



IN PARTNERSHIP WITH:
CNRS

**Institut polytechnique de
Grenoble**

Université de Grenoble Alpes

Activity Report 2017

Project-Team **TYREX**

Types and Reasoning for the Web

IN COLLABORATION WITH: Laboratoire d'Informatique de Grenoble (LIG)

RESEARCH CENTER
Grenoble - Rhône-Alpes

THEME
**Data and Knowledge Representation
and Processing**

Table of contents

1. Personnel	1
2. Overall Objectives	2
3. Research Program	2
3.1. Modeling	2
3.2. Analysis, verification and optimization	3
3.3. Design of advanced (robust, flexible, rich, novel) web applications	4
4. Application Domains	4
4.1. Web Programming Technologies	4
4.2. Multimedia and Augmented Environments	5
5. New Software and Platforms	5
5.1. sparqlgx	5
5.2. musparql	5
5.3. SPARUB	6
5.4. MRB	6
5.5. TyrAr	7
5.6. AmiAr	7
5.7. GreAR	7
5.8. Benchmarks Attitude Smartphones	7
5.9. MedAnalytics	8
6. New Results	8
6.1. Experimental evaluation of attitude estimation algorithms for smartphones	8
6.2. The SPARQLGX System for Distributed Evaluation of SPARQL Queries	9
6.3. HAP: Building Pipelines with Heterogeneous Data and Hive	9
6.4. Multi-Criteria Experimental Classification of Distributed SPARQL Evaluators	9
6.5. SPARUB: SPARQL UPDATE Benchmark	9
6.6. Optimizing SPARQL query evaluation with a worst-case cardinality estimation	9
6.7. Extending the SPARQL Algebra for the optimization of Property Paths	10
6.8. SPARQL Query Containment with ShEx Constraints	10
6.9. Selectivity Estimation for SPARQL Triple Patterns with Shape Expressions for Optimising SPARQL Query Evaluation	10
6.10. A Circuit-Based Approach to Efficient Enumeration	11
6.11. XQuery Static Type-Checking	11
6.12. Predicting At-Risk Patient Profiles from Big Prescription Data	11
7. Bilateral Contracts and Grants with Industry	11
8. Partnerships and Cooperations	12
8.1. Regional Initiatives	12
8.2. National Initiatives	12
8.2.1. ANR	12
8.2.2. PERSYVAL-lab LabEx	12
9. Dissemination	13
9.1. Promoting Scientific Activities	13
9.1.1. Scientific Events Organisation	13
9.1.1.1. Member of the Conference Program Committees	13
9.1.1.2. Reviewer	13
9.1.2. Journal	13
9.1.3. Scientific Expertise	13
9.1.4. Research Administration	13
9.2. Teaching - Supervision - Juries	14
9.2.1. Teaching	14

9.2.2. Supervision	14
9.2.3. Juries	15
10. Bibliography	15

Project-Team TYREX

Creation of the Team: 2012 November 01, updated into Project-Team: 2014 July 01

Keywords:

Computer Science and Digital Science:

- A2.1.1. - Semantics of programming languages
- A2.1.3. - Functional programming
- A2.1.7. - Distributed programming
- A2.1.10. - Domain-specific languages
- A2.2.1. - Static analysis
- A2.2.4. - Parallel architectures
- A2.4. - Verification, reliability, certification
- A3.1.1. - Modeling, representation
- A3.1.2. - Data management, quering and storage
- A3.1.3. - Distributed data
- A3.1.6. - Query optimization
- A3.1.7. - Open data
- A3.1.8. - Big data (production, storage, transfer)
- A3.2.1. - Knowledge bases
- A3.2.2. - Knowledge extraction, cleaning
- A3.3.3. - Big data analysis
- A3.4. - Machine learning and statistics
- A5.6. - Virtual reality, augmented reality
- A6.3.2. - Data assimilation
- A6.3.3. - Data processing
- A7.2. - Logic in Computer Science
- A9.1. - Knowledge
- A9.7. - AI algorithmics

Other Research Topics and Application Domains:

