Team SemLIS

Semantics, Logics, Information Systems for Data-User Interaction

Rennes

Activity Report

2016

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 Members

Head of the team

- Sébastien Ferré, Associate Professor, HDR

Administrative assistant

- Aurélie Patier

Université Rennes 1 staff

- Olivier Ridoux, Professor
- Sébastien Ferré, Associate Professor, HDR
- Annie Foret, Associate Professor, HDR
- Guillaume Aucher, Associate Professor

INSA Rennes staff

- Mireille Ducassé, Professor
- Peggy Cellier, Associate Professor (part time: 80%)

Postdocs

- Pierre Maillot, ANR IDFRAud project

PhD students

- Clément Gautrais, supervised by Peggy Cellier (25%), Thomas Guyet (25%), René Quiniou (25%), Alexandre Termier (25%)[1]

Master students

- Siham Rim Boudaoud
- Aurélien Fleuriot
- Djiby Diallo

Licence students

- Hedi-Théo Sahraoui

[1] Team Lacodam - IRISA

2 Scientific and Historical Context

The SemLIS team is a follow-up of the LIS team. LIS stands for *Logical Information Systems* [25], and the first team was created in 2007. Its main purpose was to combine logics and user navigation to break some limitations of existing systems to retrieve and update information (e.g., hierarchies vs Boolean search). The topic had emerged around 2000 with the PhD thesis of S. Ferré [27], supervised by O. Ridoux. Despite the diversity of our works since then, there is a common pattern: *the design and application of formal languages and formal methods to data and knowledge with a user-centered approach.*

Formal Concept Analysis (FCA) is a central community for the team, and we have had a strong implication in it as authors, committee members, and program chairs. FCA provided the rationale for deriving the navigation structure from data, and for driving the user-system dialogue. Our first contribution was to generalize FCA to the use of logics for representing knowledge [22]. After 2000, a number of new LIS members and PhD students contributed to widen the theoretical and application scopes. The PhD of Yoann Padioleau led to applications to file systems [34]. The PhD of Peggy Cellier led to the domain of data mining applied to software engineering [16]. The PhD of Olivier Bedel led to applications for geographical information systems [9], and a collaboration with geographers. The arrival of Annie Foret led to applications to linguistic resources, and to the domain of logics for the analysis of natural languages [31]. The PhD of Pierre Allard led to the domain of business intelligence (à la OLAP) for richer visualizations [1]. The PhD of Alice Hermann led to the community of the Semantic Web [20, 33], which now plays a central role. LIS results and prototypes have been extended from FCA to the richer knowledge representations of the Semantic Web. In 2008, Mireille Ducassé switched from software engineering applications to the use of *logical information systems* as a support system for group decision and negotiation [18, 17]. Finally, the PhD of Mouhamadou Ba led to the application domain of bioinformatic workflows [4]. In the meantime, connections were made with the community of Faceted Search, and this led to a collaborative book [35], and participation to a COST action. Finally, Guillaume Aucher recently joined us, and brought his expertise in logics for the representation of uncertain knowledge, and in epistemic reasoning for multi-agent systems [2, 3].

The development and experimentation of softwares has played a central role in the scientific development of the LIS team (e.g., CAMELIS, LISFS, GEOLIS, SEWELIS[2]). They have been built at different system levels (file system, desktop, Web application), and applied to various domains (personal data, software engineering, geographical data, linguistic resources, etc.). SEWELIS is the most recent and advanced implementation of *logical information systems*. It is a generic tool for editing, querying, and exploring a Semantic Web dataset. Its users are guided in all tasks, and are never exposed directly to the formal languages used by the system for reasoning on knowledge. SEWELIS is the current version under development, and will be the basis for a number of future works.

[2] `http://www.irisa.fr/LIS/softwares/`

3 Scientific Foundations and Former Results

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 and Guillaume Aucher study different variants of substructural logics, respectively for the analysis of natural languages, and for reasoning under uncertainty in a multi-agent context. Peggy Cellier did her PhD thesis on the application of data mining to the localization of faults in programs. 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.

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 reasoning. We also study various kinds of logics such as epistemic logics, dynamic logics, or deontic logics. In those topics, we have for instance contributed to the modular composition of logics [23], to the representation of updates [2], and to the representation of complex expressions in RDF [10].

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)

[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.

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 [11], and their extension with option and iteration [10], to a CNL (SQUALL) for querying and updating RDF graphs [30], and to the discovery of linguistic patterns from texts [13].

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 [13]. It also has competencies in learning the grammar of natural languages from a structured corpus [11]. Moreover, the LIS team was scientifically founded on Formal Concept Analysis (FCA) [GW99]. It produced FCA-based contributions for data mining [15] and machine learning [24], as well as for data exploration [28].

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], Geographical Information Systems (GIS) [LT92], and multi-agent systems [Woo09,SLB09]. In those topics, we have for instance contributed to the exploration of geographical data [9], to the discovery of functional dependencies and association rules with OLAP cubes [1], and to the extension of faceted search to RDF graphs [20].

Logic and Artificial Intelligence. Many of our daily activities which were in the past performed in the 'real' world and in interaction with other humans, are carried out in a numeric

[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.

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

[Woo09] M. Wooldridge, *An introduction to multiagent systems*, Wiley, 2009.

[SLB09] Y. Shoham, K. Leyton-Brown, *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*, Cambridge University Press, 2009.

world in interaction with non-human 'agents': e-commerce, e-voting, e-banking, e-government, etc. This transposition of our activities into the numeric world will most probably continue to a certain extent in the future, and this numeric world will therefore play a more and more important role in our everyday life. In order to better control and develop this numeric world, we need to understand, model and predict this interaction between agents. Logic and game theory are currently the most influential theoretical framework that deal with this interaction. In order to better control and develop this numeric world, we need to understand, model and predict this interaction between agents. Logic and game theory are currently the most influential theoretical framework that deal with this interaction. The work in logic and game theory is centered around the representation and the reasoning about uncertainty in a setting with multiple agents. On the one hand, because uncertainty permeates a wide range of human and artificial activities, the representation and the reasoning about uncertainty has been addressed under different forms in many branches of theoretical computer science: distributed systems, control theory, game theory, security and privacy... On the other hand, logic and epistemic logic have developed independently in the last decades a number of abstract and powerful tools to represent and reason about uncertainty. Our first contributions consist in developing and studying further these logical tools while striving to relate and unify them. Our second contributions consist in applying and using them to reformulate various theories so as to identify some common mathematical structures. This led us to investigate belief revision and planning in a multi-agent setting, and to transfer methods and results of modal logic for formalizing the notion of 'only knowing', an epistemic concept of practical relevance for knowledge bases.

Although the above five topics correspond to traditionally distinct research domains and communities, they are often found in combinations in today's research challenges and conferences. Many of our contributions actually lie at the crossing of several topics: e.g., the application of symbolic data mining to linguistic data such as texts [13], the interactive exploration and filtering of data mining results [14], the representation and querying of natural language resources such as lexicalized grammars [31], or the combination of a query language, natural language generation, and user-data interaction to help users explore the Semantic Web [6]. We believe that all those topics are essential, and need to be combined, in order to achieve our objectives.

4 Objectives and Scientific Challenges

We first state the high-level objectives of the team proposal, before illustrating them on a concrete scenario. We then list a number of specific goals at different time ranges (short/mid/long term).

4.1 High-Level Objectives

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 needs (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[3]. 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 same approach that targets user guidance as a trade-off between full automation (aka. artificial intelligence) and no automation (aka. programming). We are conscious that this set of objectives is ambitious but we think we

[3] `http://www.enseignementsup-recherche.gouv.fr/cid86746/strategie-nationale-de-recherche-rapport-de-proposit` `html`

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) [29], to encapsulate an expressive formal language into data-user interaction and natural language.

