Activity Report 2018

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.

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

> Carlos Bobed, research engineer, PEGASE project (until July)
> Pierre Maillot, research engineer, SEMELEV project

PhD students

> Francesco Bariatti (since October)
> Clément Gautrais, co-supervised with team LACODAM
> Aurélien Lamercerie, co-supervised with team HYCOMES
> Anh-Duong Nguyen, co-supervised with team LACODAM

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 [18]. 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), 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]. 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 Measure of the Structural Similarity between RDF Graphs

Participants: Pierre Maillot, Carlos Bobed-Lisbona.

In the latest years, there has been a huge effort to deploy large amounts of data, making it available in the form of RDF data thanks, among others, to the Linked Data initiative. In this context, using shared ontologies has been crucial to gain interoperability, and to be able to integrate and exploit third party datasets. However, using the same ontology does not suffice to successfully query or integrate external data within your own dataset: the actual usage of the vocabulary (e.g., which concepts have instances, which properties are actually populated and how) is crucial for these tasks. Being able to compare different RDF graphs at the actual usage level would indeed help in such situations. Unfortunately, the complexity of graph comparison is an obstacle to the scalability of many approaches.

We present a structural similarity measure in [16], designed to compare structural similarity of low-level data between two different RDF graphs according to the patterns they share. To obtain such patterns, we leverage a data mining method (KRIMP) which allows to extract the most descriptive patterns appearing in a transactional database. We adapt this method to the particularities of RDF data, proposing two different conversions for an RDF graph. Once we have the descriptive patterns, we evaluate how much two graphs can compress each other to give a numerical measure depending on the common data structures they share. We have carried out several experiments to show its ability to capture the structural differences of actual vocabulary usage.

3.2 Sparklis over PEGASE Knowledge Graph: a New Tool for Pharmacovigilance

Participants: Carlos Bobed, Sébastien Ferré, Laura Douze and Romaric Marcilly[1].

Pharmacovigilance is in charge of studying the adverse effects of pharmaceutical products. In this field, pharmacovigilance specialists experience several difficulties when searching and exploring their patient data despite the existence of standardized terminologies (MedDRA). We have presented our approach [6, 5] to enhance the way pharmacovigilance specialists perform search and exploration on their data. First, we have developed a knowledge graph that relies on the OntoADR ontology to semantically enrich the MedDRA terminology with SNOMED CT concepts, and that includes anonymized patient data from FAERS. Second, we have chosen and extended our semantic search tool, Sparklis, according to the user requirements that we have identified in pharmacovigilance. We report the results of a usability evaluation that has been performed by human factors specialists to check the benefits of our proposal.

---

[1]Univ. Lille, INSERM, CHU Lille, CIC-IT / Evalab 1403

3.3   How Hierarchies of Concept Graphs Can Facilitate the Interpretation of RCA Lattices?

Participants: Sébastien Ferré, Peggy Cellier.

Relational Concept Analysis (RCA) has been introduced in order to allow concept analysis on multi-relational data. It significantly widens the field of application of Formal Concept Analysis (FCA), and it produces richer concept intents that are similar to concept definitions in Description Logics (DL). However, reading and interpreting RCA concept lattices is notoriously difficult. Nica *et al* have proposed to represent RCA intents by cpo-patterns in the special case of sequence structures. We propose an equivalent representation of a family of RCA concept lattices in the form of a hierarchy of concept graphs [8]. Each concept belongs to one concept graph, and each concept graph exhibits the relationships between several concepts. A concept graph is generally transversal to several lattices, and therefore highlights the relationships between different types of objects. We show the benefits of our approach on several use cases from the RCA litterature.

3.4   Answers Partitioning and Lazy Joins for Efficient Query Relaxation and Application to Similarity Search

Participants: Sébastien Ferré.

Query relaxation has been studied as a way to find approximate answers when user queries are too specific or do not align well with the data schema. We are here interested in the application of query relaxation to similarity search of RDF nodes based on their description. However, this is challenging because existing approaches have a complexity that grows in a combinatorial way with the size of the query and the number of relaxation steps. We have introduced two algorithms, answers partitioning and lazy join [9], that together significantly improve the efficiency of query relaxation. Our experiments show that our approach scales much better with the size of queries and the number of relaxation steps, to the point where it becomes possible to relax large node descriptions in order to find similar nodes. Moreover, the relaxed descriptions provide explanations for their semantic similarity.

