RESEARCH CENTRE

Lille - Nord Europe

IN PARTNERSHIP WITH:

CNRS, Université de Lille

2020
ACTIVITY REPORT

Project-Team LINKS

Linking Dynamic Data

IN COLLABORATION WITH: Centre de Recherche en Informatique, Signal et Automatique de Lille

DOMAIN

Perception, Cognition and Interaction

THEME

Data and Knowledge Representation and Processing

Contents

Project-Team LINKS

Creation of the Team: 2013 January 01, updated into Project-Team: 2016 June 01

Keywords

Computer Sciences and Digital Sciences:

A2.1. – Programming Languages

A2.1.1. – Semantics of programming languages

A2.1.4. – Functional programming

A2.1.6. – Concurrent programming

A2.4. – Formal method for verification, reliability, certification

A2.4.1. – Analysis

A2.4.2. – Model-checking

A2.4.3. – Proofs

A3.1. – Data

A3.1.1. – Modeling, representation

A3.1.2. – Data management, quering and storage

A3.1.3. – Distributed data

A3.1.4. – Uncertain data

A3.1.5. – Control access, privacy

A3.1.6. – Query optimization

A3.1.7. – Open data

A3.1.8. – Big data (production, storage, transfer)

A3.1.9. – Database

A3.2.1. – Knowledge bases

A3.2.2. – Knowledge extraction, cleaning

A3.2.3. – Inference

A3.2.4. – Semantic Web

A4.7. – Access control

A4.8. – Privacy-enhancing technologies

A7. – Theory of computation

A7.2. – Logic in Computer Science

A9.1. – Knowledge

A9.2. – Machine learning

A9.7. – AI algorithmics

A9.8. – Reasoning

Other Research Topics and Application Domains:

B6.1. – Software industry

B6.3.1. – Web

B6.3.4. – Social Networks

B6.5. – Information systems

B9.5.1. – Computer science

B9.5.6. – Data science

B9.10. – Privacy

1 Team members, visitors, external collaborators

Research Scientists

- Joachim Niehren [Team leader, Inria, Senior Researcher, HDR]
- Mikael Monet [Inria, Researcher, from Oct 2020]

Faculty Members

- Iovka Boneva [Université de Lille, Associate Professor]
- Florent Capelli [Université de Lille, Associate Professor]
- Aurélien Lemay [Université de Lille, Associate Professor, HDR]
- Charles Paperman [Université de Lille, Associate Professor]
- Sylvain Salvati [Université de Lille, Professor, HDR]
- Slawomir Staworko [Université de Lille, Associate Professor, HDR]
- Sophie Tison [Université de Lille, Professor, HDR]

PhD Students

- Antonio Al Serhali [Inria, from Oct 2020]
- Nicolas Crosetti [Inria]
- Paul Gallot [Inria]
- Jose Martin Lozano [Université de Lille]
- Momar Ndiouga Sakho [Université de Lille, until Aug 2020]
- Claire Soyez-Martin [Inria, from Sep 2020]

Technical Staff

- Antonio Al Serhali [Inria, Engineer, until Sep 2020]
- Cherif Amadou Ba [Inria, Engineer, from Sep 2020]
- Momar Ndiouga Sakho [Inria, Engineer, from Nov 2020]

Interns and Apprentices

- Corentin Barloy [École Normale Supérieure de Paris, from Oct 2020]
- Leo Beuque [École centrale de Lille, until Feb 2020]
- Aymeric Come [Inria, from Jul 2020 until Aug 2020]
- Amine Laabi [Université de Lille, from Jul 2020 until Aug 2020]
- Claire Soyez-Martin [Université de Lille, from Feb 2020 until Jul 2020]

Administrative Assistant

- Nathalie Bonte [Inria]

Visiting Scientists

- Corentin Barloy [École Normale Supérieure de Paris, from Aug 2020 until Sep 2020]
- Aymeric Come [École centrale de Lille, from Aug 2020 until Sep 2020]
- Momar Ndiouga Sakho [Inria, from Sep 2020 until Oct 2020]
- Claire Soyez-Martin [Université de Lille, from Jul 2020 until Aug 2020]

2 Overall objectives

We will develop algorithms for answering logical querying on heterogeneous linked data collections in hybrid formats, distributed programming languages for managing dynamic linked data collections and workflows based on queries and mappings, and symbolic machine learning algorithms that can link datasets by inferring appropriate queries and mappings.

2.1 Presentation

The following three paragraphs summarize our main research objectives.

Querying Heterogeneous Linked Data We develop new kinds of schema mappings for semi-structured datasets in hybrid formats including graph databases, RDF collections, and relational databases. These induce recursive queries on linked data collections for which we will investigate evaluation algorithms, containment problems, and concrete applications.

Managing Dynamic Linked Data In order to manage dynamic linked data collections and workflows, we will develop distributed data-centric programming languages with streams and parallelism, based on novel algorithms for incremental query answering, study the propagation of updates of dynamic data through schema mappings, and investigate static analysis methods for linked data workflows.

Linking Data Graphs Finally, we will develop symbolic machine learning algorithms, for inferring queries and mappings between linked data collections in various graphs formats from annotated examples.

3 Research program

3.1 Background

The main objective of LINKS is to develop methods for querying and managing linked data collections. Even though open linked data is the most prominent example, we will focus on hybrid linked data collections, which are collections of semi-structured datasets in hybrid formats: graph-based, RDF, relational, and NOSQL. The elements of these datasets may be linked, either by pointers or by additional relations between the elements of the different datasets, for instance the “same-as” or “member-of” relations as in RDF.

The advantage of traditional data models is that there exist powerful querying methods and technologies that one might want to preserve. In particular, they come with powerful schemas that constraint the possible manners in which knowledge is represented to a finite number of patterns. The exhaustiveness of these patterns is essential for writing of queries that cover all possible cases. Pattern violations are excluded by schema validation. In contrast, RDF schema languages such as RDFS can only enrich the relations of a dataset by new relations, which also helps for query writing, but which cannot constraint the number of possible patterns, so that they do not come with any reasonable notion of schema validation.

The main weakness of traditional formats, however, is that they do not scale to large data collections as stored on the Web, while the RDF data models scales well to very big collections such as linked open data. Therefore, our objective is to study mixed data collections, some of which may be in RDF format, in which we can lift the advantages of smaller datasets in traditional formats to much larger linked data collections. Such data collections are typically distributed over the internet, where data sources may have rigid query facilities that cannot be easily adapted or extended.