4.2 Illustration Scenario

> *To illustrate our objectives, we imagine a scenario in the near future where a* Logical Information System (LIS) *enriched by solutions to the above objectives O1-O5 is put to use. This scenario aims at giving a more concrete shape to our objectives, and pushing the limits of imagination of the reader (as well as the author!). As a disclaimer, it certainly shares motivations with other teams, and there would be many relevant work to cite, but we refrain to do so here for the sake of narration and conciseness (citations are given in Section 4.3 along with more detailed objectives).*

Let us consider a group of people whose mission is to improve collaboration between different services of a company. Each member of the group comes from a different service, and has therefore a different point of view and knowledge on the company. Of course, each service has already an information system but each information system has its own schemas, and APIs. The first task is therefore to merge them. The group creates a new *LIS* – this is as easy as creating a new document on a collaborative platform – and selects the right wrapper for each data source in order to project them all in a same representation format (based on Semantic Web technologies) [FLEX]. The *LIS* makes it possible for all members to explore the different data sources, and to acquire some knowledge about the other services [AUTO]. Because the data sources have evolved independently, they probably use different *ontologies* – or no ontology at all! – and are therefore largely disconnected and incomparable. The members collaborate through the *LIS* to *align* the different ontologies so that the different data sources share a same vocabulary, and become tightly interlinked [SEM]. The tedious process of alignment can be done incrementally, on the need, and the *LIS* assists the process through examples and transformations rather than through formal specifications [CON].

At that point, the group can perform deep analytics of the joint activities of the different services [AUTO]. Dynamic views, dashboards, and feeds can be defined, and can be broadcasted to a large audience or sent to particular individuals according to their information needs. Those definitions are internally expressed in expressive formal languages (queries, visualizations, scripts, etc.), and yet, all interactions and specifications go through point-and-click interfaces and natural language [AUTO]. Moreover, everything is expressed concretely in terms of the actual data, and feedback about user actions are given at all time [CON].

The analytics have revealed to the group missing information, and inadequate modelling in the ontologies. The members collaborate under system guidance to incrementally improve the ontology, and to fill in the gaps [FLEX,COLL]. The *LIS* assists them in applying data transformations that preserve the consistency of the whole [CON]. When new information are input, machine learning techniques are used to *align* them on existing information [SEM]. In this way, it is easier to preserve the data invariants, even when they are implicit. When the new information are massive or unknown to the group members, rows of questions can be generated and sent to people in the company according to their profile [COLL]. When a person receives such a question, the *LIS* assists him/her like above to build an answer [AUTO]. The answers are automatically integrated by the *LIS*, and the group can analyze them globally to check and correct them [COLL]. In more complex situations, the group can set up a workflow that makes people and machines work together: the machines run automatic processes, possibly acting on the external world, and people do any task that cannot or should not be automated, like taking decisions. Workflows can themselves be defined by users, relying on the same guidance as for analytics, with suggestions derived from previous user actions [AUTO]. Again, no formal language is exposed, and users retain full control of what the *LIS* can do.

The *LIS* co-evolves with its users, without any separation in time between design and production [FLEX]. It can start simple, and then grow in complexity. It is not intelligent in itself, but it builds on the intelligence of its users, and magnifies it at the collective level [COLL]. It gains knowledge from its users, and offers them powerful information access, assistance, and automation.

4.3 Scientific Challenges and Goals

The realization of the high-level objectives O1-O5 requires solutions to a number of scientific challenges. In the following, we shortly define four major scientific challenges, and list for each of them a number of more concrete goals at different time scales: short term, mid term, long term. We do not provide a systematic correspondance between objectives and scientific challenges because a challenge often involves several objectives, and the concrete goals provide a progression in the realisation of the objectives. For each goal, we give the initials of the main participants.

4.3.1 Information Extraction

This challenge is on acquiring structured and semantic data (typically RDF), by extraction from raw data (e.g., texts, GPS tracks). It mostly concerns objective SEM. There will be a progress from relatively structured data such as GPS tracks or tables to less and less structured data. The least structured data come from texts and images, and we may rely on existing techniques (such as developed by team LinkMedia at IRISA) as a first analysis step.

RDF representation of temporal and spatial data (short term) [OR,SF]. Temporal representations are not only about dates and times but also about periods, durations, repeated events, and also relative expressions (e.g., yesterday, last Christmas). Spatial representations include geo-positions, geometries, shapes, directions, or topological relationships (e.g., "next to", "in"). A challenge is to combine expressivity (e.g., geometries like points, lines, polygons), conciseness of representations, and tractability of computations (e.g., computing durations or distances). There exists some work (e.g, Strabon [KKK12]) that can be integrated into SEWELIS, and improved further.

[KKK12] K. Kyzirakos, M. Karpathiotakis, M. Koubarakis, "Strabon: A Semantic Geospatial DBMS", *in: Int. Semantic Web Conf.*, Springer, 2012, p. 295–311.

Sequence discretization and segmentation (mid term) [PC,OR]. Sequences are a common form of data, and are typically generated automatically by sensors, as system logs, or as user histories. In order to apply symbolic approaches, and to make sense of sequential data, an important task is their discretization and segmentation. A tool like SAX [SK08] can be used to discretize numeric sequences, and pattern mining algorithms can be used to segment sequences according to repeated patterns [AS95].

RDFization of semi-structured Open Data (mid term) [MD,SF]. Open Data often comes in the form of semi-structured data: e.g., spreadsheets, CSV files [HFP+06], lists. It is also often noisy and inconsistent in its terminology and notations. A fully automatic RDFization seems illusory, and the challenge is to support the user as much as possible in the RDFization process. The tool OpenRefine[4] is a good starting point in this goal. A possible approach would be a naive RDFization followed by the application of successive RDF transformations, which would be guided by the system but under control of the user.

Information extraction from texts (long term) [AF,PC]. There exist mature solutions for named entity recognition, and for analyzing texts at the word level. However, making full use of syntax and pragmatics to generate a rich semantic representation of the content of a text remains an ambitious challenge. We would first focus on the extraction of relationships between entities in order to produce RDF graphs from texts.

4.3.2 Expressivity

This challenge is on reconciling the expressive power of formal languages (e.g., SQL) and the ease-of-use of graphical interfaces and navigation (e.g., Excel). It mostly impacts objectives AUTO (autonomy), and FLEX (flexibility). Our approach to this challenge is based on the ACN framework, Abstract Conceptual Navigation [29]. In ACN, navigation places are concepts, which are characterized by formal expressions, and navigation links are transformations of those expressions. Guidance in the navigation space should be designed to be safe (no dead-ends) and complete w.r.t. the formal language. Formal expressions and transformations can be hidden behind more natural presentations (e.g., natural language, graphics) because all user input goes through the selection of suggested transformations.

Analytical queries (short term) [SF,OR]. Analytical queries are typically offered by OLAP or spreadsheets and used in Business Intelligence (BI). SEWELIS does not yet support such queries, despite the fact that they can be expressed in SPARQL 1.1 thanks to

[4] `https://github.com/OpenRefine`

[SK08] J. Shieh, E. Keogh, "i SAX: indexing and mining terabyte sized time series", *in: ACM SIGKDD int. conf. Knowledge discovery and data mining*, ACM, p. 623–631, 2008.

[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.

[HFP+06] L. Han, T. Finin, C. Parr, J. Sachs, A. Joshi, "RDF123: a mechanism to transform spreadsheets to RDF", *in: Nat. Conf. Artificial Intelligence (AAAI)*, AAAI Press, 2006.

multi-dimensional queries, computations, and aggregations. Therefore, extending SEWELIS to analytical queries can be done by increasing SEWELIS' expressivity and guidance to cover those SPARQL features. An objective is to perform analytical queries directly against RDF graphs, while existing approaches in the Semantic Web generally propose to first extract tabular views over RDF graphs, and then apply classical approaches to those [KH11,CGMR14]. The problem with the latter approach is that domain experts rely on SW experts for the extraction of views, and that it is difficult to anticipate all useful data views. Indeed, because of their relational nature, RDF graphs can be the source of many different views on data. In our approach, the additional SPARQL features would translate to increased expressivity and navigability.