3.5   Recent Extensions to Sparklis

Participants: Sébastien Ferré.

Sparklis is a SPARQL query builder, developed in the team since 2013, that can connect to any endpoint, and that interacts with users in natural language only. Users are guided in the building of their queries so that they do not have to know the schema, and so that empty results are almost completely avoided. We have presented a demo [11] showing a number of recent extensions to Sparklis. Most notably, it now supports analytical queries, Wikidata statement qualifiers, and the display of results on a map or as a slideshow.

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

We have detailed the support for analytical queries in a technical report [20]. Indeed, 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 has focused on adapting OLAP-like approaches and question answering by modelling RDF data cubes on top of RDF graphs. We have proposed a more direct – and more expressive – approach by guiding users in the incremental building of SPARQL 1.1 queries that combine several computation features (aggregations, expressions, bindings and filters), and by evaluating those queries on unmodified (vanilla) RDF graphs. We rely on the N<A>F design pattern [Fer16a] to hide SPARQL behind a natural language interface, and to provide results and suggestions at every step. We have implemented our approach on top of SPARKLIS, and we have reported on three experiments to assess its expressivity, usability, and scalability.

3.6   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. A poster, published in the IEEE VIS conference [7], illustrates our approach on cricket players data and show how ComVisMDs compact visualization can help analyze data and find correlations as well as explain the exceptions by the way of intuitive color observation, shape of the cells, information on cell, dynamic scaling, classification and clustering.

3.7   Mining Periodic Patterns with a MDL Criterion

Participants: Esther Galbrun[2], Peggy Cellier, Nikolaj Tatti[3], Alexandre Termier[4], Bruno Crémilleux[5].

The quantity of event logs available is increasing rapidly, be they produced by industrial processes, computing systems, or life tracking, for instance. It is thus important to design effective ways to uncover the information they contain. Because event logs often record repetitive phenomena, mining periodic patterns is especially relevant when considering such data. Indeed, capturing such regularities is instrumental in providing condensed representations of the event sequences.

   We propose an approach for mining periodic patterns from event logs while relying on a Minimum Description Length (MDL) criterion to evaluate candidate patterns [13, 21]. Our goal is to extract a set of patterns that suitably characterises the periodic structure present in the data. We evaluate the interest of our approach on several real-world event log datasets.

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

Participants: Pegah Alizadeh[6], Peggy Cellier, Thierry Charnois[7], Bruno Crémilleux[8], Albrecht Zimmermann[9].

We have addressed the task of information extraction for transcript of meetings [4, 3]. Meeting documents are not usually well structured and are lacking formatting and punctuations. In addition, the information are distributed over multiple sentences. We experimentally investigate the usefulness of numerical statistics and topic modeling methods on a real data set containing multi-part dialogue texts. Such information extraction can be used for different tasks, of which we consider two: contrasting *themat-*

---

[2]Department of Computer Science, Aalto University, Finland
[3]Department of Computer Science, Aalto University, Finland
[4]Team Lacodam - IRISA
[5]Normandie Univ. - GREYC
[6]Normandie Univ. - GREYC
[7]Université de Paris 13 - LIPN
[8]Normandie Univ. - GREYC
[9]Normandie Univ. - GREYC

ically related but distinct meetings from each other, and contrasting meetings involving *the same participants* from those involving other. In addition to demonstrating the difference between counting and topic modeling results, we also evaluate our experiments with respect to the gold standards provided for the data set.

3.9   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 has been discussed by A. Foret at the LACompling conference (invited talk) and at UPC Barcelone (workhop) with a focus on the dependency grammar variant. 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. 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. Some existing contributions that could enable progress on this issue have been studied [14].

3.10   Linguistic Data and Logical Information Systems

Participants: Annie Foret.

We had recently undertaken the Logical Concept Analysis (LCA) that extends the Formal Concept Analysis (FCA) for terminology, a workflow had been developed to go from XML data to a logical information context. Through experiments on specific resources, facet designs have been tuned to facilitate the search and control on the data. With this prototype, we consider several usages of such contexts and illustrate benefits of the approach. This was done initially using the FranceTerme resource, containing terms of different scientific and technical fields; we have carried on this development considering the EuroVoc resource, a Multilingual Thesaurus of the European Union. In these data, each concept is related to a domain, with an attempt to avoid ambiguities.

