier.

- Project type: ANR
- Dates: 2016–2020
- 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.

At the beginning of this year, the UI of Sparklis was further improved based on the feedback of the ergonomists at CIC-IT. Following this, a first user study was conducted by CIC-IT in front of pharmacovigilance experts in order to compare three search tools, including Sparklis. The results show that Sparklis is powerful, can answer all use cases, and has less limitations than other tools in terms of search capabilities. However, its UI requires more learning than the other tools, although the subjects recognize that this learning could be relatively short. Unfortunately, Sparklis sometimes suffers from long response times due to insufficient support for text search in SPARQL endpoints. The study provided additional recommendations for improvement that should be implemented before conducting a deeper and larger user study.

5.1.2 ISNLP : a Cominlabs Exploratory Project

Participants: Annie Foret (coordinator), Mireille Ducassé, Aurélien Lamerцерie, Olivier Ridoux, Karen Kechis, Maxime Lambert, Kounadi Ouattara.

- Project type: exploratory action
- Dates: 2019
- PI institution: Labex Cominlabs
- Other partners: LINA Nantes, LIG Grenoble, LOUSTIC Rennes 2

The **ISNLP** project (led by Annie Foret) is a short exploratory action in 2019 funded by cominlabs. ISNLP stands for “Information Systems and Natural Language Processing”.

Initial objectives We proposed several experiments to enrich texts (in various formats) that serve as pedagogical support. This comes in two major ways:

- extraction of major parts from full texts. The extraction procedure may be automatic or interactive.
- completion of lacunar texts (such as slides or program comments). Again, automatically or interactively.

Both ways will be provided connections of major terms with referring or enlarged information (terminology, definition, formula, symbols, etc.). The targeted complexity may go from a simple glossary to a rich semantic network and may include interactive technology.

The project involved several teams and departments at IRISA-ISTIC, we provide details for the team outcomes only.

Internships coordinated by SemLIS at IRISA The outcome consists mainly in two proofs of concept (1) (2) and an annotated corpus (3).

- On Natural Language Generation (NLG). A first proof of concept concerns NLG: Kounadi Ouattara (SemLIS internship) has built a prototype going from formulas to texts in both languages French and English. The student was asked to use SimpleNLG. Some alternative tools and approaches have been discussed but left for further developments. Alternatives such as ACG or GF are high-level grammatical frameworks, still topics of research and require time to master. He developed the prototype for two test cases (taught at L2 and L3 level): a fragment of logic and a fragment of XPath.
- On Terminology. Maxime Lambert (SemLIS internship) built a workflow going from texts or pdf to a selected terminology. He investigated several NLP methods and tools, adopting a technique exposed at the ToTh conference. The prototype has been tested on a course at L2 level, it can be run on a terminal or as a web interface. This work is an important part of the project, and prepare the links to definitions and concepts concerning relevant terms or symbols. The student succeeded in dealing with multiword expressions (such as "diagramme de Venn") and symbols (such as set operations) in a pdf file (available at http://etudiant.istic.univ-rennes1.fr/current/l2ie/ofi/Classe_inversee/OFI_Chapitre_Ensembles.pdf).
- On Modelling. Our interest with Matthieu Robert (SemLIS internship), focused on system design, with the aim of assisting the engineer to design a system. The main challenge was to provide elements of study for the evaluation and validation of a processing chain around system design. We have thus defined study cases and started the realization of an annotated corpora to prepare a semantic extraction from specifications.
- On Semantic Elevation. Théo Losekoot and Matthieu Gillet from ENS Rennes are on a research internship with SemLIS on the semantic elevation of the archive of the magazine LA NATURE: about 100 years of a weekly publication from about 1870 to 1970. The aim is to render the result as RDF data on the SparkLIS platform. Possible applications are researchs in history and epistemology of the period, and a first step toward a virtual museum of science.
- On Reversible Computing. Samuel Bouaziz from ENS Rennes is on a research internship with SemLIS on the reversible computing. The aim is to animate reversible computations. Possible applications are research and teaching on reversible computing.

Note. A student from LP USETIC in Rennes 2 worked with us on Celtic languages but without fundings from the ISNLP project (due to her status). She has also been involved in the elaboration of the questionnaire proposed to students on computer science, interacting with LOUSTIC.

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

Participants: Annie Foret (coordinator), Karen Kechis, Pierre Morvan.