Automation (mid term) [SF,GA,MD]. Automation comes under different forms in information systems: inference rules based on logic, business rules triggered by some conditions and performing actions, workflows to coordinate people actions and automatic processes. In each case, a formal language has to be chosen or designed, and the challenge is to extend guidance to cover them in a safe, complete, and usable way. A key to usability, and a difficulty, is to support the definition of rules as general as possible, while giving suggestions and feedback as close to concrete examples as possible. The standard SBVR (Semantics of Business Vocabulary and Business Rules[5]) proposed by the OMG may help solving this problem and could be a good source of inspiration. SBVR is based on the abstract language of first-order deontic logic and it is intended to be the basis for formal and detailed natural language declarative description of a complex entity, such as a business. It is intended to formalize complex compliance rules, such as operational rules for an enterprise, security policy, standard compliance, or regulatory compliance rules. Such formal vocabularies and rules can also be interpreted and used by computer systems.

Exploring the immaterial (long term) [PC,GA,SF,AF]. ACN applications have so far been "material" in that concept extensions are made of concrete objects (e.g., query results). Immaterial ACN applications are where formal expressions have no concrete instance or an intractable number of instances. Two interesting examples are constraint solving problems and programming. In the case of constraints, a formal expression would be a set of user-defined constraints, and its instances would be all possible solutions. In the case of programming, a formal expression would be a program, and its instances would be all possible execution traces. The challenge is to provide guidance that is based on possible instances, without having to compute explicitly those instances, like we do in SEWELIS with query results. The choice of the formal language is also a key issue for usability given the high expressivity of constraint and programming languages. A nice starting point for interactive programming user interfaces

[5] `http://www.omg.org/spec/SBVR/1.3/PDF/`

[KH11] B. Kämpgen, A. Harth, "Transforming statistical linked data for use in OLAP systems", *in: Int. Conf. Semantic systems*, ACM, p. 33–40, 2011.

[CGMR14] D. Colazzo, F. Goasdoué, I. Manolescu, A. Roatiş, "RDF analytics: lenses over semantic graphs", *in: Int. Conf. World Wide Web*, ACM, p. 467–478, 2014.

is Scratch from MIT [RMMH+09]. In the continuation of the Web of data (aka. the Semantic Web), one could also imagine a *Web of programs*, with ideas from Unison[6]. Because SBVR permits to exchange the meanings of concepts and business rules between users and tools as well as between tools without losing information about the essence of those meanings, it may also be instrumental in achieving this long-term goal.

4.3.3 Intelligent Assistance

This challenge is on defining an intelligent assistance based on data mining, which should be as complete as possible while avoiding to overload users with information. Such an assistance can be applied to information retrieval, knowledge discovery, knowledge authoring, and decision making. This challenge mostly concerns objectives CON (confidence and control), and AUTO (autonomy).

Graph pattern mining (short term) [PC,SF]. Graph patterns are already used in SEWELIS queries, but are built step-by-step by users. The goal is here to mine interesting (e.g., frequent) graph patterns in RDF graphs to support more intelligent assistance. Unlike existing graph mining approaches, the objective would be to compute graph patterns lazily, and to avoid showing them to users. For example, SEWELIS uses tree patterns to find objects that are similar to a new object, and uses similar objects to suggest additional descriptors for the new object. Extending tree patterns to graph patterns would enable a more accurate notion of similarity. On user's request, graph patterns could be shown to users as an explanation for suggestions, and could also be used to generate inference or business rules (see 'Automation' in previous section).

Planning-based guidance (mid term) [GA,MD,SF]. In instances of our ACN framework where a navigation goal can be specified, guidance safeness may not be decidable at the level of navigation steps but only at the level of navigation paths. An example is the construction of a workflow where input and output types are specified. The insertion of a service may be valid while not making any progress from inputs to outputs. In such cases, planning techniques could be used to check, for each possible step, the existence of navigation paths leading to the navigation goal. In fact, we may resort to techniques of *distributed* planning since we have multiple services and users interacting, and even *epistemic* planning [3]: we may want for example to avoid that certain information be disclosed to some users or services, or we may want to enforce a navigation goal that specifies that a user should know a certain piece of information. From that point of view, the research carried out in planning and more particularly *epistemic* planning is definitely relevant to achieve this goal.

[6] `https://pchiusano.github.io/2014-09-14/unison.html`

[RMMH+09] M. Resnick, J. Maloney, A. Monroy-Hernández, N. Rusk, E. Eastmond, K. Brennan, A. Millner, E. Rosenbaum, J. Silver, B. Silverman *et al.*, "Scratch: programming for all", *Communications of the ACM 52*, 11, 2009, p. 60–67.

The "Imitation Game" (long term) [SF,PC,GA]. The game here is not for the system to fool the users, but to strive to anticipate user actions in order to progressively automate them. Its task is to learn from user actions and interactions, to predict the next user actions and to suggest them to users. It is analogous to predictive text input but generalized to any action and interaction with the information system. A possible approach is to apply graph pattern mining on the structured RDF representation of user actions and interactions. In our view, the system only make suggestions, and the user retains full control and decision. Automation comes either by user commitment to suggested rules, or by explicit delegation of decision, inside a fixed perimeter, based on trust. In this way, the information system is more an extension of users' intelligence than a replacement.

4.3.4 Practicality

Necessary conditions for the practicality of information systems in real situations include scalability, usability, and collaboration. The three are related to amount: amount of data that can be handled (scalability), amount of people that can use it (usability), and amount of people that can use it together (collaboration). In our case, scalability is more concerned with achieving low complexity for algorithms (ideally, linear) than with managing huge volumes of data (a billion RDF triples is already quite big). This challenge is on finding approaches and algorithms that are compatible with both large amounts of data and large amounts of people. The usability challenge mostly impacts objective AUTO (autonomy), and COLL (collaboration).

Visualization of query results (short term) [OR,SF,MD]. The goal is to go beyond textual and tabular results with graphical visualizations such as charts, maps, timelines, etc. Like for expressing queries, our objective is to combine flexibility, usability, and intelligent assistance. It should not be possible to generate visualizations that are ill-typed or look ugly because of inadequate data distribution. A declarative and compositional visualization language would make it possible to combine the different kinds of visulizations in a safe and flexible way, rather than in pre-defined idiomatic ways. Also, the visualizations would dynamically adjust through navigational steps that change the query, and hence the visualized data, like in Allard's thesis [2].

Interaction in natural language (mid term) [AF,SF]. While step-by-step navigation is safe and supportive for users, especially when they have not a precise knowledge of the system capabilities, it is less efficient than dialogue in written or spoken natural language (NL). The goal would be to add such a dialogue layer on top of the navigation layer. Standard NLP tools would be used to analyze user utterances, and the navigation layer would be used to interpret them in the specific context of the information system. In case an utterance would not be fully understood by the system, users could resort to navigation to fill in the gaps.

Distributed information systems (long term) [GA,OR]. A long term challenge of the Semantic Web is to build a Web-scale information system where both knowledge stores and reasoning tasks are distributed. It covers many domains of computer science, including systems and networks, and our focus is on studying the impact of knowledge distribution on reasoning tasks such as: answering a query, mining patterns, supporting decision making. This distributed aspect of information systems calls for methods and techniques developed in distributed artificial intelligence and more particularly in the field of multi-agent systems [Woo09,SLB09].

5 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 [AF,GA,OR,MD]. 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 [SF,MD,PC]. 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.

We detail below a few application domains for which we have already done concrete work and contributions, and for which we have a long-lasting interest.

[Woo09] M. Wooldridge, *An introduction to multiagent systems*, Wiley, 2009.
[SLB09] Y. Shoham, K. Leyton-Brown, *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*, Cambridge University Press, 2009.

5.1 Group Decision and Negotiation

Participants: Mireille Ducassé, Peggy Cellier.