We have proposed a prototype version for Breton, with a connection to a semantic graph (wordnet, babelnet). Those works have been reconsidered in the LangNum-br-fr study as a possible contribution to a new tool to help language learners [12].

3.11   Responsive and Flexible Controlled Natural Language Authoring with Zipper-Based Transformations

Participants: Sébastien Ferré.

Controlled natural languages (CNL) have the benefits to combine the readability of natural languages, and the accuracy of formal languages. They have been used to help users express facts, rules or queries. While generally easy to read, CNLs remain difficult to write because of the constrained syntax. A common solution is a grammar-based auto-completion mechanism to suggest the next possible words in a sentence. However, this solution has two limitations: (a) partial sentences may have no semantics, which prevents giving intermediate results or feedback, and (b) the suggestion is often limited to adding words at the end of the sentence. We have proposed a more responsive and flexible CNL authoring [10] by designing it as a sequence of sentence transformations. Responsiveness is obtained by having a complete, and hence interpretable, sentence at each time. Flexibility is obtained by allowing insertion and deletion on any part of the sentence. Technically, this is realized by working directly on the abstract syntax, rather than on the concrete syntax, and by using Huet's zippers to manage the focus on a query part, the equivalent of the text cursor of a word processor.

   This work has received the Best Paper Award at the International Workshop on Controlled Natural Languages (CNL).

3.12   Legal Domain

Participants: Guillaume Aucher, Annie Foret, Olivier Ridoux.

   We described [17] the theoretical principles that underlie the design of a software tool which could be used by judges for writing judgements and for making decisions about litigations. The tool is based on Binary Decision Diagrams (BDD), which are graphical representations of truth-valued functions associated to propositional formulas. Given a specific litigation, the tool asks questions to the judge; each question is represented by a propositional atom. Their answers, true or false, allow to evaluate the truth value of the formula which encodes the overall recommendation of the software about the litigation. Our approach combines some sort of 'theoretical' or 'legal' reasoning dealing with the core of the litigation itself together with some sort of 'procedural' reasoning dealing with the protocol that has to be followed by the judge during the trial: some questions or group of questions must necessarily be examined and sometimes in a specific order. That is why we consider extensions of BDDs called Multi-BDDs. They are BDDs with multiple entries corresponding to the different specific issues that must necessarily be addressed by the judge during the trial. We illustrate our ideas on a case study dealing with French union trade elections, an example that has been used throughout a project with the French Cour de cassation. We have also sketched the architecture of

a prototype software that has been developed during this project.

3.13   Scalable Active Constrained Clustering for Temporal Data

Participants: Anh-Duong Nguyen.

In this paper, we introduce a novel interactive framework to handle both instance-level and temporal smoothness constraints for clustering large temporal data. It consists of a constrained clustering algorithm, called CVQE+ [15], which optimizes the clustering quality, constraint violation and the historical cost between consecutive data snapshots. At the center of our framework is a simple yet effective active learning technique, named Border, for iteratively selecting the most informative pairs of objects to query users about, and updating the clustering with new constraints. Those constraints are then propagated inside each data snapshot and between snapshots via two schemes, called constraint inheritance and constraint propagation, to further enhance the results. Experiments show better or comparable clustering results than state-of-the-art techniques as well as high scalability for large datasets.

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.

   In 2018, Sparklis has gone through a maturation process (project SEMELEV) with funding from SATT Ouest Valorisation and FEDER. The objective is to make Sparklis an easily reusable JavaScript library. Pierre Maillot has been recruited as a research engineer for conducting the development and documentation tasks. A new startup, Askelys, has been founded by Romain Thomas and Marc Eluard in order to commercialize services and products on top of Sparklis. This innovation was rewarded 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 improved for the purpose of the PEGASE project, starting from recommendations made by project partners who are human factor specialists (Laura Douze and Romaric Marcilly from CIC-IT / Evalab Lille).

4.1.2 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.

4.1.3 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 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.

4.1.4 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.

5 Contracts and collaborations

5.1 National Initiatives

5.1.1 PEGASE: Improved Pharmacovigilance and Signal Detection with Groupings

Participants: Sébastien Ferré, Carlos Bobed-Lisbona, Annie Foret, Peggy Cellier, Mireille Ducassé.