- B6.1. - Software industry
- B6.3.1. - Web
- B6.5. - Information systems
- B8.2. - Connected city
- B9.4.1. - Computer science
- B9.4.5. - Data science
- B9.7.2. - Open data
- B9.9. - Risk management
- B9.9.2. - Financial risks

1. Personnel

Research Scientists

Nabil Layaida [Team leader, Inria, Senior Researcher, HDR]

Pierre Geneves [CNRS, Researcher, HDR]

Faculty Members

Nils Gesbert [Institut polytechnique de Grenoble, Associate Professor]

Cécile Roisin [Univ Grenoble Alpes, Professor, HDR]

Post-Doctoral Fellows

Damien Graux [Inria, until Apr 2017]

Thibaud Michel [Univ Grenoble Alpes, from Dec 2017]

PhD Students

Abdullah Abbas [Univ Grenoble Alpes, until Sep 2017]

Fateh Boulmaiz [Univ Grenoble Alpes, from Nov 2017]

Louis Jachiet [Ecole Normale Supérieure Paris]

Raouf Kerkouche [Inria, from Oct 2017]

Muideen Lawal [Univ Grenoble Alpes, from Oct 2017]

Thibaud Michel [Univ Grenoble Alpes, until Oct 2017]

Intern

Joseph Marotte [Inria, from Jun 2017 until Aug 2017]

Administrative Assistant

Helen Pouchot-Rouge-Blanc [Inria]

2. Overall Objectives

2.1. Objectives

The TyReX team aims at developing a vision of a web where content is enhanced and protected, applications made easier to build, maintain and secure. It seeks to open new horizons for the development of the web, enhancing its potential, effectiveness, and dependability. In particular, we aim at making contributions by obtaining fundamental results, by building advanced experimental applications showcasing these results and by contributing to web standards. One fundamental problem of our time is a lack of formalisms, concepts and tools for reasoning simultaneously about content or data, programs, and communication aspects. Our main scientific goal is to establish a unifying development framework for designing advanced (robust, flexible, rich, efficient and novel) web applications.

To tackle our overall goal, we decomposed the problem along three dimensions, each corresponding to a more specific objective and research theme:

1. models, to deal with the issues of heterogeneous data and application complexity by abstracting away from document formats and programming language syntax;
2. analysis, verification and optimization; and
3. design of advanced distributed web application, to address programming in mobile and large-scale distributed systems.

3. Research Program

3.1. Modeling

Modeling consists in capturing various aspects of document and data processing and communication in a unifying model. Our modeling research direction mainly focuses on three aspects.

The first aspect aims at reducing the impedance mismatch. The impedance mismatch refers to the complexity, difficulty and lack of performance induced by various web application layers which require the same piece of information to be represented and processed differently. The mismatch occurs because programming languages use different native data models than those used for documents in browsers and for storage in databases. This results in complex and multi-tier software architectures whose different layers are incompatible in nature. This, in turn, results in expensive, inefficient, and error-prone web development. For reducing the impedance mismatch, we will focus on the design of a unifying software stack and programming framework, backed by generic and solid logical foundations similar in spirit to the NoSQL approach.

The second aspect aims at harnessing heterogeneity. Web applications increasingly use diverse data models: ordered and unordered tree-like structures (such as XML), nested records and arrays (such as JSON), graphs (like RDF), and tables. Furthermore, these data models involve a variety of languages for expressing constraints over data (e.g. XML schema, RelaxNG, and RDFS to name just a few). We believe that this heterogeneity is here to stay and is likely to increase. These differences in representations imply loads of error-prone and costly conversions and transformations. Furthermore, some native formats (e.g. JSON) are repurposed from an internal representation to a format for data exchange. This often results in a loss of information and in errors that need to be tracked and corrected. In this context, it is important to seek methods for reducing risks of information loss during data transformation and exchange. For harnessing heterogeneity, we will focus on the integration of data models through unified formal semantics and in particular logical interpretation. This allows using the same programming language constructs on different data models. At the programming language level, this is similar to languages such as JSoNq for JSON and XML.