Group decision and negotiation focuses on complex and self-organizing processes that constitute multiparticipant, multicriteria, ill-structured, dynamic, and often evolutionary problems. Group decision and negotiation refers to the whole process or flow of activities relevant to reaching a group decision, and not merely to the final choice - aspects of the process in group decision and negotiation include scanning, communication and information sharing, problem definition (representation) and evolution, alternative generation, and social-emotional interaction. Group decision support systems (GDSS) and negotiation support systems (GDNSS) are amongst the major approaches to address the problems.

In the current thread of research, we are showing that Logical Information Systems provide an innovative technological support for most of the above mentioned aspects of GDSS. In particular, the navigation and filtering capabilities of LIS help detect inconsistencies and missing knowledge during meetings. The updating capabilities of LIS enable participants to add objects, features and links between them on the fly. As a result the group has a more complete and relevant set of information. Furthermore, the compact views provided by LIS help participants embrace the whole required knowledge. The group can therefore build a shared understanding of the relevant information previously distributed amongst the participants. Lastly, the navigation and filtering capabilities of LIS are relevant to quickly converge on a reduced number of targets. A future trend of research will be to investigate how LIS can also support negotiation.

5.2 Geographical Information Systems

Participants: Olivier Ridoux, Peggy Cellier, Sébastien Ferré.

Geographical Information Systems (GIS) is an important, fast developing domain of Information technology, and it is almost absent from INRIA projects. It is especially important for local communities (e.g. region and city councils).

Geographical information systems [LT92] handle information that are localized in space (*geolocalized*). GIS form an area which incorporates various technologies such as web, databases, or imaging. One characteristic of GIS is their organization as *layers*. This is inherited from the plastic sheets that where used until recently for drawing maps. A layer represents the road system, another the fluvial system, another the relief, etc. This is another instance of the tyranny of the dominant decomposition, and is not satisfactory: to which layer belong bridges, into which layer can we represent a multimodal network? Moreover, mining GIS is known to be difficult for the same reason; the layer structure makes inter layer relationships difficult to discover.

The first advantage of applying LIS to GIS is to allow cross-layer navigation. Another advantage is to permit a logical handling of scales. In current GIS systems, scales are treated as different layers, and it is difficult to keep the consistency between all layers that describe the

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

same object. Another advantage that we have observed in a preliminary work is that LIS helps cleaning a data-base. This was not expected, and opens an interesting research area. Another characteristic of GIS is an intensive usage of topological relations (touches, overlaps, etc) and geographical relations (North, upstream, etc). Logic offers a rich language for expressing these relations and combining them.

5.3 Software Engineering

Participants: Olivier Ridoux, Annie Foret.

This application domain can be seen as a particular case of the “social sciences” domain where the object of study is not a living organism or a territory but a software, and where the domain experts are developers. That domain plays an important role in the history of the team. The formal nature of computer programs (syntactic structure, type systems, versioning) lends itself to a structured and semantic representation. However, software also contains elements in natural language (e.g., comments, commit messages). Possible applications are the retrieval of software components and their composition, or the analysis of code or program traces.

6 Softwares

6.1 Camelis, Sewelis, and Sparklis

Participants: Sébastien Ferré.

Camelis is a stand-alone application that allows to store, retrieve and update objects through a graphical interface. Its main purpose is to experiment with the LIS paradigm. In particular, it has been very useful for refining the query-answer principle in special circumstances (e.g. when there are many answers, or when there are few answers). It is currently used as a personal storage device for handling photos, music, bibliographical references, etc, up to tens of thousands of objects. It implements as closely as possible the LIS paradigm. It is generic w.r.t. logics, and is compatible with our library of logic functors, LogFun. It is available on Linux and Windows, and comes with a user manual.

An important extension, Sewelis, has been developped to browse RDF(S) graphs, a Semantic Web standard. It uses a query language whose expressivity is similar to SPARQL, the reference query language of the Semantic Web. The LIS navigation has been proved safe (i.e., does not lead to dead-ends), and complete (i.e., can reach all conjunctive queries), so that users can perform complex searches easily and safely [20]. Sewelis also supports the guided creation and update of objects, according to the UTILIS approach [33]. In 2015, some code of Portalis (see 2014 report) was integrated to Sewelis to make it a multi-user Web application. A beta version is available online.

Sparklis is a re-implementation of the querying capabilities of Sewelis as a Web application on top of SPARQL endpoints. Its main advantages compared to Sewelis are: (1) no setup, just load a page in a browser, (2) direct exploration of remote SPARQL endpoints, (3) scaling up to large triples that contain up to billions triples, like DBpedia, and (4) verbalization of queries in natural language for better readability. In 2016, Sparklis has been extended to cover SPARQL expressions and nested aggregations. This enabled us to take part in the QALD6 challenge (Question Answering on Linked Data), on the task of "statistical questions". A presentation and a demo were given at ESWC in Heraklion. A Spanish version has been developed with the help of José Enrique Ortiz Vivar from Chile, in addition to English and French. A demo has been given on the occasion of the "5 years" of ISTIC. A major result is the adoption of Sparklis at Persée, a bibliographic portal for Humanities and social sciences whose users are researchers in those domains. Our collaboration with Persée has already led to a number of improvements of Sparklis in terms of features and usability. A valorisation process is also ongoing with Ouest Valorisation, and led by Romain Thomas.

6.2 Typed grammars

Participants: Annie Foret [contact point], Denis Béchet [LINA-Nantes].

A Pregroup ToolBox is under development on the gforge Inria as a collaborative work with LINA. It includes a generic pregroup parser (LINA) and grammar lexicon definitions and manipulation tools based on XML. An interface with Camelis has been developped (from Camelis to the Pregroup XML format, and the other way round). It has been used to define and experiment grammar prototypes for different natural languages.

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

6.3 SDMC: Sequential Data Mining under Constraints

Participants: Pierre Holat [LIPN], Nicolas Béchet [IRISA/Vannes], Peggy Cellier, Thierry Charnois [LIPN], Bruno Crémilleux [GREYC].

SDMC is a tool to extract sequential patterns [8][7]. It is specially adapted for text mining, allowing part-of-speech extraction and linguistic constraints like gap and membership. Already used by other institutions, the tool can be useful for many text mining tasks like clustering or named entity recognition.

SDMC is available as a Web application: `http://tal.lipn.univ-paris13.fr/sdmc/`.

6.4 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 [30]. 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

[7] N. Béchet, P. Cellier, T. Charnois, B. Crémilleux, “SDMC : un outil en ligne d’extraction de motifs séquentiels pour la fouille de textes”, in: Extraction et gestion des connaissances (EGC’2013 - Session démo), A. F. Gelbukh (editor), 2013.

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 under two forms: one that translates SQUALL sentences to SPARQL, and another one that directly return query answers from a SPARQL endpoint.

6.5 PEW: Possible World Explorer

Participants: Sébastien Ferré, Sebastian Rudolph.

The Possible World Explorer (PEW) 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 [26]. 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 asserting negative axioms (about things that should not exist), PEW also allows for the definition of named classes (OWL equivalent class axioms), and for the creation of named individuals as instances of class expressions (OWL class assertion axioms).

PEW has been extended to better support the edition of ontologies *de novo* [11]. It allows for the extension of the signature by creating new classes, properties, and individuals; and it covers a wide range of OWL axioms. The user interface has also been extended with shortcuts to speed up the removal of impossible worlds.

7 Contracts and Collaborations

7.1 IDFRAud: An Operational Automatic Framework for Identity Document Fraud Detection and Profiling (ANR)

Participants: Sébastien Ferré, Peggy Cellier, Pierre Maillot.

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 funds 2 postdoc years and research costs (e.g., missions, machines).

This year, two consortium meetings were held on April 12th, and September 21st, as well as a number of working meetings with AriadNEXT on various topics (knowledge base, knowledge acquisition, document analysis workflow inference, data mining of false documents, ...). Pierre Maillot organized at IRCGN a field user study with ID document experts to evaluate the benefits of FORMULIS, an intelligent form-based user interface for the guided acquisition of RDF metadata [13]. On 15-16 December, a colloquium was organized by ENSP to present the current results of the projects to stakeholders from various institutions and companies. On 7 December, the project passed its mid-term evaluation by ANR, and was found excellent in terms of collaboration between partners, and in terms of industrial transfer. A poster and demonstration were presented at the 20th International Conference on Knowledge Engineering and Knowledge Management [13].