- 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.

This year, the PEGASE knowledge based was improved and finalised, and Sparklis' UI was further improved from the feedback of the ergonomists at CIC-IT. Those results

were presented at two conferences, EKAW [5] and SWAT4HCLS [6]. The next step is to conduct user experiments with pharmacovigilance experts.

5.1.2   IDFRAud: An Operational Automatic Framework for Identity Document Fraud Detection and Profiling

Participants: Sébastien Ferré, Peggy Cellier, Pierre Maillot.

- Project type: ANR
- Dates: 2015–2018
- PI institution: Univ. Rennes 1
- Other partners: AriadNEXT, IRCGN ("Institut de Recherche Criminelle de la Gendarmerie Nationale"), and ENSP (National School of Police)

The ANR IDFRAud aims at allowing the automated recognition of ID documents, and the detection of false documents, by applying techniques for document analysis, classification, and knowledge management. The leader is Montaser Awal (replacing Abdullah Almaksour in 2016) from the AriadNEXT innovating company, and other partners are the IRISA laboratory, IRCGN ("Institut de Recherche Criminelle de la Gendarmerie Nationale"), and ENSP (National School of Police). Sébastien Ferré is the scientific head for the IRISA partner. The project started in February 2015. For the SemLIS team, the project funded 2 postdoc years and research costs (e.g., missions, machines).

This year was the end of the IDFRAud project. It was concluded by a colloquium in Lyon on 22-23 March where project results were presented in front of various stakeholders in the domain of security related to ID documents.

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

Participants: Annie Foret.

- Project type: Ministère de la culture, DGLF
- Dates: 2018
- 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.

5.2   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 Unvisersity of Caen), Thierry Charnois (Professor at LIPN), Albrecht Zimmerman (Associate professor at University of Caen), Pegah Alizadeh (Post-doc on the project) and Ludovic Jean-Baptiste (Ingineer on the project since November) about extracting information from meeting transcripts.

  Since the end of 2016, she also collaborates with Esther Galbrun (Researcher at Aalto University, Finland) and Alexandre Termier about analysing traces.

- Mireille Ducassé collaborates with Ivane Javakhishvili Tbilisi State University, in Georgia (Caucasus). A proposal has been made to the Georgian National Science Foundation regarding "A Georgian Language Based Semantic Search Engine Algorithm Development" by Manana Khachidze and Magda Tsintsadze with a contribution of SemLIS. Results are pending. Magda Tsintsadze visited SemLIS from May 14th to May 18th. Mireille Ducassé also collaborates with the B. K. Birla Institute of Engineering & Technology, in India. Shridhar Dandin visited SemLIS from May 18th to June 5th

- 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).

- Sébastien Ferré collaborates with Persée, a research unit in the domain of humanities and social sciences, on the use of Sparklis (see 4.1.1).

6   Dissemination

6.1   Promoting scientific activities

6.1.1   Scientific Events Organisation

General Chair, Scientific Chair

- Annie Foret was co-chair (member of the steering committee) of the international conference FG (Formal Grammar) [1].

Member of the Organizing Committees

- Peggy Cellier and Annie Foret were member of the organization committee of TALN 2018, which held in Rennes in May 2018. The president of the committee was Vincent Claveau.

6.1.2 Scientific Events Selection

Chair of Conference Program Committees

- Peggy Cellier was, with Anne-Laure Ligozat, program chair for RJC 2018 (Rencontres Jeunes Chercheur·euse·s) which groups together two conferences (RECITAL and RJCRI).

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:

  - 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),
  - EGC démo (Conférence internationale sur l'extraction et la gestion des connaissances),
  - EKAW (International Conference on Knowledge Engineering and Knowledge Management),
  - ICCS (Concept Lattices and Applications),
  - IJCAI-ECAI (International Joint Conference on Artificial Intelligence and the European Conference on Artificial Intelligence).

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

- Sébastien Ferré was a member of the program committee of several conferences and workshops:

  - WWW (The Web Conference),
  - ISWC (Int. Semantic Web Conf.),
  - CLA (Concept Lattices and their Applications),
  - CNL (Controlled Natural Languages),
  - HAICTA post-proceedings (International Conference on Information and Communication Technologies in Agriculture, Food & Environment)

- Pierre Maillot was a program comittee member for the ACM Symposium on Applied Computing 2018, Semantic Web and Applications (SWA) track.