Finally, the third aspect aims at making applications and data more compositional. Most web programming technologies are currently limited from a compositional point of view. For example, tree grammars (like schema languages for XML) are monolithic in the sense that they require the full description of the considered structures, instead of allowing the assembly of smaller and reusable building blocks. As a consequence, this translates into monolithic web applications, which makes their automated verification harder by making modular analyses more difficult. The need for compositionality is illustrated in the industry by the increasing development of fragmented W3C specifications organised in ad-hoc modules. For making applications and data more compositional, we will focus on the design of modular schema and programming languages. For this purpose, we will notably rely on succinct yet expressive formalisms (like two-way logics, polymorphic types, session types) that ease the process of expressing modular specifications.

3.2. Analysis, verification and optimization

This research direction aims at guaranteeing two different kinds of properties: safety and efficiency.

The first kind of properties concerns the safety of web applications. Software development was traditionally split between critical and non-critical software. Advanced (and costly) formal verification techniques were reserved to the former whereas non-critical software relied almost exclusively on testing, which only offers a ‘best-effort’ guarantee (removes most bugs but some of them may not be detected). The central idea was that in a non-critical system, the damage a failure may create is not worth the cost of formal verification. However, as web applications grow more pervasive in everyday life and gain momentum in corporates and various social organizations, and touch larger numbers of users, the potential cost of failure is rapidly and significantly increasing. In that sense, we can consider that web applications are becoming more and more critical. The growing dependency on the web as a tool, combined with the fact that some applications involve very large user bases, is becoming problematic as it seems to increase rapidly but silently. Some errors like crashes and confidential information leaks, if not discovered, can have massive effects and cause significant financial or reputation damage.

The second kind of properties concerns the efficiency of web applications. One particular characteristic of web programming languages is that they are essentially data-manipulation oriented. These manipulations rely on query and transformation languages whose performance is critical. This performance is very sensitive to data size and organization (constraints) and to the execution model (e.g. streaming evaluators). Static analysis can be used to optimize runtime performance by compile-time automated modification of the code (e.g.

substitution of queries by more efficient ones). One major scientific difficulty here consists in dealing with problems close to the frontier of decidability, and therefore in finding useful trade-offs between programming ease, expressivity, complexity, succinctness, algorithmic techniques and effective implementations.

3.3. Design of advanced (robust, flexible, rich, novel) web applications

The generalized use of mobile terminals deeply affects the way users perceive and interact with their environment. The ubiquitous use of search engines capable of producing results in fractions of a second raised user expectations to a very high level: users now expect relevant information to be made available to them instantly and directly by context sensitivity to the environment itself. However, the information that needs to be processed is becoming more and more complex compared to the traditional web. In order to unlock the potential introduced by this new generation of the web, a radical rethinking of how web information is produced, organized and processed is necessary.

Until now, content rendering on the web was mainly based on supporting media formats separately. It is still notably the case in HTML5 for example where, for instance, vector graphics, mathematical content, audio and video are supported only as isolated media types. With the increasing use of web content in mobile terminals, we also need to take into account highly dynamic information flowing from sensors (positioning and orientation moves) and cameras. To reach that goal, web development platforms need to ease the manipulation of such content with carefully designed programming interfaces and by developing supporting integrative methods.

More precisely, we will focus on the following aspects: (1) **Build Rich content models**. This requires combining in a single model several content facets such as 3D elements, animations, user interactions, etc. We will focus on feature-compositional methods, which have become a prerequisite for the production of compelling web applications. (2) **Physical environment modeling and integration**. This consists of modeling and representing urban data such as buildings, pathways, points of interest. It requires developing appropriate languages and techniques to represent, process and query such environment models. In particular, we will focus on tracking positional user information and design techniques capable of combining semantic annotations, content, and representation of the physical world. (3) **Native streams support**. This consists of capturing new data flows extracted from various sensors in mobile terminals and various equipments. (4) **Cross-platform abstractions**. We will contribute to the design of appropriate abstractions to make applications run in a uniform way across various devices and environments. Our goal is to provide a viable alternative to current (platform-specific) mobile application development practices.