- 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 IAT: Vers l'intelligence artificielle territoriale (towards AI for local administration)

The IAT project concerns artificial intelligence applied to territorial administration. IAT stands for "intelligence artificielle territoriale", the team is a participant in this project launched in 2018. The project is described here <https://www-druid.irisa.fr/iat/>: it concerns massive learning for the visualization and prediction of administrative decisions. The planned contribution of the participants of the SemLIS team focuses on knowledge extraction, symbolic learning and natural language processing.

5.2 Collaborations

- Since the end of 2016, Peggy Cellier is involved in the FUI project REUs (started at the end of 2016) in collaboration with the GREYC (Caen) about information extraction from meeting reports. For this collaboration she works with Bruno Crémilleux (Professor at University of Caen), Thierry Charnois (Professor at LIPN), Albrecht Zimmerman (Associate professor at University of Caen) and Ludovic Jean-Baptiste (Engineer on the project since November 2018) about extracting information from meeting transcripts.

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 integration of pattern mining tools into the Scikit-learn python library.

- Mireille Ducassé collaborates with Ivane Javakhishvili Tbilisi State University, in Georgia (Caucasus). An informal collaboration is starting with Tina Margalitzadze from the Lexicography team in relation with the Kartu-verbs project (see 4.1.7).
- Annie Foret collaborates with LINA (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

- Sébastien Ferré was named President of the French-speaking days of Knowledge Engineering (IC) 2020, which is part of PFIA 2020.
- Peggy Cellier was workshop and tutorial chair for the international conference ECML PKDD 2019 (The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases) with Kurt Driessens.

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 a member of the program committee of several conferences:
 - BigFCA workshop,
 - CLA (Int. Conf. on Concept Lattices and Their Applications),
 - ECML PKDD (The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases),
 - PhD Forum of ECML PKDD,
 - EGC démo (Conférence sur l’extraction et la gestion des connaissances),
 - ICCS (Concept Lattices and Applications),
 - ICFCA (Int. Conference on Formal Concept Analysis),
 - IJCAI (International Joint Conference on Artificial Intelligence),

- RECITAL (Conférence des jeunes chercheurs en TAL).

She also served as a "Senior PC" for EGC and TALN.

- Sébastien Ferré was a member of the program committee of several conferences and workshops:
 - WWW (The Web Conference),
 - IJCAI (Int. Joint Conf. Artificial Intelligence),
 - ESWC (Semantic Web),
 - ICFCA (Int. Conf. Formal Concept Analysis),
 - BigFCA (Workshop on Formal Concept Analysis in the Big Data Era),
 - AWD (Workshop on the Web of Data)
- Annie Foret is a member of the following program committees :
 - **FG 2020** (Formal Grammar International Conference). "FG provides a forum for the presentation of new and original research on formal grammar, mathematical linguistics and the application of formal and mathematical methods to the study of natural language."
 - **ICGI 2020** 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."
 - **NLA 2020** Special Session on Natural Language and Argumentation 2020 (NLA'20) at DCAI 2020: International Conference on Distributed Computing and Artificial Intelligence.
 - **CLTW 2019** "The Celtic Language Technology Workshop (CLTW) series of workshops provides a forum for researchers interested in developing NLP (Natural Language Processing) resources and technologies for Celtic languages."
 - Annie Foret made a review for **ECAI 2020** (European Conference on Artificial Intelligence)

6.1.3 Journal

Reviewer - Reviewing Activities

- Sébastien Ferré made reviews for the following journals:
 - JWS (Journal of Web Semantics)
 - IJFS (Journal of Intelligent and Fuzzy Systems)
 - KAIS (Knowledge And Information Systems)
- Annie Foret was a reviewer for

- a book chapter of "Logic and Algorithms in Computational Linguistics 2018 (LACompLing2018)", Volume of Springer series, Studies in Computational Intelligence (SCI)
- the journal Information and Computation (special issue pour LATA 2018, Language and Automata Theory and Applications)