- Annie Foret was a member of the following program commitees :

  - FG (Formal Grammar)
  - LACompLing (Logic and Algorithms in Computational Linguistics)
  - CoNLL (The SIGNLL Conference on Computational Natural Language Learning)
  - RJC (Rencontre Jeunes Chercheurs correspondant à la 13e édition de la Rencontre des Jeunes Chercheurs en Recherche d'Information (RJCRI) et à la 20e édition des Rencontre Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL))

6.1.3 Journal

Reviewer - Reviewing Activities

- Peggy Cellier made reviews for the DAM journal (Discrete Applied Mathematics).

- Sébastien Ferré made reviews for the following journals:

  - KAIS (Knowledge And Information Systems)
  - RIA (Revue d'Intelligence Artificielle)

- Annie Foret was a reviewer in 2018 for the Journal of Logic, Language and Information; this is the official publication of the Association for Logic, Language, and Information.

6.1.4 Invited Talks

- Sébastien Ferré was invited to give a one-day tutorial about Sparklis in front of researchers in social sciences and humanities (Le Mans, 10 March).

- 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".

  Annie Foret was invited to give a talk at a workshop on the occasion of Carlos Cardó PhD defense in UPC, Barcelona, the title was: "Categorial Dependency Grammars with Iterated Sequences".

- Mireille Ducassé visited Ivane Javakhishvili Tbilisi State University, Georgia, in February, August and November 2018 within the framework of an Erasmus+ International Credit Mobility project. She gave a talk about a preliminary project about Georgian verb conjugation.

- Peggy Cellier was invited to give a talk at "Séminaire au vert" by team Lacodam at Frehel, 19 –21 June 2018, the title was "Pattern Mining for NLP".

6.1.5 Research Administration

- Olivier Ridoux is the head of the DKM department since October 2014. He is a member of the EcoInfo CNRS service group on sustainable development and information technology. As such he participated in the Lean ICT working group of the Shift Project think tank. The working group published the prospective and recommendations report "Pour une sobriété numérique" [19].

- Olivier Ridoux and Sébastien Ferré are members 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 an associate professor position at the University Jean Monnet at Saint-Etienne (ref. MCF0511).

6.2 Teaching, supervision

6.2.1 Administration

- Mireille Ducassé is the dean of of international affairs of the INSA Rennes since December 2010. As such, she is a member of the direction of the INSA Rennes. Since March 2014 she is 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, University of Rennes, the forthcoming federation of Higher-Education institutions 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 Morrocco ; *Institut de Technologie du Cambodge* in Cambodgia, : *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.

- Olivier Ridoux is an elected member of the administration board of ESIR (École supérieure d'ingénieurs de Rennes), and responsible for the Humanities and Innovation studies in ESIR (end on June 2018). As of September 2018, he joined ISTIC (CS department of University of Rennes 1), and became an elected member of its administrative board and is co-responsible, with Annie Foret, for the second year CS studies (bachelor).

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

- 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 septermber 2018. In 2018-2019, she is responsible with Olivier Ridoux of the second year computer science studies at Rennes 1 university.

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

  Since September 2013 until September 2018, she was responsible of the internships of computer science students (Licence 3 - 75 students, Master 1 - 65 students, and Master 2 - 65 students).

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

  She participated to the recruitement committes of 1st 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).

6.2.2   Teaching

- Mireille Ducassé, at INSA, is responsible of three courses, taught in English: *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 10 hours Prolog courses at Master 2 level at Ivane Javakhishvili Tbilisi State University 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 both at Polytech Nantes (24h, 25 students) and at ENSAI Rennes (15h, 21 students).

- 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, compiler design, and innovation at ESIR, and compiler techniques, logic and constraint programming, and epistemology at ISTIC.

- At INSA, Peggy Cellier is responsible of three courses: *Databases and web development* (Licence 3), *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).