4. Application Domains

4.1. Web Programming Technologies

Despite the major social and economic impacts of the web revolution, current web programming methods and content representation are lagging behind and remain severely limited and in many respects archaic. Dangerously, designing web applications even becomes increasingly complex as it relies more and more on a jungle of programming languages, tools and data formats, each targeted toward a different application layer (presentation, application and storage). This often yields complex and opaque applications organized in silos, which are costly, inefficient, hard to maintain and evolve, and vulnerable to errors and security holes. In addition, the communication aspects are often handled independently via remote service invocations and represent another source of complexity and vulnerability. We believe that we reached a level where there is an urgent need and a growing demand for alternative programming frameworks that capture the essence of web applications: advanced content, data and communication. Therefore, successful candidate frameworks must capture rich document formats, data models and communication patterns. A crucial aspect is to offer correction guarantees and flexibility in the application architecture. For instance, applications need to be checked, optimized and managed as a whole while leveraging on the consistency of their individual components and data fragments. For all these reasons, we believe that a new generation of tools must be created and developed in order to overcome the aforementioned limitations of current web technologies.

4.2. Multimedia and Augmented Environments

The term Augmented Environments refers collectively to ubiquitous computing, context-aware computing, and intelligent environments. The goal of our research on these environments is to introduce personal Augmented Reality (AR) devices, taking advantage of their embedded sensors. We believe that personal AR devices such as mobile phones or tablets will play a central role in augmented environments. These environments offer the possibility of using ubiquitous computation, communication, and sensing to enable the presentation of context-sensitive information and services to the user. AR applications often rely on 3D content and employ specialized hardware and computer vision techniques for both tracking and scene reconstruction and exploration. Our approach tries to seek a balance between these traditional AR contexts and what has come to be known as mobile AR browsing. It first acknowledges that mobile augmented environment browsing does not require that 3D content be the primary means of authoring. It provides instead a method for HTML5 and audio content to be authored, positioned in the surrounding environments and manipulated as freely as in modern web browsers. The applications we develop to guide and validate our concepts are pedestrian navigation techniques and applications for cultural heritage visits. Features found in augmented environments are demanding for the other activities in the team. They require all kinds of multimedia information, that they have to combine. This information has to be processed efficiently and safely, often in real time, and it also, for a significant part, has to be created by human users.

5. New Software and Platforms

5.1. sparqlgx

KEYWORDS: RDF - SPARQL - Distributed computing

SCIENTIFIC DESCRIPTION: SPARQL is the W3C standard query language for querying data expressed in RDF (Resource Description Framework). The increasing amounts of RDF data available raise a major need and research interest in building efficient and scalable distributed SPARQL query evaluators.

In this context, we propose and share SPARQLGX: our implementation of a distributed RDF datastore based on Apache Spark. SPARQLGX is designed to leverage existing Hadoop infrastructures for evaluating SPARQL queries. SPARQLGX relies on a translation of SPARQL queries into executable Spark code that adopts evaluation strategies according to (1) the storage method used and (2) statistics on data. Using a simple design, SPARQLGX already represents an interesting alternative in several scenarios.

FUNCTIONAL DESCRIPTION: Distributed SPARQL query evaluator

RELEASE FUNCTIONAL DESCRIPTION: - Faster load routine which widely improves this phase performances by reading once the initial triple file and by partitioning data in the same time into the correct predicate files. - Improving the generated Scala-code of the translation process with mapValues. This technic allows not to break the partitioning of KeyValueRDD while applying transformations to the values instead of the traditional map that was done prior. - Merging and cleaning several scripts in bin/ such as for example sgx-eval.sh and sde-eval.sh - Improving the compilation process of compile.sh - Cleaner test scripts in tests/ - Offering the possibility of an easier deployment using Docker.

- Participants: Damien Graux, Thomas Calmant, Louis Jachiet, Nabil Layaïda and Pierre Genevès
- Contact: Pierre Genevès
- Publications: [Optimizing sparql query evaluation with a worst-case cardinality estimation based on statistics on the data](#) - [The SPARQLGX System for Distributed Evaluation of SPARQL Queries](#)
- URL: <https://github.com/tyrex-team/sparqlgx>

5.2. musparql

KEYWORDS: SPARQL - RDF - Property paths

FUNCTIONAL DESCRIPTION: reads a SPARQL request and translates it into an internal algebra. Rewrites the resulting term into many equivalent versions, then chooses one of them and executes it on a graph.