The main assumption that we impose in order to enable the logical approach, is that the given linked data collection must be correct in most dimensions. This means that all datasets are well-formed with respect to their available constraints and schemas, and clean with respect to the data values in most of the components of the relations in the datasets. One of the challenges is to integrate good quality RDF datasets into this setting, another is to clean the incorrect data in those dimensions that are less proper. It remains to be investigated in how far these assumptions can be maintained in realistic applications, and how much they can be weakened otherwise.

For querying linked data collections, the main problems are to resolve the heterogeneity of data formats and schemas, to understand the efficiency and expressiveness of recursive queries, that can follow links repeatedly, to answer queries under constraints, and to optimize query answering algorithms based on static analysis. When linked data is dynamically created, exchanged, or updated, the problems are how to process linked data incrementally, and how to manage linked data collections that change dynamically. In any case (static and dynamic) one needs to find appropriate schema mappings for linking semi-structured datasets. We will study how to automatize parts of this search process by developing symbolic machine learning techniques for linked data collections.

3.2 Research axis: Querying Data Graphs

Linked data is often abstracted as datagraphs: nodes carry information and edges are labeled. Internet, the semantic web, open data, social networks and their connections, information streams such as twitter are examples of such datagraphs. An axis of Links is to propose methods and tools so as to extract information from datagraphs. We dwell in a wide spectrum of tools to construct these methods: circuits, compilation, optimization, logic, automata, machine learning. Our goal is to extend the kinds of information that can be extracted from datagraphs while improving the efficiency of existing ones.

This axis is split within two themes. The first one pertains to the use of lwo level representation by means of circuits to compute efficiently complex numerical aggregates that will find natural applications in AI. The second one proposes to explore path oriented query language and more particularly their efficient evaluation by means of efficient compilation and machine learning methods so as to have manageable statistics.

3.2.1 AI: Circuits for Data Analysis

Circuits are consice representations of data sets that recently found a unifying interest in various areas of artificial intelligence. A circuit may for instance represent the answer set of a database query as a dag whose operators are disjoint unions (for disjunction) and cartesian products (for conjunction). Similarly, it may also represent the set of all matches of a pattern in a graph. The structure of the circuit may give rise to efficient algorithms to process large data sets based on representation that are often much smaller. Among others, such applications range from knowledge representation/compilation, counting the number of solutions of queries, efficient query answering, factorized databases.

In a first line of research, we want to study novel problems on circuits, in which database queries are relevant to data analysis tasks from artificial intelligence, in machine learning or data mining in particular. In particular we propose to study optimiation problems on answer sets of database queries based on circuits, i.e., how to find obtimal solutions in the answer set for a given set of conditions. Decompressing small circuits into large answer sets would make the optimization problem unfeasible in many cases. We believe that circuits can structure certain optimization problems in such a way that it can be phrased concisely and then solved efficiently.

Second, we propose to develop a tighter integration between circuits and databases. Indeed query-related circuits are generally produced from a database. This requires that the data is copied within the circuits. This memory cost is accompanied with the loss of the environment of the DBMS which allows many optimization and uses many low level optimizations that are hard to implement. We propose then to encode circuits directly within the database using materialized views and index structures. We shall also develop the required computational tools for maintaining and exploiting these embedded circuits.

3.2.2 Path Query Optimization

Graph databases are easily queries using path descriptions. Most often these paths are described by means of regular expressions. This makes path queries difficult as the use of Kleene star makes them recursive. In relational DBMS, recursion is almost never used and it is not advised to use it. The natural theoretical tool that pertains to recursion in the context of relational data Datalog. There has been a wealth of optimization algorithms that have been proposed for queries written in Datalog. We propose to use Datalog as a low level language to which we will compile path queries of various kinds. The idea is that the compiler will try to obtain Datalog programs that will have low execution complexity taking advantages of optimization techniques such as magic supplementary set rewriting, pre-computed indexes and also statistics computed from the graph. Our goal is to develop a compiler that will be able to efficiently evaluates path queries on large graphs which in particular will explore only a part of it.

3.3 Research axis: Monitoring Data Graphs

Traditional database applications are programs that interact with database via updates and queries. We are interested in developing programming language techniques so as to interact with datagraphs rather than with traditional relational databases. Moreover, we shall take into account the dynamic aspects of datagraphs which shall evolve through updates. The methods we shall develop will monitor changes in datagraphs and react according to the modifications.

3.3.1 Functional Programming Languges for Data Graphs

The first question is which kind of programming language to use to enable monitoring processes for data graphs based on query answering. While languages of path queries found quite some interest on data graphs, less attention has been given to the programming language tasks, that needed to be solved to produce structured output and to compose various queries with structured output into a pipeline. We believe that transferring the generalization of ideas developed for data trees in the context of XML to data graphs will allow to solve such problems in a systematic manner.

Our approach will consist in developing a functionl programming language based on first principles (the lambda calculus, graph navigation, logical connective) that generalizes full XPath 3.0 to the context of graphs. Here we can rely on own previous work for data trees, such as the language X-Fun and λ-XP. After the language for data graphs is designed we shall study its behavior empirically by means of an implementation. This implementation will help us to design optimization methods so as to evaluate the queries in that language. We think that in the context of querying functional programs play a central which means that the query language will not allow side effects when computing. This will allow us to use a wealth of techniques so as to optimize the computation. Indeed, we can try to compile data structures to imperative ones when possible and also exploit possibilities of parallel executions in certain cases. Functional programming comes with nice verification techniques that we are going to use in several contexts: 1. in optimizing queries (e.g. stop the evaluation when it is possible to know that no more data can contribute to the output) 2. verify that the query behaves correctly. The verification methods we shall focus on will be mainly related to automata and transducers.

Finally we shall also develop a programming language that allows to describe services that use datagraphs as a backend for storing data. Here again, functional programming seems a good candidate, we would need however to orchestrate the concurrent executions of queries so as to ensure the correct behavior of services. This means that we should have concurrent constructs that are built in the language. The high level of concurrence enabled by the notion of *futures* seems an interesting candidate to adapt to the context of service orchestration.

3.3.2 Hyperstreaming Program Evaluation

Complex-event processing requires to monitor data graphs that are produced on input streams and to write data graphs to some output stream, which can then be used as inputs again. A major problem here is to reduce the high risk of blocking, which arises when the writing of some of the output stream suspends on a data value that will become available only in the future on some input stream. In such cases, all monitoring processes reading the output stream may have to suspend as well. In order to reduce the risc of blocking, we propose to develop the hyperstreaming approach further, of which we layed the foundations in the evaluation period based on automata techniques. The idea is to generalize streams to hyperstreams, i.e. to add holes to streams that can be filled by some other stream in the future. In order to avoid suspension as possible, a monitoring procesess on hyperstream must then be able to jump over the holes, and to perform some speculative compuation. The objective for the next period are to develop tools for hyperstreaming query answering and to lift these to hyperstreaming program evaluation. Furthermore, on the conceptual side, the notion of certain query answers on hyperstreams needs to be lifted to certain program outputs on hyperstreams.