6.1.4 Invited Talks

- Mireille Ducassé visited Ivane Javakhishvili Tbilisi State University in Tbilisi and Akaki Tsereteli University in Kutaisi, Georgia, in November 2019 within the framework of an Erasmus+ International Credit Mobility project. She gave talks about the Kartu-verbs project (see 4.1.7) and she was invited by the French embassy to give a presentation about doctoral studies in France.
- Sébastien Ferré and Annie Foret were invited to give talks at the 100th Ernst-Schröder Kolloquium/Seminar in Darmstadt, on 8-9 March. Sébastien gave talks about "Reconciling Expressivity and Usability in Information Access with Conceptual Navigation" and "SPARKLIS: a versatile vehicle to explore and query knowledge graphs (demo included)". Annie gave a talk about "Formal/Logical Concept Analysis and linguistic information units". The talk [slides](#) are available from the center [site](#).
- Sébastien Ferré was also invited to give a lecture at the EIT Digital summer school on Digital Cities in Rennes, on 2 July. Title: "Semantic Web, Information Retrieval, Data Analysis, Data-User Interaction: Unleashing the Power of Knowledge Graphs for Data Consumers".
- Annie Foret was invited to give a talk at [LACompLing2018](#) (the Symposium on Logic and Algorithms in Computational Linguistics 2018) Stockholm, 28 –31 August 2018, the title was "On Categorial Grammatical Inference and Logical Information Systems". A related post-conference paper is published as a book chapter [4].
- Annie Foret gave a talk on the LangNum-br-fr project challenges and outcome, at a transversal axis "Art, Patrimoine et Culture" day seminar, Rennes, 11-2-2019, on "Comment aider des apprenants d'une langue comme le breton à l'ère du numérique". The talk [slides](#) are available on the axis seminar [site](#).

6.1.5 Research Administration

- Olivier Ridoux was head of the DKM department of IRISA until Summer 2019. He created 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 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 associate professor positions at IUT Blois, Université Paris 13, Ecole des Mines de Nancy and Université de Rennes 2.

- 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. Two one-day meetings took place in 2019: the first one about "Identifiers for Open Science" (March 29); and the second one about "shared library management system" (October 18th).
- Francesco Bariatti takes 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

- Olivier Ridoux is a member of the **EcoInfo** CNRS service group (GDS) on sustainable development and information technology (aka Green IT).

6.2 Teaching, supervision

6.2.1 Administration

- Peggy Cellier organized the bibliographic and internship defense for the Research Master in Computer Science (SIF).

She has also been involved in the IDPE (Ingénieur diplômé par l'état) diploma.

She participated to the recrutement committees of 3rd year students (interviews of candidates).

She also helped the three persons in charge of each year at Computer Science department at INSA (3INFO, 4INFO and 5INFO) in the process of student selection for the options through the use of two tools (Wallet, Whishlis).

- Mireille Ducassé is the dean of international affairs of INSA Rennes since December 2010. As such, she is a member of the direction of INSA Rennes. Until November 2019 she was also the coordinator of the international relations committee of Groupe INSA. She is tightly involved in the working committee regarding international affairs for the constitution of UniR, the forthcoming University of Rennes.

She is, in particular, responsible for exchange programs involving around 400 student mobilities and 30 staff mobilities per year. She set up a number of dual degrees programs over the past years. She supervises an Erasmus+ consortium for Groupe INSA and International credit mobility programs with *Tbilisi State*

University and *Akaki Tsereteli State University* of Kutasi in Georgia ; *Université Euro-Méditerranéenne de Fès* and *Institut National des Postes et Télécommunications* in Morocco ; *Institut de Technologie du Cambodge* in Cambodia, : *Université Cheikh Anta Diop* of Dakar and *Université Gaston Berger* de Saint Louis in Senegal ; as well as *Université Libanaise* in Lebanon. She is directly in charge of the management of the projects with Georgia.

- Sébastien Ferré is vice-director of the MIAGE at ISTIC.
Along with Simon Malinowski, he is in charge of creating a new track on Data Science in the EIT Digital Master School at Univ. Rennes 1. It will open in September 2020, and will share teaching resources with Master Miage.
- Annie Foret is an elected member of the scientific 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 2018-2019 and 2019-2020 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

- At INSA, Peggy Cellier is responsible of four courses: *Databases and web development* (Licence 3 INFO), *Databases* (Licence 3 Math) *Data-Based Knowledge Acquisition: Symbolic Methods* (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).
Peggy Cellier in collaboration with Alexandre Termier and Laurent Guillo managed a group of 7 students in a M1 project about the implementation of pattern mining algorithms based on compression.
- Mireille Ducassé, at INSA Rennes, is responsible of three courses, taught in English if international students are present: *Formal Methods for Software Engineering* (with the “B formal method”) and *Constraint Programming* at Master 1 level, as well as *Participatory Design* at Master 2 level. She gave a 8-hours Prolog course at Master 2 level both at Ivane Javakhishvili Tbilisi State University and Akaki Tsereteli University in Kutaisi in Georgia to prepare future incoming students.
- Sébastien Ferré teaches symbolic data mining, Semantic Web, and compiler techniques at the master level. He also teaches functional programming at license level. This year, he also taught Semantic Web at master 2 level at ENSAI Rennes (15h, 23 students).