- Participant: Louis Jachiet
- Contact: Nabil Layaïda
- Publication: [Extending the SPARQL Algebra for the optimization of Property Paths](#)
- URL: <https://gitlab.inria.fr/tyrex/musparql>

5.3. SPARUB

SPARQL UPDATE Benchmark generator:

KEYWORDS: SPARQL - RDF

SCIENTIFIC DESCRIPTION: One aim of the RDF data model, as standardized by the W3C, is to facilitate the evolution of data over time without requiring all the data consumers to be changed. To this end, one of the latest additions to the SPARQL standard query language is an update language for RDF graphs. The research on efficient and scalable SPARQL evaluation methods increasingly relies on standardized methodologies for benchmarking and comparing systems. However, current RDF benchmarks do not support graph updates. We propose and share SPARUB: a benchmark for the SPARQL update language on RDF graphs. The aim of SPARUB is not to be yet another RDF benchmark. Instead it provides the means to automatically extend and improve existing RDF benchmarks along a new dimension of data updates, while preserving their structure and query scenarios.

FUNCTIONAL DESCRIPTION: SPARUB is a simple tool to generate additional scenarios of test from an already existing N-Triples dataset and some SPARQL queries while focusing on the SPARQL UPDATE fragment (which is part of SPARQL 1.1). It simply extends already existing benchmarking methods taking an RDF dataset and (optionally) SPARQL queries to provide a complete scenario of test. Moreover, a list of predefined metrics is also available to extract interesting figures of the tests.

Technically, SPARUB is a bash script `sparub.sh` which takes a triple file and an optional list of SPARQL queries as arguments. It will then generate a scenario divided into several steps to benchmark an RDF storage system allowing the SPARQL evaluation on the various functionalities of the SPARQL UPDATE standard extension.

- Participants: Damien Graux, Pierre Genevès and Nabil Layaïda
- Contact: Pierre Genevès
- Publication: [SPARUB: SPARQL UPDATE Benchmark](#)
- URL: <https://github.com/tyrex-team/sparub>

5.4. MRB

Mixed Reality Browser

KEYWORDS: Augmented reality - Geolocation - Indoor geolocalisation - Smartphone

FUNCTIONAL DESCRIPTION: MRB displays PoI (Point of Interest) content remotely through panoramics with spatialized audio, or on-site by walking to the corresponding place, it can be used for indoor-outdoor navigation, with assistive audio technology for the visually impaired. It is the only browser of geolocalized data to use XML as a native format for PoIs, panoramics, 3D audio and to rely on HTML5 both for the iconic and full information content of PoIs. Positioning in MRB is based on a PDR library, written in C++ and Java and developed by the team, which provides the user's location in real time based on the interpretation of sensors. Three main modules have been designed to build this positioning system: (i) a pedometer that estimates the distance the user has walked and his speed, (ii) a motion manager that enables data set recording and simulation but also the creation of virtual sensors or filters (e.g. gyroscope drift compensation, linear acceleration, altimeter), and (iii) a map-matching algorithm that provides a new location based on a given OpenStreetMap file description and the current user's trajectory.

- Participant: Thibaud Michel
- Contact: Nabil Layaïda
- Publications: [On Mobile Augmented Reality Applications based on Geolocation - Attitude Estimation with Smartphones](#)
- URL: <http://tyrex.inria.fr/projects/mrb.html>

5.5. TyrAr

Geo Augmented Reality on a Smartphone

KEYWORDS: Augmented reality - Smartphone - Geolocation

FUNCTIONAL DESCRIPTION: This application is an AR viewer to name the mountains, cities and historical buildings over the camera feed of the smartphone. The user can turn on himself with his device to discover names and information about Points of Interest (POIs). POIs are directly extracted from the OSM database thanks to the Overpass Turbo API. POIs are displayed on the screen with their name, an icon and an extra information. City POIs exhibit the number of inhabitants, mountains are associated to their altitude and historical buildings display their date of construction.