3.4 Research axis: Graph Data Integration

We intend to continue to develop tools for integration of linked data with RDF being their principal format. Because from its conception the main credo of RDF has been "just publish your data," the problem at hand faces two important challenges: data quality and data heterogeneity.

3.4.1 Data Quality with Schemas and Repairing with Inference

The data quality of RDF may suffer due to a number of reasons. Impurities may arise due to data value errors (misspellings, errors during data entry etc.). Such data quality problems have been thoroughly investigated in literature for relational databases and solutions include dictionary methods,... However, it remains to be seen if the challenges of adapting the existing solutions for relational databases can be easily addressed.

One particular challenge comes from the fact that RDF allows a higher degree of structural freedom in how information is represented as opposed to relation databases, where the choice is strongly limited to flat tables. We plan to investigate suitability of existing data cleaning methods to tackle the problems of data value impurities in RDF. The structural freedom of RDF is a source of data quality issues on its own. With the recent emergence of schema formalisms for RDF, it becomes evident that significant parts of existing RDF repositories do not necessarily satisfy schemas prepared by domain experts.

In the first place, we intend to investigate defining suitable measures of quality for RDF documents. Our approaches will be based on a schema language, such as ShEx and SHACL, and we shall explore suitable variants of graph alignment and graph edit distance to capture similarity between the existing RDF document and its possible repaired versions that satisfy the schema.

The central issue here is repairing an RDF document w.r.t. schema by identifying essential fragments of the RDF that fail to satisfy the schema. Once such fragments are identified, repairing actions can be applied however there might be a significant number of alternatives. We intend to explore enumeration approaches where the space of repairing alternatives is intelligently browsed by the user and the most suitable one chosen. Furthermore, we intend to propose a rule language for choosing the most suitable repairing action and will investigate inference methods to derive from interactions with user the optimal order in which various repairing actions are presented to the user and derive the rules for the choice of the preferred repairing action for repeating types of fragments that do not satisfy the schema.

3.4.2 Integration and Graph Mappings with Schemas and Inference

The second problem pertaining to integration of RDF data sources is their heterogeneity. We intend to continue to identify and study suitable classes of mappings between RDF documents conforming to potentially different and complementary schemas. We intend to assist the user in constructing such mappings by developing rich and expressive graphical languages for mappings. Also, we wish to investigate inference of RDF mappings with the active help of an expert user. We will need to define interactive protocols that allows the input to be sufficiently informative to guide the inference process while avoiding the pitfalls of user input being too ambiguous and causing combinatorial explosion. We intend to identify

RDF Data Quality. Approach based on a schema language (ShEx or SHACL) used to identify errors and giving a notion of a measure of quality of an RDF database. Impurities in RDF may come from data value errors (misspellings etc.) but also from the fact that RDF imposes fewer constraints on how data is structured which is a consequence of a significantly different use philosophy (just publish your data anyway you want). Repairing of RDF errors would be modeled with a localized rules (transformations that operate within a small radius of an affected node) and if several rules apply, preferences are used to identify the most desirable one. Both the repairing rules and preferences can be inferred with the help of inference algorithms in an interactive setting. Smart tools for LOD integration. Assuming that the LOD sources are of good quality, we want to build tools that assist the user in constructing mappings that integrate data in the user database. For this, we want to define inference algorithms which are guided by schemas, and which are based on comprehensible interactions with the user. For this, we need to define interactions that are rich enough to inform the algorithm, while simple enough to be understandable by a non-expert user. In particular, that means that we need to present data (nodes in a graph for instance) in a readable way. Also, we want to investigate how the - possibly inferred - schema can be used to guide the inference.

4 Application domains

4.1 Linked data integration

There are many contexts in which integrating linked data is interesting. We advocate here one possible scenario, namely that of integrating business linked data to feed what is called Business Intelligence. The latter consists of a set of theories and methodologies that transform raw data into meaningful and useful information for business purposes (from Wikipedia). In the past decade, most of the enterprise data was proprietary, thus residing within the enterprise repository, along with the knowledge derived from that data. Today's enterprises and businessmen need to face the problem of information explosion, due to the Internet's ability to rapidly convey large amounts of information throughout the world via end-user applications and tools. Although linked data collections exist by bridging the gap between enterprise data and external resources, they are not sufficient to support the various tasks of Business Intelligence. To make a concrete example, concepts in an enterprise repository need to be matched with concepts in Wikipedia and this can be done via pointers or equalities. However, more complex logical statements (i.e. mappings) need to be conceived to map a portion of a local database to a portion of an RDF graph, such as a subgraph in Wikipedia or in a social network, e.g. LinkedIn. Such mappings would then enrich the amount of knowledge shared within the enterprise and let more complex queries be evaluated. As an example, businessmen with the aid of business intelligence tools need to make complex sentimental analysis on the potential clients and for such a reason, such tools must be able to pose complex queries, that exploit the previous logical mappings to guide their analysis. Moreover, the external resources may be rapidly evolving thus leading to revisit the current state of business intelligence within the enterprise.

4.2 Data cleaning

The second example of application of our proposal concerns scientists who want to quickly inspect relevant literature and datasets. In such a case, local knowledge that comes from a local repository of publications belonging to a research institute (e.g. HAL) need to be integrated with other Web-based repositories, such as DBLP, Google Scholar, ResearchGate and even Wikipedia. Indeed, the local repository may be incomplete or contain semantic ambiguities, such as mistaken or missing conference venues, mistaken long names for the publication venues and journals, missing explanation of research keywords, and opaque keywords. We envision a publication management system that exploits both links between database elements, namely pointers to external resources and logical links. The latter can be complex relationships between local portions of data and remote resources, encoded as schema mappings. There are different tasks that such a scenario could entail such as (i) cleaning the errors with links to correct data e.g. via mappings from HAL to DBLP for the publications errors, and via mappings from HAL to Wikipedia for opaque keywords, (ii) thoroughly enrich the list of publications of a given research institute, and (iii) support complex queries on the corrected data combined with logical mappings.