7.2 PEGASE: Improved pharmacovigilance and signal detection with groupings

Participants: Sébastien Ferré, Annie Foret, Peggy Cellier, Mireille Ducassé.

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 ergonoms of the medical domain from CIC-IT Evalab.

The project has been selected by ANR in August, and the first consortium meeting took place at HEGP on December 12th.

7.3 Cour de Cassation

Participants: Guillaume Aucher, Annie Foret.

Guillaume Aucher and Annie Foret were members of a group working with "The French Cour de Cassation". This multi-disciplinary project on "Développement d'un prototype de logiciel d'aide à la prise de décision lors de la rédaction de jugements" has received funds from ENS CHRC (Collège de Recherche Hubert Curien) (proposal submitted by François Schwarzentruber at ENS Rennes) ; the project also received funds from Rennes1 University as a "Défi émergent" (2014-2015, submitted by A. Foret). A summary of the work carried out in this project and the results and prototype produced have been submitted for publication.

7.4 Other Collaborations

- Annie Foret is an external collaborator of LINA (research lab. Nantes), in TALN team (Natural Language Processing), and 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).
- Peggy Cellier collaborates with the members of the ANR project HYBRID[8] on the part about data mining for natural language processing.

  She is also involved in the FUI project REUs (started at the end of 2016) in collaboration with the GREYC (Caen) about information extraction from meeting reports.

  Since the end of 2016, she also collaborates with the members of the industrial project ITRAMI (started in 2015) led by Alexandre Termier about the analysis of traces.
- Sébastien Ferré collaborates with Persée, a research unit in the domain of humanities and social sciences, on the use of Sparklis (see 6.1).

8 New Results

8.1 Sparklis: An Expressive Query Builder for SPARQL Endpoints with Guidance in Natural Language

Participants: Sébastien Ferré.

Linked data is increasingly available through SPARQL endpoints, but exploration and question answering by regular Web users largely remain an open challenge. Users have to choose between the expressivity of formal languages such as SPARQL, and the usability of tools based on navigation and visualization. In a previous work, we have proposed Query-based Faceted Search (QFS) as a way to reconcile the expressivity of formal languages and the usability of faceted search. In this work [6], we further reconciled QFS with scalability and portability by building QFS over SPARQL endpoints. We also improved expressivity and readability. Many SPARQL features are now covered: multidimensional queries, union, negation, optional, filters, aggregations, ordering. Queries are now verbalized in English (and French), so that no knowledge of SPARQL is ever necessary.

All of this is implemented in a portable Web application, Sparklis, and has been evaluated on many endpoints and questions. No endpoint-specific configuration is necessary as the data

[8] http://hybride.loria.fr

schema is discovered on the fly by the tool. Online since April 2014, thousands of queries have been formed by hundreds of users over more than a hundred endpoints. We have analysed in detail usage logs, which show that a number of users have applied Sparklis to their own data, in particular in bioinformatics, and managed to build complex questions in order to satisfy their information needs.

8.2 Bridging the Gap Between Formal Languages and Natural Languages with Zippers

Participants: Sébastien Ferré.

The Semantic Web is founded on a number of Formal Languages (FL) whose benefits are precision, lack of ambiguity, and ability to automate reasoning tasks such as inference or query answering. This however poses the challenge of mediation between machines and users because the latter generally prefer Natural Languages (NL) for accessing and authoring knowledge. We have introduced the N<A>F design pattern based on Abstract Syntax Trees (AST), Huet's zippers and Montague grammars to zip together a natural language and a formal language [9]. Unlike question answering, translation does not go from NL to FL, but as symbol N<A>F suggests, from ASTs (A) of an intermediate language to both NL (N<A) and FL (A>F). ASTs are built interactively and incrementally through a user-machine dialog where the user only sees NL, and the machine only sees FL. The design pattern is demonstrated with an application to a subset of SPARQL. Its effectiveness and generality are then demonstrated by reporting on three other applications: querying SPARQL endpoints, authoring RDF descriptions, and completing OWL ontologies.

8.3 Graph-FCA in Practice

Participants: Sébastien Ferré, Peggy Cellier.

With the rise of the Semantic Web, more and more relational data are made available in the form of knowledge graphs (e.g., RDF, conceptual graphs). A challenge is to discover conceptual structures in those graphs, in the same way as Formal Concept Analysis (FCA) discovers conceptual structures in tables. Graph-FCA has been introduced in a previous work [19] as an extension of FCA for such knowledge graphs. This year, algorithmic aspects and use cases are explored in order to study the feasibility and usefulness of G-FCA [8]. We consider two use cases. The first one extracts linguistic structures from parse trees, comparing two graph models. The second one extracts workflow patterns from cooking recipes, highlighting the benefits of n-ary relationships and concepts.

8.4 An RDF Design Pattern for the Structural Representation and Querying of Expressions

Participants: Sébastien Ferré.

Expressions, such as mathematical formulae, logical axioms, or structured queries, account for a large part of human knowledge. It is therefore desirable to allow for their representation and querying with Semantic Web technologies. We have proposed an RDF design pattern [10] that fulfills three objectives. The first objective is the structural representation of expressions in standard RDF, so that expressive structural search is made possible. We have proposed simple Turtle and SPARQL abbreviations for the concise notation of such RDF expressions. The second objective is the automated generation of expression labels that are close to usual notations. The third objective is the compatibility with existing practice and legacy data in the Semantic Web (e.g., SPIN, OWL/RDF). We have shown the benefits for RDF tools to support this design pattern with the extension of SEWELIS, a tool for guided exploration and edition, and its application to mathematical search.

8.5 Semantic Authoring of Ontologies by Exploration and Elimination of Possible Worlds

Participants: Sébastien Ferré.

We have proposed a novel approach to ontology authoring that is centered on semantics rather than on syntax [11]. Instead of writing axioms formalizing a domain, the expert is invited to explore the possible worlds of her ontology, and to eliminate those that do not conform to her knowledge. Each elimination generates an axiom that is automatically derived from the explored situation. We have implemented the approach in prototype PEW (Possible World Explorer), and conducted a user study comparing it to Protégé. The results show that more axioms are produced with PEW, without making more errors. More importantly, the produced ontologies are more complete, and hence more deductively powerful, because more negative constraints are expressed.

8.6 Displaying Updates in Logic

Participants: Guillaume Aucher.

We generalized the language of substructural logics interpreted over the ternary relational semantics and introduce a logic called update logic [2]. This is motivated by our intention to capture within the logical framework of substructural logics various logic-based formalisms dealing with common sense reasoning and logical dynamics. This initiative is based on the key observation that an update can be represented abstractly by the ternary relation of the substructural framework. Thus, we introduce three triples of connectives that are interconnected by means of cyclic permutations. The usual fusion, implication and co-implication connectives form one of these triples. We defined a proper display calculus for update logic that generalizes the display calculus for modal logic. We did not resort to structural connectives for truth constants and we show how we can obtain our display rules using Gaggle Theory. We proved the soundness and strong completeness of our display calculus via a Henkin-style construction. Using correspondence results from substructural logics, we also obtained sound and complete display calculi for a wide variety of classical and substructural logics. Then, we defined dual substructural connectives and we provide a display calculus for update logic extended with these dual connectives. Finally, we focused on the specific case of bi-intuitionistic logic for which we provided a novel sound and strongly complete display calculus.

8.7 Dynamic Epistemic Logic in Update Logic

Participants: Guillaume Aucher.