- Participant: Thibaud Michel
- Contact: Nabil Layaïda
- Publication: [On Mobile Augmented Reality Applications based on Geolocation](#)
- URL: <http://tyrex.inria.fr/projects/mrb.html>

5.6. AmiAr

Smart Home Augmented Reality on a Smartphone

KEYWORDS: Augmented reality - Smart home - Smartphone - Indoor geolocalisation

FUNCTIONAL DESCRIPTION: This application is a proof of concept of a Geo AR system in a smart apartment. This setup has been conducted in EquipEx Amiqual4Home. The goal here is to control objects in the apartment using widgets over the video feed from the camera. For example, a user points a lamp with his smartphone, a widget appears, then he uses a slider in this widget to modify the light intensity.

- Participant: Thibaud Michel
- Contact: Nabil Layaïda
- Publication: [On Mobile Augmented Reality Applications based on Geolocation](#)

5.7. GreAR

Grenoble AR-Tour based on geolocation.

KEYWORDS: Augmented reality - Geolocation - Smartphone

FUNCTIONAL DESCRIPTION: This application is an AR navigator specifically designed for pedestrians. This application was initially developed during the Venturi FP7 (2011-2015) project and has been updated with our AR framework since then. Between two visually driven AR experiences (at the time, developed by partners), the navigator provides the user with an audio and visual guidance through a pre-defined touristic path in Grenoble. The position of the user is obtained through a fusion of GPS signal (when available), pedometer estimates and a map-matching algorithm exploiting OpenStreetMap. As the GPS signal is poor in several parts of the old city the integration of the pedometer enables the navigator to obtain a sufficiently reliable position estimate, crucial for AR applications and geofencing. Within the application, there are several options given to the user to view the navigation path through the city, ranging from a satellite image of the streets to a vector map. In the navigation pane, the geofences relating to the AR experiences and other points of interest can be seen.

- Participant: Thibaud Michel
- Contact: Nabil Layaïda
- Publication: [On Mobile Augmented Reality Applications based on Geolocation](#)
- URL: <http://tyrex.inria.fr/projects/mrb.html>

5.8. Benchmarks Attitude Smartphones

KEYWORDS: Performance analysis - Sensors - Motion analysis - Experimentation - Smartphone

SCIENTIFIC DESCRIPTION: We investigate the precision of attitude estimation algorithms in the particular context of pedestrian navigation with commodity smartphones and their inertial/magnetic sensors. We report on an extensive comparison and experimental analysis of existing algorithms. We focus on typical motions of smartphones when carried by pedestrians. We use a precise ground truth obtained from a motion capture system. We test state-of-the-art attitude estimation techniques with several smartphones, in the presence of magnetic perturbations typically found in buildings. We discuss the obtained results, analyze advantages and limits of current technologies for attitude estimation in this context. Furthermore, we propose a new technique for limiting the impact of magnetic perturbations with any attitude estimation algorithm used in this context. We show how our technique compares and improves over previous works.

- Participants: Hassen Fourati, Nabil Layaïda, Pierre Genevès and Thibaud Michel
- Partner: GIPSA-Lab
- Contact: Pierre Genevès
- URL: <http://tyrex.inria.fr/mobile/benchmarks-attitude/>

5.9. MedAnalytics

KEYWORDS: Big data - Predictive analytics - Distributed systems

FUNCTIONAL DESCRIPTION: We implemented a method for the automatic detection of at-risk profiles based on a fine-grained analysis of prescription data at the time of admission. The system relies on an optimized distributed architecture adapted for processing very large volumes of medical records and clinical data. We conducted practical experiments with real data of millions of patients and hundreds of hospitals. We demonstrated how the various perspectives of big data improve the detection of at-risk patients, making it possible to construct predictive models that benefit from volume and variety. This prototype implementation is described in the 2017 preprint available at: <https://hal.inria.fr/hal-01517087/document>.