6.2.3   Supervision

- PhD: **Clément Gautrais**, Signatures : detecting and characterizing complex recurrent behavior in sequential data, defended October 2018, supervised by Peggy Cellier (25%), Thomas Guyet (25%), René Quiniou (25%), Alexandre Termier (25%)[10]

- 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 Caillaud[11]

- PhD in progress: **Anh-Duong Nguyen**, Compression Based Pattern Mining, started March 2018, supervised by Peggy Cellier (25%), Alexandre Termier (25%)[12], Romaric Gaudel (50%)[13]

- internship (L3): **Grégoire Pacreau**, on "Utilisation de modèles MDL « raffinés » en pattern mining", 2 months, supervised by Peggy Cellier, Romaric Gaudel and Alexandre Termier.

- internship (M2): **Erwan Bourrand**, "AlphaGo goes shopping", 4.5 months, supervised by Peggy Cellier, Romaric Gaudel and Alexandre Termier.

- internship: **Karen Kechis**, on "languages and digital" projects to help learners, 12 weeks, supervised by Annie Foret.

- internship: **Keti Meipariani, Veriko Nikuradze and Mikheil Maisuradze** from Tbilisi STate University, Georgia on "A Base of Inflected Verb Forms for the Georgian Language (Caucasus): A Sewelis Application", 5 months, supervised by Mireille Ducassé.

6.2.4   Juries

- Mireille Ducassé served as president in the PhD committee of Clémence Frioux on "Investigating host-microbiota cooperation with gap-filling optimization problems", Anne Siegel supervisor, University of Rennes 1, 19-11-2018

- Olivier Ridoux served as president in the PhD committee of Tristan Charrier on "Complexité théorique du raisonnement en logique épistémique dynamique et étude d'une approche symbolique", Sophie Pinchinat and François Schwartzentruber supervisors, University of Rennes 1, 5-12-2018

---

[10]Team Lacodam - IRISA
[11]Team Hycomes - IRISA
[12]Team Lacodam - IRISA
[13]ENSAI

- Annie Foret served as president in the PhD committee of Carlos Cardó on "Algebraic Dependency Grammar", 15-03-2018, at UPC Barcelona (Universitat Politècnica de Catalunya).

  She also gave a talk at a workshop organized at UPC on this occasion, her talk's title was: "Categorial Dependency Grammars with Iterated Sequences".

6.3 Popularization

- Olivier Ridoux participated in Journée Sciences et Musiques of IRISA on October 20th, 2018, to present its eMaestro projet (an electronic assistant to music direction).

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.

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, 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.

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] A. Foret, G. M. Kobele, S. Pogodalla (editors), *Formal Grammar 2018 - 23rd International Conference, FG 2018, Sofia, Bulgaria, August 11-12, 2018, Proceedings*, Lecture Notes in Computer Science, *10950*, Springer, 2018, `https://doi.org/10.1007/978-3-662-57784-4`.

[2] A.-L. Ligozat, P. Cellier, A.-L. Minard, V. Claveau, C. Grouin, P. Paroubek (editors), *Actes de la conférence Traitement Automatique de la Langue Naturelle, TALN 2018*, 2018, `https://hal.archives-ouvertes.fr/hal-01843585`.

Publications in Conferences and Workshops

[3] P. Alizadeh, P. Cellier, T. Charnois, B. Crémilleux, A. Zimmermann, "Étude Expérimentale d'Extraction d'Information dans des Retranscriptions de Réunions", *in : Traitement automatique du langage naturel (TALN)*, Rennes, France, May 2018.

[4] P. Alizadeh, P. Cellier, T. Charnois, B. Crémilleux, A. Zimmermann, "An Experimental Approach For Information Extraction in Multi-Party Dialogue Discourse", *in : Proceedings of the 19th International Conference on Computational Linguistics and Intelligent Text Processing (CiCling 2018)*, LNCS, Springer, 2018.

[5] C. Bobed, L. Douze, S. Ferré, R. Marcilly, "PEGASE: A Knowledge Graph for Search and Exploration in Pharmacovigilance Data", *in : EKAW Posters and Demon-*

strations, P. Cimiano, O. Corby (editors), *CEUR Workshop Proceedings, 2262*, p. 33–36, 2018.

[6] C. Bobed, L. Douze, S. Ferré, R. Marcilly, "Sparklis over PEGASE knowledge graph: a new tool for pharmacovigilance", *in: Int. Conf. Semantic Web Applications and Tools for Life Sciences (SWAT4LS)*, A. Waagmeester, C. J. O. Baker, A. Splendiani, O. D. Beyan, M. S. Marshall (editors), *CEUR Workshop Proceedings, 2275*, 2018.