We showed in [3] that dynamic epistemic logic (DEL) is a substructural logic and that it is an extension of the update logic introduced in [2]. We identified axioms and inference rules that completely characterize the DEL product update, and we provide a sequent calculus for DEL. Finally, we showed that DEL with a finite number of atomic events is as expressive as epistemic logic. In parallel, we provided a sequent calculus for update logic which turns out to be a generalization of the non-associative Lambek calculus.

8.8 Using Bids, Arguments and Preferences in Sensitive Multi-unit Assignments: A p-Equitable Process

Participants: Mireille Ducassé, Peggy Cellier.

Bonus distribution in enterprises or course allocation at universities are examples of sensitive multi-unit assignment problems, where a set of resources is to be allocated among a set of agents having multi-unit demands. Automatic processes exist, based on quantitative information, for example bids or preference ranking, or even on lotteries. In sensitive cases, however, decisions are taken by persons also using qualitative information. At present, no multi-unit assignment system supports both quantitative and qualitative information. In this paper, we propose an interactive process for multi-assignment problems where, in addition to bids and preferences, agents can give arguments to motivate their choices. Bids are used to automatically make pre-assignments, qualitative arguments and preferences help decision makers break ties in a founded way. A group decision support system, based on Logical Information Systems, allows decision makers to handle bids, arguments and preferences in a unified interface. We say that a process is *p-equitable* for a property p if all agents satisfying p are treated equally. We formally demonstrate that the proposed process is p-equitable for a number of properties on bids, arguments and preferences. It is also Pareto-efficient and Gale-Shapley-stable with respect to bids. It has been tested on a course allocation case study that spans over two university years. The decision makers were confident about the process and the resulting assignment. Furthermore, the students, even the ones who did not get all their wishes, found the process to be equitable [5].

8.9 Solving Data Mismatches in Bioinformatics Workflows by Generating Data Converters

Participants: Mouhamadou Ba, Sébastien Ferré, Mireille Ducassé.

Heterogeneity of data and data formats in bioinformatics entail mismatches between inputs and outputs of different services, making it difficult to compose them into workflows. To reduce those mismatches, bioinformatics platforms propose ad'hoc converters, called shims. When shims are written by hand, they are time-consuming to develop, and cannot anticipate all needs. When shims are automatically generated, they miss transformations, for example data composition from multiple parts, or parallel conversion of list elements. We have proposed to systematically detect convertibility from output types to input types. Convertibility detection relies on a rule system based on abstract types, close to XML Schema. Types allow to abstract data while precisely accounting for their composite structure. Detection is accompanied by an automatic generation of converters between input and output XML data. We show the applicability of our approach by abstracting concrete bioinformatics types (e.g., complex biosequences) for a number of bioinformatics services (e.g., blast). Our automatically generated converters help to resolve data mismatches when composing workflows. We conducted an experiment on bioinformatics services and datatypes, using an implementation of our approach, as well as a survey with domain experts. The detected convertibilities and produced converters were validated as relevant from a biological point of view. Furthermore the automatically produced graph of potentially compatible services exhibited a connectivity higher than with the ad'hoc approaches. Indeed, the experts discovered unknown possible connexions [4].

8.10 Formulis: Dynamic Form-Based Interface For Guided Knowledge Graph Authoring

Participants: Pierre Maillot, Sébastien Ferré, Peggy Cellier, Mireille Ducassé.

Knowledge acquisition is a central issue of the Semantic Web. Knowledge cannot always be automatically extracted from existing data, thus contributors have to make efforts to create new data. In [13], we propose FORMULIS, a dynamic form-based interface designed to make RDF data authoring easier. FORMULIS guides contributors through the creation of RDF data by suggesting fields and values according to the previously filled fields and the previously created resources. This system has been tested in an experiment involving the IRCGN forged ID department.

8.11 Understanding Customer Attrition at an Individual Level: a New Model in Grocery Retail Context

Participants: Clément Gautrais, Peggy Cellier, Thomas Guyet, René Quiniou, Alexandre Termier.

We present a new model to detect and explain customer defection in a grocery retail context [12]. This new model analyzes the evolution of each customer basket content. It therefore provides actionable knowledge for the retailer at an individual scale. In addition, this model is able to identify customers that are likely to defect in the future months.

8.12 Exploration of Data from "Défi EGC 2016" with a Logical Information System

Participants: Peggy Cellier, Sébastien Ferré, Annie Foret, Olivier Ridoux.

We have taken part to the contest "Défi EGC 2016"[9]. The challenge was very open. Indeed, given the proceedings of the EGC conferences from 2004, the task was "Surprise us!". We have combined automatic NLP workflows (for french and english), pattern mining and a logical information system to provide an exploration tool for the data [7]. The NLP workflows clean data and add information about the words, the pattern mining step extracts a kind of topics and the LIS allows to navigate in all this knowledge to explore the data.

9 Dissemination

9.1 Invited Talks and Visits

- Carlos Bobed, a postdoc from University of Zaragoza, Spain, is visiting the SemLIS team since November. He has worked on the Semantic Web, question answering, and mobile applications. He is working with Sébastien Ferré on making Sparklis more linguistic-aware. His visit is expected to last until June 2017.
- Thomas Bolander (DTU, Denmark) visited Guillaume Aucher from the 20th of June 2016 until the 24th of June 2016 at IRISA.
- Guillaume Aucher visited Ram Ramanujam (IMSc, Chennai, India) from the 5th of December 2016 until the 23rd of December 2016 in the context of the 'Projet Emergent' entitled "Imperfect Information Games in Epistemic Temporal Logic".

  He was invited by Valentin Goranko (http://www2.philosophy.su.se/goranko/) from the Stockholm department of philosophy for a research collaboration from the 5th of June 2016 until the 15th of June 2016.

  He was invited and participated in the symposium "A decade of ICR" from the 17th until the 18th of March 2016 (http://icr.uni.lu/symposium2016/).

  He was invited and participated in the Journées MAFTEC (https://www.irit.fr/~Frederic.Maris/maftec/) of the pré-GDR IA from the 21st until the 23rd of September 2016 in Toulouse.

  He participated in the school "Linear Logic" and the workshop "Linear logic and philosophy" (https://ll2016.sciencesconf.org/) from the 7th until the 9th of November 2016 in Lyon.

  He participated in the Lorentz center workshop entitled "Unified correspondence" in Leiden in February 2016 (https://www.lorentzcenter.nl/lc/web/2016/757/info.php3?wsid=757).
- Mireille Ducassé visited the Northwestern Polytechnical University of X'ian, and the Beijing Jiaotong University, China, in May 2016 in order to set up dual degree programs for Groupe INSA. She visited the Université Euro-Méditerrannée of Fez, Morocco in order to set up collaboration projects. She visited the Tbilisi State University, Georgia (Caucasus) within the framework of an Erasmus+ International mobility project.

[9] http://www.egc.asso.fr/Manifestations_dEGC/71-FR-Defi_EGC_2016_Communaute_EGC_quelle_histoire_et_quel_avenir

9.2 Young Researchers

- Pierre Maillot is member of the team as postdoc since December 2015 as part of the ANR IDfraud in a two years position. He has defended his thesis in November 2015 on new methods for distributed queries and data quality maintenance in the Semantic Web. During his stay, he has developed FORMULIS, a system for intelligent guided RDF data creation based on SEWELIS. FORMULIS was presented in a poster/demo at EKAW16 [13]. As part of the IDFraud project, he has also participated to technology transfer with the AriadNext company by presenting them Semantic Web technologies.

9.3 Involvement in Scientific Communities

- Peggy Cellier and Sébastien Ferré are the Conference Chairs of the 14th International Conference on Formal Concept Analysis (ICFCA), which will take place at IRISA, Rennes on 13-16 June 2017.

- Sébastien Ferré is a member of the Editorial Board of the International Conference on Formal Concept Analysis (ICFCA). He was also in 2016 a member of the program committee of: the conference CLA (Concept Lattices and Applications), and several workshops: FCA4AI at IJCAI and SoWeDo at IC. Finally, he served as an external reviewer for the Journal of Web Semantics, Discrete Applied Mathematics, and Fundamenta Informaticae.