4.3 Real-time complex event processing

Complex event processing serves for monitoring nested word streams in real time. Complex event streams are gaining popularity with social networks such as with Facebook and Twitter, and thus should be supported by distributed databases on the Web. Since this is not yet the case, there remains much space for future industrial transfer related to Links' second axis on dynamic linked data.

5 Social and environmental responsibility

5.1 Footprint of research activities

Sophie Tison elected member of conseil de l'EATCS (European Association for Theoretical Science)

5.2 Impact of research results

Databases and methods from Artificial Intelligence are used in mostly all web services.

6 Highlights of the year

All recently hired permanent members of Links published papers in major conferences on database theory, artificial intelligence, and computer science theory:

PODS 2021: Principles of Database Systems. Paper accepted by **Corentin Barloy, Charles Paperman**, et. al.] Stackless Processing of Streamed Trees [19]. The top most database theory conference. Corentin is starting his PhD project in LINKS with Charles.

AAAI 2021 Conference on Artificial Intelligence. Two papers accepted.

Florent Capelli et al. Certifying Top-Down Decision-DNNF Compilers [22]. Cooperation avec Pierre Marquis de Lens. Furthermroe, Florent got a ANR project accepted on related topics.

Mikaël Monet et al. The Tractability of SHAP-Score-Based Explanations over <Deterministic and Decomposable Boolean Circuits [17]. Furthermore, Mikael was hired as CRNC Inria this year.

MFCS 2020: Mathematical Foundations of Computer Science. Paper published by **Paul Gallot, Aurélien Lemay, and Sylvain Salvati**. Linear high-order deterministic tree transducers with regular look-ahead [23]. This contribution will be part of Paul's PhD thesis.

7 New software and platforms

7.1 New software

7.1.1 ShEx validator

Name: Validation of Shape Expression schemas

Keywords: Data management, RDF

Functional Description: Shape Expression schemas is a formalism for defining constraints on RDF graphs. This software allows to check whether a graph satisfies a Shape Expressions schema.

Release Contributions: ShExJava now uses the Commons RDF API and so support RDF4J, Jena, JSON-LD-Java, OWL API and Apache Clerezza. It can parse ShEx schema in the ShEcC, ShEJ, ShExR formats and can serialize a schema in ShExJ.

To validate data against a ShExSchema using ShExJava, you have two different algorithms: - the refine algorithm: compute once and for all the typing for the whole graph - the recursive algorithm: compute only the typing required to answer a validate(node,ShapeLabel) call and forget the results.

URL: `http://shexjava.lille.inria.fr/`

Contacts: Iovka Boneva, Jeremie Dusart

7.1.2 gMark

Name: gMark: schema-driven graph and query generation

Keywords: Semantic Web, Data base

Functional Description: gMark allow the generation of graph databases and an associated set of query from a schema of the graph.gMark is based on the following principles: - great flexibility in the schema definition - ability to generate big size graphs - ability to generate recursive queries - ability to generate queries with a desired selectivity

URL: https://github.com/graphMark/gmark

Contact: Aurélien Lemay

7.1.3 SmartHal

Keyword: Bibliography

Functional Description: SmartHal is a better tool for querying the HAL bibliography database, while is based on Haltool queries. The idea is that a Haltool query returns an XML document that can be queried further. In order to do so, SmartHal provides a new query language. Its queries are conjunctions of Haltool queries (for a list of laboratories or authors) with expressive Boolean queries by which answers of Haltool queries can be refined. These Boolean refinement queries are automatically translated to XQuery and executed by Saxon. A java application for extraction from the command line is available. On top of this, we have build a tool for producing the citation lists for the evaluation report of the LIFL, which can be easily adapter to other Labs.

URL: http://smarthal.lille.inria.fr/

Contact: Joachim Niehren

7.1.4 QuiXPath

Keywords: XML, NoSQL, Data stream

Scientific Description: The QuiXPath tools supports a very large fragment of XPath 3.0. The QuiXPath library provides a compiler from QuiXPath to FXP, which is a library for querying XML streams with a fragment of temporal logic.

Functional Description: QuiXPath is a streaming implementation of XPath 3.0. It can query large XML files without loading the entire file in main memory, while selecting nodes as early as possible.

URL: https://project.inria.fr/quix-tool-suite/

Contact: Joachim Niehren

7.1.5 X-FUN

Keywords: Programming language, Compilers, Functional programming, Transformation, XML

Functional Description: X-FUN is a core language for implementing various XML, standards in a uniform manner. X-Fun is a higher-order functional programming language for transforming data trees based on node selection queries.

Authors: Joachim Niehren, Pavel Labath

Contact: Joachim Niehren

Participants: Joachim Niehren, Pavel Labath

7.1.6 ShapeDesigner

Name: ShapeDesigner

Keywords: Validation, Data Exploration, Verification

Functional Description: ShapeDesigner allows construct a ShEx or SHACL schema for an existing dataset. It combines algorithms to analyse the data and automatically extract shape constraints, and to edit and validate shape schemas.

URL: https://gitlab.inria.fr/jdusart/shexjapp

Contacts: Jeremie Dusart, Iovka Boneva

7.2 New platforms

8 New results

8.1 Querying Data Graphs

8.1.1 Circuits for Data Analysis in Artificial Intelligence

Knowledge compilation to Boolean circuits is a general technique in artificial intelligence to obtain tractable algorithms for subclasses of algorithmic problems that are computationally hard. For instance, a variant of Yannakakis' algorithm can be used to compile acyclic conjunctive database queries to Boolean circuits. These will then be decomposable and deterministic, and thus tractable in polynomial time, while for the general class of conjunctive queries, testing the existence of a query answer on a relational database is coNP-complete. Another class of instances, where knowledge complilation is used in AI, concern satisfiablity problems. Beside of satisfiability, knowledge compilation is equally relevant to aggregation and enumeration problems.

In their article in Theory of Computing Systems [11], Capelli, Monet et al. present a systematic picture connecting Boolean circuits to width measures through upper and lower complexity bounds. This is joined work with the Inria Valda team. In particular, their upper bounds show that bounded-treewidth circuits can be constructively converted to special circuits known as d-SDNNFs, in time linear in the circuit size and singly exponential in the treewidth. A much more general survey of complexity question in artificial intelligence is given in a book chapter by Tison et al. [26].

Capelli et al. present at AAAI [22] a method to certify the output knowledge compilers and #SAT-solvers. This is a cooperation with the University of Lens. The idea is to output a certificate that can be checked in polynomial time and can be used to certify that a given CNF formula has K models. Their experiments were encouraging showing that a large majority of CNF formulas for which the #SAT-solver D4 terminates have certificates that can be checked more quickly than the compilation time.