- 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.
- Francesco Bariatti this year, at INSA Rennes, taught Database (License 2), Data-Based Knowledge Acquisition (Licence 3 INFO), Operative systems and parallel programming (Master 1) and Semantic Web (Master 2). At ISTIC, he taught Database (License 2) and Web HTML/CSS/JS (to secondary-school teachers).
- Olivier Ridoux teaches formal language theory, compiler design, and innovation at ESIR, and compiler techniques, logic and constraint programming, operating system, and epistemology at ISTIC. He participated in the ISTIC program for high-school teachers.

6.2.3 Supervision

- PhD in progress: **Francesco Bariatti**, Semantic Lifting of Complex Data by Hypergraph Compression, started October 2018, supervised by Sébastien Ferré (50%) and Peggy Cellier (50%)
- PhD in progress: **Aurélien Lamercerie**, From texts carrying deontic modalities to their formal representations, started November 2017, supervised by Annie Foret and Benoît Caillaud¹⁰
- PhD: **Anh-Duong Nguyen**, Compression Based Pattern Mining, started March 2018 and ended March 2019, supervised by Peggy Cellier (25%), Alexandre Termier (25%)¹¹, Romaric Gaudel (50%)¹²
- internship (L3): **Nicolas Fouqué**, on "Approximate Evaluation of SPARQL queries in JENA, and application to concepts of neighbors", 2 months, supervised by Sébastien Ferré and Peggy Cellier.
- internship (L3): **Naima Nazari**, on "Visualization of patterns", 3 months, supervised by Alexandre Termier and Peggy Cellier.
- internship (M1): **Sophie Bouvry**, on "A natural language interface to query the Semantic Web", 3 months, supervised by Sébastien Ferré and Peggy Cellier.
- internship (M1): **Josie Signe**, on "Programming by demonstration with pattern mining", 2.5 months, supervised by Alexandre Termier and Peggy Cellier.
- research internship (M1 SIF): **Théo Losekoot, Matthieu Gillet and Samuel Bouaziz**, 6 months, supervised by Olivier Ridoux.

¹⁰Team Hycomes - IRISA

¹¹Team Lacodam - IRISA

¹²ENSAI

- internship (LP USETIC, Rennes 2): **Karen Kechis**, on "languages and digital" projects to help learners, 12 weeks in 2018-2019, supervised by Annie Foret.
- internship (LP USETIC, Rennes 2): **Pierre Morvan**, on "languages and digital" projects to help learners, 12 weeks in 2019-2020, supervised by Annie Foret.
- internship (L3info): **Kounadi Ouattara**, 3 months, within the **ISNLP** project on AIED, on "Natural Language Generation for pedagogical documents enrichments", supervised by Annie Foret.
- internship (M1): **Maxime Lambert**, 2,5 months, within the **ISNLP** project on AIED, on "Terminology extraction and linking for pedagogical documents enrichments", supervised by Annie Foret.
- internship (ESIR1): **Matthieu Robert**, 2 months, within the **ISNLP** project on AIED, on "the modeling of specifications (in natural language) to controlled language / automata", supervised by Aurélien Lamerçerie.
- internship: **Keti Meipariani, Mariam Asatiani, and Mikheil Maisuradze** from Tbilisi State University, as well as **Ana Idadze** from Akaki Tsereteli University in Kutaisi, Georgia, on the Kartu-Verbs project (see 4.1.7), 5 months, supervised by Mireille Ducassé.

6.2.4 Juries

- Sébastien Ferré served as examiner in the PhD committee of Lucas Bourneuf on "A search space of graph motifs for graph compression: from Powergraphs to triplet concepts", Jacques Nicolas supervisor, University of Rennes 1, 17-12-2019.
He also served in the "CSID" of Camille Guerry (Univ. Rennes 1) and Priscilla Keip (IMT Mines-Alès).
- Peggy Cellier served as examiner in the PhD committee of Aimene Belfodil on "An Order Theoretic Point-of-view on Subgroup Discovery", Céline Robardet and Mehdi Kaytoue supervisors, INSA Lyon, 30-09-2019.
Peggy Cellier served as examiner in the PhD committee of Nyoman Juniarta on "Fouille de données complexes et biclustering avec l'analyse formelle de concepts", Amedeo Napoli and Miguel Courceiro supervisors, University of Lorraine, 18-12-2019.
She also served in the "CSID" of Cheikh Brahim (Univ. Rennes 1), Clément Dalloux (Univ. Rennes 1), Cyrielle Mallart (Univ. Rennes 1) and Erwan Bourrand (Univ. Rennes 1).
- Olivier Ridoux served as president in the PhD committee of Manuel Leduc on "Efficient external domain-specific software language engineering", Olivier Barais and Benoît Combemale (IRIT, Toulouse) supervisors, University of Rennes 1, 18-12-2019.
He also served in the CSID committee of Nicolas Guillaudeux (Univ. Rennes 1).