  Sébastien served as a referee in the PhD committee (Université de Bretagne Occidentale, January 12th, 2017) of Julie Sauvage-Vincent on "Un langage contrôlé pour les Instructions nautiques du Service Hydrographique et Océanographique de la Marine".

- Peggy Cellier is a member of the Editorial Board of the International Conference on Formal Concept Analysis (ICFCA). She served as Senior PC for the french conference EGC (Extraction et Gestion de la connaissance). She was also in 2016 a member of the program committee of: the conference CLA (Concept Lattices and Applications), ECMLPKDD (European Conference on Machine Learning and Data Mining), ICCS (International Conference on Conceptual Structures) and several workshops: DEMO and GAST at EGC and EXCES at SAGEO. Finally, she served as an external reviewer for LOD (Numéro spécial de la Revue Ingénierie des Systèmes d'Informations : "Web de données : publication, liage et capitalisation") and for the Journal of DMA (Discrete Applied Mathematics).

  Peggy co-organized with Thierry Charnois, Andreas Hotho, Stan Matwin, Marie-Francine Moens and Yannick Toussaint, the workshop "DMNLP" [1] at ECMLPKDD (Riva del Garda, Italy) in Septembre 2016 (`http://dmnlp.loria.fr`).

  She served as an examiner in the PhD defense committee (University of Rennes 1, July 18th) of William Correa Beltran on "Découverte et exploitation de proportions analogiques dans les bases de données relationnelles".

  Peggy is a supervisor of the PhD of Clément Gautrais with Thomas Guyet, René Quiniou and Alexandre Termier. She co-supervised the internships of: Hedi-Théo Sahraoui with Pierre Holat, Thierry Charnois and Sébastien Ferré on "A LIS navigation interface for text mining" ; Aurélien Fleuriot with Bruno Crémilleux and Alexandre Termier on "Study of the behaviour of customers from semi-supervised retail data" ; and Djiby Diallo with Mireille Ducassé on "A tool to manage electronic student purse".

She was also member of the thesis supervision committee of Yann Dauxais (University of Rennes 1, October 21th).

- Guillaume Aucher co-organized the workshop “Dynamics in Logic” in Delft in October 2016 (`http://www.appliedlogictudelft.nl/events1/dynamics-in-logic-iv-2016/`).

  He was a member of the PC of IJCAI 2016, KR 2016 and DARe 2016. He served as an external reviewer for the Journal of Logic and Computation and CSL 2016.

  He was member of the *thesis supervision committee* of PhD student Diego Agustin Ambrossio (University of Luxembourg).

- Annie Foret is a member of the Editorial Board of the "Formal Grammar" International Conference (FG), affiliated with ESSLLI. In 2016, she was chair of the Formal Grammar conference, and member of the following conference program committees : Formal Grammar (FG 2016), Logical Aspects of Computational Linguistics (LACL 2016), Traitement automatique des langues celtiques (CLTW 2016). In 2016, she has been a reviewer for the following journals : special issue TLL2015 (Tools for Teaching Logic), Knowledge and Information Systems.

  Annie Foret was a member of the PHD committee of Aleksandre Maskharashvili on "Discourse Modeling with Abstract Categorial Grammars" in December 2016 Nancy.

- Pierre Maillot served as an external reviewer for the IJCAI workshop FCA4AI in 2016 an the french-speaking conference Extraction et Gestion de Connaissance (EGC 2017).

9.4 Academic Responsabilities

- Olivier Ridoux is the head of the DKM department since October 2014, and member of IRISA "Conseil de laboratoire" (laboratory board). Olivier Ridoux is a member of the EcoInfo CNRS service group on sustainable development and information technology (`http://ecoinfo.cnrs.fr`).

- Mireille Ducassé is the director of international relations of the INSA of Rennes since december 2010. As such, she is a member of the direction of the Insa of Rennes. Since March 2014 she is also the coordinator of the international relations of Groupe INSA. In particular, she is responsible for exchange programs involving around 400 student mobilities. She set up a number of dual degrees programs over the past years. She set up International programs with Tbilisi State University in Georgia, UEMF, Morrocco and ITC in Cambodgia. She is responsible of a BRAFITEC agreement for Groupe INSA with Universidade Federal da Paraiba (UFPB) and Universidade Federal de Campina Grande (UFCG), Brazil.

- Sébastien Ferré was in charge of the creation of the new SemLIS team, including a presentation to the scientific direction of IRISA, and finally a presentation to the scientific council of IRISA, after which SemLIS was officialy created. He is vice-director of the MIAGE at ISTIC, and is in charge of Master 1 internships (80 students).

- Olivier Ridoux and Sébastien Ferré are members of the committee of the DKM scientific department (Data and Knowledge Management) at IRISA.

- Guillaume Aucher represents the ISTIC at the SCD (Service Commun de la Documentation) of the University of Rennes 1.

- 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.

- Peggy Cellier is an elected member of the "Conseil de Composante IRISA/INSA" at INSA. She is also an elected member of the "Conseil de laboratoire" at IRISA.

  She organized the bibliographic and internship defense for the Research in Computer Science (MRI) specialism. Since September, 2013 she is responsible of the internships of computer science students (Licence 3, Master 1 and Master 2).

9.5 Teaching

- Guillaume Aucher teaches the course "Foundations of Artificial Intelligence: Logic and Knowledge Representation" (60h, M1) and the course "Logic and Commonsense Reasonning" in the department of philosophy (28h, M1-M2). He created both of these courses. He also participates in the courses of imperative programming (first year), calculability (second year) and formal language theory (third year) in the computer science department.

- 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.

- Sébastien Ferré teaches symbolic data mining, Semantic Web, and compilation techniques at the master level. He also teaches functional programming at license level. This year, he gave a 22h course on the Semantic Web at master level in Polytech Nantes.

- Annie Foret teaches university courses including formal logic and formal methods for computer scientists, object programming (third year), xml technology and related notions (Rennes 1 second-fifth year, fourth year and industry specific sessions) and databases (Rennes 1 and e-miage).

- Olivier Ridoux teaches data-bases, algorithmics, the theory of formal languages and compilation in the engineering school ESIR. He is also in charge of the innovation training in the school, in which he also teaches sustainable development w.r.t. IT, and disruptive innovation (à la Clayton Christensen) w.r.t. scientific revolution (à la Thomas Kuhn). He also teaches logic and constraint programming at the Master level, and an introduction to the principles of IT systems at the Bachelor level. He is also in charge of the institutional communication of ESIR.

- At INSA, Peggy Cellier is responsible of two courses: *Databases and web development* (Licence 3) and *Data-Based Knowledge Acquisition: Symbolic Methods* (Licence 3). She also teaches some other courses: *Database* (Licence 2), *Use and functionalities of an operating system* (Licence 3). This year, she gave a 10h course on data mining at master level in ENSSAT (Lannion). She has also been involved in the IDPE (Ingénieur diplômé par l'état) diploma.
- Pierre Maillot supervises practicals of knowledge acquisition in third year and of constraint programming in fourth year at INSA Rennes.

10 Bibliography

Books and Monographs

[1] P. Cellier, T. Charnois, A. Hotho, S. Matwin, M. Moens, Y. Toussaint (editors), *Proceedings of the 3rd Workshop on Interactions between Data Mining and Natural Language Processing, DMNLP 2016, co-located with the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, ECML-PKDD 2016, Riva del Garda, Italy, September 23, 2016*, *1646*, CEUR-WS.org, 2016, `http://ceur-ws.org/Vol-1646`.

Articles in referred journals and book chapters

[2] G. Aucher, "Displaying Updates in Logic", *Journal of Logic and Computation*, March 2016, `https://hal.inria.fr/hal-01476234`.

[3] G. Aucher, "Dynamic Epistemic Logic in Update Logic", *Journal of Logic and Computation*, March 2016, `https://hal.inria.fr/hal-01476249`.