- Participants: Pierre Genevès and Thomas Calmant
- Partner: CHU Grenoble
- Contact: Pierre Genevès
- Publication: [Predicting At-Risk Patient Profiles from Big Prescription Data](#)

6. New Results

6.1. Experimental evaluation of attitude estimation algorithms for smartphones

- **Context:** Pervasive applications on smartphones increasingly rely on techniques for estimating attitude. Attitude is the orientation of the smartphone with respect to Earth's local frame. Modern smartphones embed sensors such as accelerometer, gyroscope, and magnetometer which make it possible to leverage existing attitude estimation algorithms.
- **Contribution:** we focused on smartphone attitude estimation. We proposed the first benchmark using a motion lab with a high precision (the Inria Kinovis platform) for the purpose of comparing and evaluating filters from the literature on a common basis. This allowed us to provide the first in-depth comparative analysis in this context. In particular, we focused on typical motions of smartphones when carried by pedestrians. Furthermore, we proposed a new parallel filtering technique for limiting the impact of magnetic perturbations with any attitude estimation algorithm used in this context. We showed how our technique compares and improves over previous works. We made our benchmark available (see Benchmarks Attitude Smartphones in Software section) and paid attention to the reproducibility of results. We analyzed and discussed the obtained results and reported on lessons learned [14], [9], [24].

6.2. The SPARQLGX System for Distributed Evaluation of SPARQL Queries

In this work [19], we propose SPARQLGX: an implementation of a distributed RDF datastore based on Apache Spark. SPARQLGX is designed to leverage existing Hadoop infrastructures for evaluating SPARQL queries efficiently. SPARQLGX relies on an automated translation of SPARQL queries into optimized executable Spark code. We show that SPARQLGX makes it possible to evaluate SPARQL queries on billions of triples distributed across multiple nodes, while providing attractive performance figures. We report on experiments which show how SPARQLGX compares to state-of-the-art implementations and we show that our approach scales better than other systems in terms of supported dataset size. With its simple design, SPARQLGX represents an interesting alternative in several scenarios.

6.3. HAP: Building Pipelines with Heterogeneous Data and Hive

The increasing number of available datasets gives opportunities to build large and complex applications which aggregate results coming from several sources. These emerging use cases require new systems where the combinations of heterogeneous sources are both allowed and efficient. To tackle these challenges, we built a system [17] offering a simple high-level set of primitives – called HAP – to easily describe processing chains. These descriptions are then compiled into optimized SQL queries executed on Hive.

6.4. Multi-Criteria Experimental Classification of Distributed SPARQL Evaluators

In this work [13], we provide a new perspective on distributed sparql evaluators, based on a multi-criteria ranking obtained through extensive experiments. Specifically, we propose a set of five principal features which we use to rank evaluators. Each system exhibits a particular combination of these features. Similarly, the various requirements of practical use cases can also be decomposed in terms of these features. Our suggested set of features provides a more comprehensive description of the behavior of a distributed evaluator when compared to traditional performance metrics. We show how it helps in more accurately evaluating to which extent a given system is appropriate for a given use case. For this purpose, we systematically benchmarked a panel of 10 state-of-the-art implementations. We ranked them using this reading grid to pinpoint the advantages and limitations of current sparql evaluation systems.

6.5. SPARUB: SPARQL UPDATE Benchmark

One aim of the RDF data model, as standardized by the W3C, is to facilitate the evolution of data over time without requiring all the data consumers to be changed. To this end, one of the latest addition to the SPARQL standard query language is an update language for RDF graphs. The research on efficient and scalable SPARQL evaluation methods increasingly relies on standardized methodologies for benchmarking and comparing systems. However, current RDF benchmarks do not support graphs updates. We propose SPARUB: a benchmark for the SPARQL UPDATE language on RDF graphs [18]. The aim of SPARUB is not to be yet another RDF benchmark. Instead it provides the mean to automatically extend and improve existing RDF benchmarks along a new dimension of data updates, while preserving their structure and query scenarios.

6.6. Optimizing SPARQL query evaluation with a worst-case cardinality estimation

SPARQL is the w3c standard query language for querying data expressed in the Resource Description Framework (RDF). There exists a variety of SPARQL evaluation schemes and, in many of them, estimating the cardinality of intermediate results is key for performance, especially when the computation is distributed and the datasets very large. For example it helps in choosing join orders that minimize the size of intermediate subquery results.