In their article in Discrete Applied Mathematics, Capelli et. al [14] study the problem of faster enumerating models of models of DNF formulas. The aim is to provide enumeration algorithms with a delay that depends polynomially on the size of each model and not on the size of the formula. In particular, they provide a constant delay algorithm for k-DNF formulas with fixed k.

8.1.2 Uncertainty and Explanations

Monet et al. [18] propose in a paper at NeurIPS a new formalization of the interpretability of classes of models of machine learning algorithms based on computational complexity theory. This work is done in cooperation with the Universidad de Chile. They can prove in their framework that shallow neural networks are more interpretable than deeper neural networks.

Monet et al. [17] study in a paper at AAAI Shapely values for providing explanations to classification results over machine learning models. This work is also done in cooperation with Chile, but now with the Universidad Catolica. While in general computing Shapley values is a computationally intractable problem, it has recently been claimed that the SHAP-score can be computed in polynomial time over the class of decision trees. They show that the SHAP-score can be computed in polynomial time over deterministic and decomposable Boolean circuits.

8.1.3 Path Query Optimization

Niehren, Salvati et al. [24] propose a new algorithm for answering nested regular path queries on data graphs efficiently. Previous jumping algorithms were limited to data trees, while the new jumping evaluator can be applied to data graphs. This generalization is obtained by a novel compilation scheme of path queries to datalog programs.

8.2 Monitoring Data Graphs

8.2.1 Functional Programming Langugues for Data Trees

Gallot, Lemay, and Salvati [23] introduced high-order deterministic tree transducers at the 45th International Symposium on Mathematical Foundations of Computer Science (MFCS). This is a natural generalization of known models top-down tree transductions including macro tree transuders and streaming tree transducers. They show that the class of linear high-order tree transducers with look-ahead captures the functional tree-to-tree transformations definable in monadic second-order logic. They also give a specialized procedure for the composition of those transducers that preserves linearity.

Paperman et al. [13] present an article at Logical Methods in Computer Science, in which they study the continuity of functional transducers on words. This is an international cooperation with Chicago and Paris.

8.2.2 Query Answering on Streams

Complex event processing requires to answer queries on streams of complex events, i.e., nested words or, equivalently, linearizations of data trees, but also to produce dynamically evolving data structures as output.

Niehren and Boneva supervised the PhD thesis of Sakho [28] on certain query answering on hyperstreams. They studied the complexity of hyperstreaming query evaluation in a articel published at Information and computation [12]. While it is generally in EXP, the complexity goes down to P-time when representing queries by deterministic automata on nested words, and restricting hypestreams to be linear.

In an article published at Algorithms [16] extending on a paper published at CSR [20], they could show that regular path queries on XML documents in the usual XPathMark benchmark can be compiled to reasonably small deterministic automata on nested words. For this they propose new compilers to the novel class of deterministic stepwise hedge automata and proposed a minimization algorithm for them. We note that streaming evaluators for such automata are heavily stack based.

Paperman with his future PhD student Barloy study stackless stream processing for nested words in a cooperation with the University of Warsaw [19]. They characterize in a paper accepted at the International Conference of Foundations of Database Systems (PODS) the subclass of regular path queries that can be evaluated stacklessly - with and without registers.

8.3 Graph Data Integration

Staworko and Boneva supervised the PhD thesis of Lozano [27] on data exchange from relational database to RDF graphs subject to shape schemas in ShEx. In [25] they show that the consistency problem is coNP-complete, i.e. checking whether every source instance of the relational database admits a target solution, i.e., a RDF graph that satisfies the source-to-target dependencies. They also study the problem of certain query answering, of finding answer of any target solution. For this they introduce the notion of universal simulation solution that allows to compute certain query answers for forwards path queries.

In a cooperation with the Unversity of Oviedo in Spain, Boneva and Staworko conducted a usability experiment on three different graph schema languages for heterogeneous data mapping [15]. Their results show that users of our own language ShExML tend to perform better than those of YARRRML and SPARQL-Generate.

8.4 Others

Paperman et at. [13] published a paper on polynomial recursive sequence at the 47th International Colloquium on Automata, Languages and Programming (ICALP). For researching this results, Paperman invited the 4 other authors for a 5 day working meeting in Lille. ICALP is one of the major conferences in theoretical computer science, so this result could be marked as another highlight of the year.

9 Bilateral contracts and grants with industry

9.1 Bilateral contracts with industry

Stawarko Academic member of Linked Data Benchmark Council (LDBC).

Staworko Member of Work Group on Property Graph Schemas (standardisation effort).

Tison Vice présidente de l'association Force Awards.

10 Partnerships and cooperations

10.1 International Initiatives

Declared Inria international partners

Saint Petersburg, Russia Salvati and Niehren cooperate with the University of Saint Petersburg following a visit of R. Azimov leading to a comon publication at BDA'2020 [24]. This cooperation was funded by a invitation for R. Azimov by the Cristal lab in 2019.

Informal international partners

Santiago, Chile Monet cooperates with Marcelo Arenas and Pablo Barceló from Pontificia Universidad Católica de Chile and with Luca Bertossi from Universidad Adolfo Ibanez (also Chile) on counting problems for incomplete databases and on the computation of SHAP-score explainations for circuit classes from knowledge compilation. This yield joint publications at NeuIPS'2020 [18] and AAAI'2021 [17].

Warsaw, Poland Paperman cooperates with Filip Murlak on query evaluation on streams. A joint paper is accepted for publication at PODS'2021 [19].

Wroclaw, Poland S. Staworko has regular exchange with Piotr Wieczorek from the University of Wroclaw, which lead to a joined publication at PODS 2019.

Tel Aviv, Israel Monet also has regular exchanges with Benny Kimelfeld from Technion (Israel) and Daniel Deutch from Tel Aviv University on computing Shapley values for database query answers.

10.2 International research visitors

10.2.1 Visits of international scientists

Rustam Azimov Saint Petersburg State University, 3 months visit Oct-Dec. Funded by the French-Russian Ambassady. Cancelled for Corona.

Nofar Cameli Technion, Israel. Links' online seminar. Dec 14, 2020.

Alexandre Vigny Bremen University, Germany. Links' online seminar. Dec 10, 2020.

Pierre Pradic Oxford University, England. Links online seminar. Dec 4. 2020.

10.2.2 Sabbatical programme

Florent Capelli Delegation Inria, 2019-2020.

Slawek Staworko Demi delegation Inria, 2020-21.