[7] S. Dandin, M. Ducassé, "ComVisMD-A Visualization Tool for Compact Display of Multidimensional Data: An illustration on Cricket Players Data", *in: IEEE VIS 2018 International Conference, Abstract*, 2018. Video previews available under https://vimeo.com/groups/vis18previews.

[8] S. Ferré, P. Cellier, "How Hierarchies of Concept Graphs Can Facilitate the Interpretation of RCA Lattices?", *in: Int. Conf. Concept Lattices and Their Applications (CLA)*, D. I. Ignatov, L. Nourine (editors), *CEUR 2123*, CEUR-WS.org, p. 69–80, 2018, `http://ceur-ws.org/Vol-2123/paper6.pdf`.

[9] S. Ferré, "Answers Partitioning and Lazy Joins for Efficient Query Relaxation and Application to Similarity Search", *in: Int. Conf. The Semantic Web (ESWC)*, A. Gangemi *et al.* (editors), *LNCS 10843*, Springer, p. 209–224, 2018, `https://doi.org/10.1007/978-3-319-93417-4\_14`.

[10] S. Ferré, "Responsive and Flexible Controlled Natural Language Authoring with Zipper-Based Transformations", *in: Int. Work. Controlled Natural Language (CNL)*, B. Davis, C. M. Keet, A. Wyner (editors), *Frontiers in Artificial Intelligence and Applications, 304*, IOS Press, p. 21–30, 2018, `https://doi.org/10.3233/978-1-61499-904-1-21`.

[11] S. Ferré, "What's New in SPARKLIS", *in: EKAW Posters and Demonstrations*, P. Cimiano, O. Corby (editors), *CEUR Workshop Proceedings, 2262*, p. 41–44, 2018.

[12] A. Foret, "Breton-français et numérique, projet LangNum-br-fr (phase conception)", *in: Conférence Langues et numérique 2018*, Paris, France, July 2018, `https://hal.inria.fr/hal-01959451`.

[13] E. Galbrun, P. Cellier, N. Tatti, A. Termier, B. Crémilleux, "Mining Periodic Patterns with a MDL Criterion", *in: The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2018)*, 2018.

[14] A. Lamercerie, "Formal analysis of natural language requirements for the design of cyber-physical systems", *in: Conférence TALN*, 2018, `https://hal.inria.fr/hal-01970134`.

[15] S. T. Mai, S. Amer-Yahia, A. D. Chouakria, K. T. Nguyen, A. Nguyen, "Scalable Active Constrained Clustering for Temporal Data", *in: Int. Conf. Database Systems for Advanced Applications (DASFAA)*, p. 566–582, 2018.

[16] P. Maillot, C. Bobed, "Measuring structural similarity between RDF graphs", *in: Proceedings of the 33rd Annual ACM Symposium on Applied Computing, SAC 2018, Pau, France, April 09-13, 2018*, p. 1960–1967, 2018, `https://doi.org/10.1145/3167132.3167342`.

Internal Reports

[17] G. Aucher, A. Baire, J. Berbinau, A. Foret, J.-B. Lenhof, M.-L. Morin, O. Ridoux, F. Schwarzentruber, "Principles for a Judgement Editor Based on

BDD", *Research report*, Université de Rennes 1, France, November 2018, Report of a collaborative project between the Cour de cassation, the University of Rennes 1, the ENS Rennes, the INRIA and the University of Torino, from 2013 to 2015., `https://hal.inria.fr/hal-01914593`.

[18] 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`.

[19] H. FERREBŒUF, O. RIDOUX *et al.*, "LEAN ICT, Pour une sobriété numérique", *research report*, The SHIFT Project, France, Mars 2018, Rapport de travail intermédiaire.

[20] S. FERRÉ, "A SPARQL 1.1 Query Builder for the Data Analytics of Vanilla RDF Graphs", *Research report*, IRISA, team SemLIS, 2018, `https://hal.inria.fr/hal-01820469`.

[21] E. GALBRUN, P. CELLIER, N. TATTI, A. TERMIER, B. CRÉMILLEUX, "Mining Periodic Patterns with a MDL Criterion", *research report*, 2018, `http://arxiv.org/abs/1807.01706`.