- Annie Foret served as *reviewer (rapporteur)* and as jury member for the *habilitation (HDR)* of Mathieu MANGEOT, 5-12-2019, in Chambéry, on "Génie lexicosémantique multilingue contributif".
- Annie Foret is in the pre-PhD committee (CSI) of Hugo Talibart, in bioinformatics, on "Learning grammars with long-distance correlations on proteins", supervised by J. Nicolas and F. Costes at IRISA-Rennes.
- Annie Foret is in the pre-PhD committee (CSI) 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 March 2019, within the framework of "A la découverte de la recherche", supervised par Rennes Rectorate, Mireille Ducassé gave two presentations in high schools about "Logical Information Systems: Artificial Intelligence to Leverage Natural Intelligence". In October 2019, she made an intervention to the "Girls'R Coding" event organized by ADN Bretagne.
- In October 2019, Peggy Cellier was a member of the jury for the "Science en court" festival¹³.
- As a member of GDS ÉcoInfo, Olivier Ridoux participated in the **360 Possibles** event organized by Région Bretagne on June 12-14, 2019, as an invited speaker on Green IT. He participated in workshops with **Incubateur du barreau de Rennes** and with **Cercle des Jeunes Entrepreneurs**. He also collaborated in the writing of an article for the journal *Alternatives Économiques* (« Les métaux rares mettent le monde sous tension »).

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.

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.

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.

¹³<http://sciences-en-courts.fr>

- 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 Categorical 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, *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.

Books and Monographs

[1]

Articles in referred journals and book chapters

- [2] C. BOBED, P. MAILLOT, P. CELLIER, S. FERRÉ, “Data-driven Assessment of Structural Evolution of RDF Graphs”, *Semantic Web: Interoperability, Usability, Applicability*, 2019, To appear, <http://www.semantic-web-journal.net/content/data-driven-assessment-structural-evolution-rdf-graphs-0>.
- [3] S. FERRÉ, P. CELLIER, “Graph-FCA: An extension of formal concept analysis to knowledge graphs”, *Discrete Applied Mathematics*, 2019, <http://www.sciencedirect.com/science/article/pii/S0166218X19301532>.
- [4] 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>.

Publications in Conferences and Workshops

- [5] A. CHAUVET, A. FORET, “Systèmes d’information pour les chercheurs en droit”, in : *Colloque des Convergences du Droit et du Numérique, Actes du colloque des Convergences du droit et du numérique*, Université de Bordeaux, Bordeaux, France, July 2019, <https://hal.inria.fr/hal-02195921>.
- [6] S. FERRÉ, “Link Prediction in Knowledge Graphs with Concepts of Nearest Neighbours”, in : *The Semantic Web (ESWC)*, P. Hitzler, M. Fernández, K. Janowicz, A. Zaveri, A. J. G. Gray, V. López, A. Haller, K. Hammar (editors), *LNCS 11503*, Springer, p. 84–100, 2019, https://doi.org/10.1007/978-3-030-21348-0_6.
- [7] S. FERRÉ, “Prédiction de liens dans les graphes de connaissances avec les concepts de plus proches voisins”, in : *Extraction et Gestion des connaissances (EGC)*, M. Rousset, L. Boudjeloud-Assala (editors), *RNTI, E-35*, Hermann-Éditions, p. 69–80, 2019.

Internal Reports

- [8] M. DUCASSÉ, “Kartu-Verbs: A Semantic Web Base of Inflected Verb Forms for the Georgian Language”, *Technical report*, IRISA-INSA, December 2019, Available under HAL <https://hal.archives-ouvertes.fr/hal-02460975>.

Miscellaneous

- [9] A. FORET, “ISNLP on AI for Education - Exploratory action 2019 : Exploratory experiments combining Information Systems and Natural Language Processing”, Cominlabs days, November 2019, Poster, <https://hal.inria.fr/hal-02462991>.

- [10] A. LAMERCERIE, “Une algèbre des automates d’acceptation propositionnelle déterministes comme théorie d’interface pour la conception de systèmes cyberphysiques”, MSR 2019 - 12ème Colloque sur la Modélisation des Systèmes Réactifs, Nov 2019, Angers, France, November 2019, Poster, <https://hal.archives-ouvertes.fr/hal-02432696>.