[4] M. Ba, S. Ferré, M. Ducassé, "Solving Data Mismatches in Bioinformatics Workflows by Generating Data Converters", *Transactions on Large-Scale Data and Knowledge-Centered Systems (TLDKS) LNCS 9510*, 2016, p. 88–115.

[5] 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.

[6] 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, `http://www.semantic-web-journal.net/content/sparklis-expressive-query-builder-sparql-endpoints-guidance-natural-language-1`.

Publications in Conferences and Workshops

[7] P. Cellier, S. Ferré, A. Foret, O. Ridoux, "Exploration des Données du Défi EGC 2016 à l'aide d'un Système d'Information Logique", *in: Journées Francophones Extraction et Gestion des Connaissances, EGC*, C. de Runz, B. Crémilleux (editors), *RNTI E-30*, Hermann-Éditions, p. 443–448, 2016, `http://editions-rnti.fr/?inprocid=1002198`.

[8] 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.

[9] 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.

[10] S. Ferré, "An RDF Design Pattern for the Structural Representation and Querying of Expressions", *in: Int. Conf. Knowledge Engineering and Knowledge Management*, *LNAI 10024*, Springer, 2016, https://hal.inria.fr/hal-01405495.

[11] 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.

[12] C. Gautrais, P. Cellier, T. Guyet, R. Quiniou, A. Termier, "Understanding Customer Attrition at an Individual Level: a New Model in Grocery Retail Context", *in: Proceedings of the 19th International Conference on Extending Database Technology, EDBT 2016, Bordeaux, France, March 15-16, 2016, Bordeaux, France, March 15-16, 2016.*, p. 686–687, 2016.

[13] P. Maillot, S. Ferré, P. Cellier, M. Ducassé, F. Partouche, "FORMULIS: Dynamic Form-Based Interface For Guided Knowledge Graph Authoring", *in: 20th International Conference on Knowledge Engineering and Knowledge Management, Posters & Demonstrations*, 2016.

Selection of publications by the team in previous years

[1] 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.

[2] P. Allard, *Logical modeling of multidimensional analysis of multivalued relations - Application to geographic data exploration*, PhD Thesis, Thèse de l'Université de Rennes 1 - École doctorale MATISSE, 12 décembre 2011, supervised by S. Ferré and O. Ridoux.

[3] G. Aucher, T. Bolander, "Undecidability in Epistemic Planning", *in: IJCAI*, 2013.

[4] G. Aucher, B. Maubert, S. Pinchinat, "Automata Techniques for Epistemic Protocol Synthesis", *in: 2nd International Workshop on Strategic Reasoning*, F. Mogavero, A. Murano, M. Y. Vardi (editors), *EPTCS*, *146*, p. 11, 2014-04-06, France, April 2014, https://hal.inria.fr/hal-01098740.

[5] G. Aucher, "Supervisory Control Theory in Epistemic Temporal Logic", *in: AAMAS 2014*, Paris, France, May 2014, https://hal.inria.fr/hal-01098771.

[6] G. Aucher, "Intricate Axioms as Interaction Axioms", *Studia Logica*, 2015, p. 28, https://hal.inria.fr/hal-01193284.

[7] M. Ba, S. Ferré, M. Ducassé, "Solving Data Mismatches in Bioinformatics Workflows by Generating Data Converters", *Transactions on Large-Scale Data and Knowledge-Centered Systems (TLDKS) LNCS 9510*, 2016, p. 88–115.

[8] N. Béchet, P. Cellier, T. Charnois, B. Crémilleux, "Sequence mining under multiple constraints", *in: ACM Symposium on Applied Computing*, R. L. Wainwright, J. M. Corchado, A. Bechini, J. Hong (editors), ACM, p. 908–914, 2015.

[9] 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.

[10] 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.

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

[12] D. Béchet, A. Foret, "A Pregroup Toolbox for Parsing and Building Grammars of Natural Languages", *Linguistic Analysis Journal 36*, 2010.

[13] 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.

[14] P. Cellier, S. Ferré, M. Ducassé, T. Charnois, "Partial orders and logical concept analysis to explore patterns extracted by data mining", *in: Int. Conf. on Conceptual Structures for Discovering Knowledge*, Springer, p. 77–90, 2011.

[15] 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.

[16] P. Cellier, "Formal concept analysis applied to fault localization", *in: International Conference on Software Engineering (ICSE 2008) Companion*, Robby (editor), ACM, p. 991–994, 2008.

[17] M. Ducassé, P. Cellier, "Fair and Fast Convergence on Islands of Agreement in Multicriteria Group Decision Making by Logical Navigation", *Group Decision and Negotiation 23*, 4, July 2014, p. 673–694, `http://dx.doi.org/10.1007/s10726-013-9372-4`.

[18] M. Ducassé, S. Ferré, "Fair(er) and (almost) serene committee meetings with Logical and Formal Concept Analysis", *in: Proceedings of the International Conference on Conceptual Structures*, P. Eklund, O. Haemmerlé (editors), Springer-Verlag, July 2008. Lecture Notes in Artificial Intelligence 5113.

[19] S. Ferré, "A Proposal for Extending Formal Concept Analysis to Knowledge Graphs", *in: Int. Conf. Formal Concept Analysis (ICFCA)*, J. Baixeries, C. Sacarea, M. Ojeda-Aciego (editors), *LNCS 9113*, Springer, p. 271–286, 2015.

[20] 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.

[21] S. Ferré, R. D. King, "A dichotomic search algorithm for mining and learning in domain-specific logics", *Fundamenta Informaticae – Special Issue on Advances in Mining Graphs, Trees and Sequences 66*, 1-2, 2005, p. 1–32.

[22] S. Ferré, O. Ridoux, "A Logical Generalization of Formal Concept Analysis", *in: International Conference on Conceptual Structures*, G. Mineau, B. Ganter (editors), *Lecture Notes in Computer Science*, 1867, Springer, p. 371–384, August 2000.

[23] S. Ferré, O. Ridoux, "A Framework for Developing Embeddable Customized Logics", *in: Int. Work. Logic-based Program Synthesis and Transformation*, A. Pettorossi (editor), *LNCS 2372*, Springer, p. 191–215, 2002.

[24] 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.

[25] S. Ferré, O. Ridoux, "An Introduction to Logical Information Systems", *Information Processing & Management 40*, 3, 2004, p. 383–419.

[26] S. Ferré, S. Rudolph, "Advocatus Diaboli - Exploratory Enrichment of Ontologies with Negative Constraints", *in: Int. Conf. Knowledge Engineering and Knowledge Management (EKAW)*, A. ten Teije et al. (editor), *LNAI 7603*, Springer, p. 42–56, 2012.

[27] S. Ferré, *Systèmes d'information logiques : un paradigme logico-contextuel pour interroger, naviguer et apprendre*, Thèse d'université, Université de Rennes 1, October 2002, Accessible en ligne à l'adresse http://www.irisa.fr/bibli/publi/theses/theses02.html., `http://www.irisa.fr/bibli/publi/theses/theses02.html`.

[28] S. Ferré, "Camelis: a logical information system to organize and browse a collection of documents", *Int. J. General Systems 38*, 4, 2009.

[29] 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.

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

[31] A. Foret, S. Ferré, "On Categorial Grammars as Logical Information Systems", *in: Int. Conf. Formal Concept Analysis*, L. Kwuida, B. Sertkaya (editors), *LNCS 5986*, Springer, p. 225–240, 2010.

[32] A. Foret, "A modular and parameterized presentation of pregroup calculus", *Information and Computation Journal 208*, 5, may 2010, p. 395–604.

[33] A. Hermann, S. Ferré, M. Ducassé, "An Interactive Guidance Process Supporting Consistent Updates of RDFS Graphs", *in: Int. Conf. Knowledge Engineering and Knowledge Management (EKAW)*, A. ten Teije et al. (editor), *LNAI 7603*, Springer, p. 185–199, 2012.

[34] Y. Padioleau, O. Ridoux, "A Logic File System", *in: Proc. USENIX Annual Technical Conference*, 2003.

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