10.3 European Initiatives

10.4 National initiatives

- **ANR JCJC KCODA** (2021-25):

Participants Florent Capelli *(correspondent)*, Charles Paperman, Sylvain Salvati.

Le but de KCODA est d'étudier comment des représentations succinctes peuvent être utilisées pour résoudre efficacement des problèmes d'optimisation et d'IA modernes qui utilisent beaucoup de données. Nous proposons d'utiliser des structures de données provenant du domaine de la compilation de connaissances qui permettent de représenter de gros jeux de données succinctement en factorisant certaines parties tout en permettant une analyse efficaces des données représentées. Le premier but de KCODA est de comprendre comment on peut résoudre efficacement des problèmes d'optimisation et d'apprentissage pour des données représentées par ces structures. Le second but de KCODA est d'offrir une meilleure intégration de ces techniques dans les systèmes de gestion de bases de données en proposant de nouveaux algorithmes permettant de construire des représentations factorisées des données des réponses d'une requête de BD et en proposant des encodages de ces représentations à l'intérieur de la BD.

- **ANR Colis** (2015-21): Correctness of Linux Scripts.

Participants Joachim Niehren *(correspondent)*, Aurélien Lemay, Paul Gallot, Sylvain Salvati.

The coordinator is R. Treinen from the Université Paris 7 and the other partner is the Tocata project of Inria Saclay (C. Marché).

Objective: This project aims at verifying the correctness of transformations on data trees defined by shell scripts for Linux software installation. The data trees here are the instance of the file system which are changed by installation scripts.

- **ANR DataCert** (2015-21):

Participants Iovka Boneva *(correspondent)*, Sophie Tison, Jose Martin Lozano.

Partners: The coordinator is E. Contejean from the Université Paris-Sud and the other partner is the Université de Lyon.

Objective: the main goals of the Datacert project are to provide deep specification in Coq of algorithms for data integration and exchange and of algorithms for enforcing security policies, as well as to design data integration methods for data models beyond the relational data model.

- **ANR Headwork** (2016-22):

Participants Joachim Niehren *(correspondent)*, Momar Sakho, Nicolas Crosetti, Florent Capelli.

Scientific partners: The coordinateur is D. Gross-Amblard from the Druid Team (Rennes 1). Other partners include the Dahu team (Inria Saclay) and Sumo (Inria Bretagne).

Industrial partners: Spipoll, and Foulefactory.

Objective: The main object is to develop data-centric workflows for programming crowd sourcing systems in flexible declarative manner. The problem of crowd sourcing systems is to fill a database with knowledge gathered by thousands or more human participants. A particular focus is to be put on the aspects of data uncertainty and for the representation of user expertise.

- **ANR Delta** (2016-21):

Participants Joachim Niehren *(correspondent)*, Sylvain Salvati, Aurélien Lemay.

Partners: The coordinator is M. Zeitoun from LaBRI, other partners are LIF (Marseille) and IRIF (Paris-Diderot).

Objective: Delta is focused on the study of logic, transducers and automata. In particular, it aims at extending classical framework to handle input/output, quantities and data.

- **ANR Bravas** (2017-22):

Participants Sylvain Salvati *(correspondent)*.

Scientific Partners: The coordinator is Jérôme Leroux from LaBRI, Université de Bordeaux. The other partner is LSV, ENS Cachan.

Objective: The goal of the BraVAS project is to develop a new and powerful approach to decide the reachability problems for Vector Addition Systems (VAS) extensions and to analyze their complexity. The ambition here is to crack with a single hammer (ideals over well-orders) several long-lasting open problems that have all been identified as a barrier in different areas, but that are in fact closely related when seen as reachability.

10.5 Regional initiatives

Dynamic Semantic Crossords, a project of CPER Data (2020-21):

Participants Joachim Niehren *(correspondent)*, Cherif Ba.

Objective: The objective is to integrate streaming algorithms into the Links' demonstrator of dynamic semantic networks.

Knwledge Compilation, a cooperation with Lens, CPER Data (2020-21)

Participants Florent Capelli *(correspondent)*.

F.Capelli cooperates on knowledge compilation with J.-M.Lagniez et P.Marquis. A joined paper got published at AAAI'2021 [22]. This cooperation is partially funded by the CPER Data.

CPER Cornelia on Artificial Intelligence (2021-2025)

Participants Joachim Niehren *(correspondent)*.

The whole Links' project is partner of this new CPER project.

PhD project Nicolas Crosetti (2018-...) Cofunded by the Region Haut de France. In coopertion with Jan Ramon from Inria Magnet.

Participants Sophie Tison *(supervisor)*, Florent Capelli, Joachim Niehren.

11 Dissemination

11.1 Promoting Scientific Activities

11.1.1 Scientific Events: Organisation

Capelli co-organisation of working group Alga (Automata, Logic, Games & Algebra) of the GDR IM of the CNRS

Capelli co-organisateur of Working group IMIA (Informatique Mathématique Intelligence Artificielle) of the GDR IM of the CNRS.

Member of the Organizing Committees

Capelli Summer School and Workshop Kocoon on Knowledge Compilation. Organised with Marquis and Mengel from Lens. Cancelled for Corona. More info at: kocoon.gforge.inria.fr/

11.1.2 Scientific Events: Selection

Member of the Conference Program Committees

Capelli Program Committee of the 35ieme AAAI Conference on Artificial Intelligence (AAAI'21)

Capelli Program Committee of the International Joint Conference on Artificial Intelligence (IJCAI'21).

Capelli Program Committee of SAT'20.

Monet Program Committee of the 35ieme AAAI Conference on Artificial Intelligence (AAAI'21)

Niehren Program Committee of the 8th International Conference on Computational Methods in Systems Biology (CMSB 2020).

Staworko Program Committee of the 23th International Conference on Extending Database Technology (EDBT 2020)

11.1.3 Journal

Member of the Editorial Boards

Niehren Editorial Board de Fundamenta Informaticae

Tison Editorial Board de RAIRO-ITA

Salvati Managing Editor of the Journal JLLI (Springer)

11.1.4 Scientific Expertise

Salvati Member of Inria's Evaluation Committee.

Tison Elected member of CNU 27.

11.1.5 Research Administration

Tison Membre de l'équipe coordinatrice de l'ISite Université de Lille - Nord Europe

Tison Membre élue du Conseil d'administration de l'Université de Lille.

Salvati Membre de la commission mixte (et restreinte) du département d'informatique et de CRIStAL pour le recrutements. Université de Lille.

11.2 Teaching - Supervision - Juries

11.2.1 Teaching Responibilities

Salvati co-directeur d'étude du Master MIAGE FA,

Salvati directeur d'étude de la licence informatique-mathématique, Université de Lille.

Salvati co-responsable du parcours renforcé recherche de la licence d'informatique, Université de Lille.

Salvati membre du conseil de département, FIL, Université de Lille.

Paperman responsabilité du parcours WebAnalyste du master MIASH, Université de Lille.

Tison member of the selection board for «Capes» in computer science.

Staworko Coordinator of International Relationships at the Department of Computer Science, Université de Lille.

Capelli responsabilité des L1, UFR LEA, Université de Lille.

Capelli membre élu du conseil d'UFR LEA, Université de Lille.

Capelli responsable Parcoursup d'UFR LEA, Université de Lille.

Capelli Concours d'entrée à l'ENS, Testeur du sujet de l'épreuve d'informatique-mathématique, 2020 et Correcteur.

11.2.2 Teaching Activities

Boneva teaches computer science in DUT Informatique of Université de Lille

Capelli teaches computer science in UFR LEA of Université de Lille for around 200h per year (Licence and Master). He is also responsible of remediation of Licence 1 in its UFR.

Lemay teaches computer science in UFR LEA of Université de Lille for around 200h per year (Licence and Master). He is also responsible for computer science and numeric correspondent for its UFR.

Niehren gives lessons for the 2nd year students of the Master MOCAD (Université de Lille): on information extraction (21h).

Paperman teaches computer science for a total of around 200h per year. He gives lessons in UFR MISASH (Université de Lille), in Licence and Master. He also gives a database lesson of 25h in Master MOCAD (Université de Lille).

Salvati teaches computer science for a total of around 230h per year in computer science departement of Université de Lille. That includes Introduction to Computer Science (L1, 50h), Logic (L3, 50h), Algorithmic and operational research (L3, 36h), Functional Programming (L3, 35h), Research Option (L3, 10h), Semantic Web (M2, 30h), Advanced Databases (M1, 20h).

Staworko teaches computer science for a total of around 200h in UFR MIME (Université de Lille).

Tison teaches computer science for a total of around 120h at the Université de Lille. That includes a course on Advanced Algorithms and Complexity (50h, M1), Business Intelligence (36h, M1), Databases (21h L2).

11.2.3 Supervision

Sakho PhD thesis defended in July. Certain Query Answering on Hyperstreams [28]. Supervised by Niehren and Boneva.

Lozano PhD thesis defended in December. Data Exchange from Relational Databases to RDF with Target Shape Schemas [27]. Supervised by Staworko and Boneva.

Gallot PhD project in progress since 2017. On safety of data transformations. Supervised by Salvati and Lemay.

Crosetti PhD project in progress since 2018. Privacy Risks of Aggregates in Data Centric-Workflows. Supervised by Tison, Capelli, Niehren. With Ramon from Inria Magnet.

Soyez-Martin PhD project started 2020. On Streaming with vectors and circuits. Supervised by Salvati and Paperman.

Al Serhaly PhD project started 2020. On hyperstream programming. Supervised by Niehren.

11.2.4 Juries

PhDs committees

Tison Membre du jury de thèse de Théo Grente (Caen)

Tison Membre du jury de thèse de Alexandre Mansard (La Réunion),

Tison Membre du jury de thèse de Mohammed Houssem Eddine Hachmaoui (Saclay, présidente du jury)

HDR committees

Niehren Rapporteur de l'HDR de Loïc Paulevé, Université de Saclay.

Salvati Membre du jury d'HDR de Olivier Gauwin, Université de Bordeaux.

12 Scientific production

12.1 Major publications

[1] A. Amarilli and C. Paperman. 'Topological Sorting with Regular Constraints'. In: *45th International Colloquium on Automata, Languages, and Programming (ICALP 2018)*. Prague, Czech Republic, July 2018. URL: https://hal.archives-ouvertes.fr/hal-01950909.

[2] M. Arenas, P. Barceló, L. Bertossi and M. Monet. 'The Tractability of SHAP-Score-Based Explanations over Deterministic and Decomposable Boolean Circuits'. In: *Thirty-Fifth AAAI Conference on Artificial Intelligence*. Held online, France, Feb. 2021. URL: https://hal.inria.fr/hal-03147623.

[3] C. Barloy, F. Murlak and C. Paperman. 'Stackless Processing of Streamed Trees'. In: *2021 PODS*. Xi'an, Shaanx, China, June 2021. DOI: 10.4230/LIPIcs. URL: https://hal.archives-ouvertes.fr/hal-03021960.

[4] I. Boneva, J. G. Labra Gayo and E. G. Prud 'hommeaux. 'Semantics and Validation of Shapes Schemas for RDF'. In: *ISWC2017 - 16th International semantic web conference*. Vienna, Austria, Oct. 2017. URL: https://hal.archives-ouvertes.fr/hal-01590350.

[5] P. Bourhis, M. Leclère, M.-L. Mugnier, S. Tison, F. Ulliana and L. Gallois. 'Oblivious and Semi-Oblivious Boundedness for Existential Rules'. In: *IJCAI 2019 - International Joint Conference on Artificial Intelligence*. Macao, China, Aug. 2019. URL: https://hal-lirmm.ccsd.cnrs.fr/lirmm-02148142.

[6] F. Capelli, J.-M. Lagniez and P. Marquis. 'Certifying Top-Down Decision-DNNF Compilers'. In: *Thirty-Fifth AAAI Conference on Artificial Intelligence*. Online, France, Feb. 2021. URL: https://hal.inria.fr/hal-03111679.

[7] F. Capelli and S. Mengel. 'Tractable QBF by Knowledge Compilation'. In: *36th International Symposium on Theoretical Aspects of Computer Science (STACS 2019)*. https://arxiv.org/abs/1807.04263. Berlin, Germany, Mar. 2019. URL: https://hal.archives-ouvertes.fr/hal-01836402.

[8] P. D. Gallot, A. Lemay and S. Salvati. 'Linear high-order deterministic tree transducers with regular look-ahead'. In: *MFCS 2020 : The 45th International Symposium on Mathematical Foundations of Computer Science*. Andreas Feldmann, Michal Koucky and Anna Kotesovcova. Prague, Czech Republic, Aug. 2020. DOI: 10.4230/LIPIcs.MFCS.2020.34. URL: https://hal.archives-ouvertes.fr/hal-02902853.

[9] J. Niehren and M. Sakho. 'Determinization and Minimization of Automata for Nested Words Revisited'. In: *Algorithms* (Feb. 2021). URL: https://hal.inria.fr/hal-03134596.

[10] S. Staworko and P. Wieczorek. 'Containment of Shape Expression Schemas for RDF'. In: *SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems (PODS)*. Amsterdam, Netherlands, June 2019. URL: https://hal.inria.fr/hal-01959143.

12.2 Publications of the year

International journals

[11] A. Amarilli, F. Capelli, M. Monet and P. Senellart. 'Connecting Knowledge Compilation Classes and Width Parameters'. In: *Theory of Computing Systems* (1st Aug. 2020). DOI: 10.1007/s00224-019-09930-2. URL: https://hal.inria.fr/hal-02163749.

[12] I. Boneva, J. Niehren and M. Sakho. 'Regular Matching and Inclusion on Compressed Tree Patterns with Constrained Context Variables'. In: *Information and Computation* (24th Feb. 2021). URL: https://hal.inria.fr/hal-03151014.

[13] M. Cadilhac, O. Carton and C. Paperman. 'Continuity of functional transducers: a profinite study of rational functions'. In: *Logical Methods in Computer Science* (21st Feb. 2020). DOI: 10.23638/LMCS-16(1:24)2020. URL: https://hal.archives-ouvertes.fr/hal-03111682.

[14] F. Capelli and Y. Strozecki. 'Enumerating models of DNF faster: breaking the dependency on the formula size'. In: *Discrete Applied Mathematics* (4th June 2020). DOI: 10.1016/j.dam.2020.02.014. URL: https://hal.inria.fr/hal-01891483.

[15] H. García-González, I. Boneva, S. Staworko, J. E. Labra-Gayo and J. M. Cueva Lovelle. 'ShExML: improving the usability of heterogeneous data mapping languages for first-time users'. In: *PeerJ Computer Science* 6 (23rd Nov. 2020), p. 27. DOI: 10.7717/peerj-cs.318. URL: https://hal.archives-ouvertes.fr/hal-03110745.

[16] J. Niehren and M. Sakho. 'Determinization and Minimization of Automata for Nested Words Revisited'. In: *Algorithms* (24th Feb. 2021). URL: https://hal.inria.fr/hal-03134596.

International peer-reviewed conferences

[17] M. Arenas, P. Barceló, L. Bertossi and M. Monet. 'The Tractability of SHAP-Score-Based Explanations over Deterministic and Decomposable Boolean Circuits'. In: Thirty-Fifth AAAI Conference on Artificial Intelligence. Held online, France, 2nd Feb. 2021. URL: https://hal.inria.fr/hal-03147623.

[18] P. Barceló, M. Monet, J. A. Perez and B. Subercaseaux. 'Model Interpretability through the Lens of Computational Complexity'. In: NeurIPS 2020. Held online, United States, 7th Dec. 2020. URL: https://hal.inria.fr/hal-03052508.

[19] C. Barloy, F. Murlak and C. Paperman. 'Stackless Processing of Streamed Trees'. In: PODS 2021 : Symposium on Principles of Database Systems. Proceedings of the Symposium on Principles of Database Systems, PODS 2021. Xi'an, Shaanx, China, 20th June 2021. DOI: 10.4230/LIPIcs. URL: https://hal.archives-ouvertes.fr/hal-03021960.

[20] I. Boneva, J. Niehren and M. Sakho. 'Nested Regular Expressions can be Compiled to Small Deterministic Nested Word Automata'. In: CSR 2020 - 15th International Computer Science Symposium in Russia. Ekaterinburg, Russia, 29th June 2020. URL: https://hal.inria.fr/hal-02532706.

[21] M. Cadilhac, F. Mazowiecki, C. Paperman, M. Pilipczuk and G. Sénizergues. 'On polynomial recursive sequences'. In: ICALP 2020 - 47th International Colloquium on Automata, Languages and Programming. Saarbrücken / Virtual, Germany, 2020. DOI: 10.4230/LIPIcs.ICALP.2020.117. URL: https://hal.archives-ouvertes.fr/hal-03098614.

[22] F. Capelli, J.-M. Lagniez and P. Marquis. 'Certifying Top-Down Decision-DNNF Compilers'. In: Thirty-Fifth AAAI Conference on Artificial Intelligence. Online, France, 2nd Feb. 2021. URL: https://hal.inria.fr/hal-03111679.

[23] P. D. Gallot, A. Lemay and S. Salvati. 'Linear high-order deterministic tree transducers with regular look-ahead'. In: *45th International Symposium on Mathematical Foundations of Computer Science (MFCS 2020)*. MFCS 2020 : The 45th International Symposium on Mathematical Foundations of Computer Science. Prague, Czech Republic, 28th Aug. 2020. DOI: 10.4230/LIPIcs.MFCS.2020.34. URL: https://hal.archives-ouvertes.fr/hal-02902853.

National peer-reviewed Conferences

[24] R. Azimov, J. Niehren and S. Salvati. 'Jumping Evaluation of Nested Regular Path Queries'. In: Bases de données avancées. Online, France, 1st Sept. 2020. URL: https://hal.inria.fr/hal-02492780.

Scientific book chapters

[25] I. Boneva, S. Staworko and J. M. Lozano Aparicio. 'Consistency and Certain Answers in Relational to RDF Data Exchange with Shape Constraints'. In: *Consistency and Certain Answers in Relational to RDF Data Exchange with Shape Constraints*. 17th Aug. 2020, pp. 97–107. DOI: 10.1007/978-3-030-54623-6_9. URL: https://hal.archives-ouvertes.fr/hal-03110741.

[26] O. Bournez, G. Dowek, R. Gilleron, S. Grigorieff, J.-Y. Marion, S. Perdrix and S. Tison. 'A Guided Tour of Artificial Intelligence Research - Volume III: Interfaces and Applications of Artificial IntelligenceTheoretical Computer Science: Computational Complexity'. In: *A Guided Tour of Artificial Intelligence Research - Volume III: Interfaces and Applications of Artificial IntelligenceTheoretical Computer Science: Computational Complexity (10.1007/978-3-030-06170-8)*. 2020. URL: https://hal.archives-ouvertes.fr/hal-02995771.

Doctoral dissertations and habilitation theses

[27] J. M. Lozano Aparicio. 'Data Exchange from Relational Databases to RDF with Target Shape Schemas'. Universite de Lille, Lille, FRA.; Universite de Lille, 14th Dec. 2020. URL: https://tel.archives-ouvertes.fr/tel-03118044.

[28] M. Sakho. 'Certain Query Answering on Hyperstreams'. Université de Lille; Inria, 24th July 2020. URL: https://tel.archives-ouvertes.fr/tel-03028074.