In this context [23], we propose a new cardinality estimation based on statistics about the data. Our cardinality estimation is a worst-case analysis tailored for SPARQL and capable of taking advantage of the implicit schema often present in RDF datasets (e.g. functional dependencies). This implicit schema is captured by statistics therefore our method does not need for the schema to be explicit or perfect (our system performs well even if there are a few “violations” of these implicit dependencies). We implemented our cardinality estimation and used it to optimize the evaluation of SPARQL queries: equipped with our cardinality estimation, the query evaluator performs better against most queries (sometimes by an order of magnitude) and is only ever slightly slower.

6.7. Extending the SPARQL Algebra for the optimization of Property Paths

In this work [22], [21], we propose a new algebra, μ -algebra, inspired by works on the relational algebra, SQL and NoSQL languages (especially SPARQL) along with a prototype implementation of a SPARQL optimizer based on this algebra¹. Our algebra has the following properties: (1) It subsumes the SPARQL Algebra (under the set semantics) with a more general recursion. (2) SPARQL with Property Paths can be efficiently translated to this algebra. (3) We have a type system and rewriting rules for terms of this algebra that allow optimization, notably of terms involving recursion. We illustrate the differences and the benefits of our approach on recursive query optimization. While a generic approach often comes at the cost of performance, we experimentally show that this approach actually leads to more efficient evaluation of queries with Property Paths. We also show that our approach produces Query Execution Plans (QEP) that are not considered by other existing approaches.

6.8. SPARQL Query Containment with ShEx Constraints

Data structured in the Resource Description Framework (RDF) are increasingly available in large volumes. This leads to a major need and research interest in novel methods for query analysis and compilation for making the most of RDF data extraction. SPARQL is the widely used and well supported standard query language for RDF data. In parallel to query language evolutions, schema languages for expressing constraints on RDF datasets also evolve. Shape Expressions (ShEx) are increasingly used to validate RDF data, and to communicate expected graph patterns. Schemas in general are important for static analysis tasks such as query optimisation and containment.

In this work [10], [8], we investigate the means and methodologies for SPARQL query static analysis in the presence of ShEx schema constraints. Our contribution consists in considering the problem of SPARQL query containment in the presence of ShEx constraints. We propose a sound and complete procedure for the problem of containment with ShEx, considering several SPARQL fragments. Particularly our procedure considers OPTIONAL query patterns, that turns out to be an important feature to be studied with schemas. We provide complexity bounds for the containment problem with respect to the language fragments considered. We also propose alternative method for SPARQL query containment with ShEx by reduction into First Order Logic satisfiability, which allows for considering SPARQL fragment extension in comparison to the first method. This is the first work addressing SPARQL query containment in the presence of ShEx constraints.

6.9. Selectivity Estimation for SPARQL Triple Patterns with Shape Expressions for Optimising SPARQL Query Evaluation

ShEx (Shape Expressions) is a language for expressing constraints on RDF graphs. In this work [11], [8], [15], we optimize the evaluation of conjunctive SPARQL queries, on RDF graphs, by taking advantage of ShEx constraints. Our optimization is based on computing and assigning ranks to query triple patterns, dictating their order of execution. The presence of intermediate joins between the query triple patterns is the reason why ordering is important in increasing efficiency. We first define a set of well formed ShEx schemas, that possess interesting characteristics for SPARQL query optimization. We then define our optimization method by exploiting information extracted from a ShEx schema. We finally report on evaluation results performed showing the advantages of applying our optimization on the top of an existing state-of-the-art query evaluation system.

6.10. A Circuit-Based Approach to Efficient Enumeration

In this work [12], we study the problem of enumerating the satisfying valuations of a circuit while bounding the delay, i.e., the time needed to compute each successive valuation. We focus on the class of structured d-DNNF circuits originally introduced in knowledge compilation, a sub-area of artificial intelligence. We propose an algorithm for these circuits that enumerates valuations with linear preprocessing and delay linear in the Hamming weight of each valuation. Moreover, valuations of constant Hamming weight can be enumerated with linear preprocessing and constant delay. Our results yield a framework for efficient enumeration that applies to all problems whose solutions can be compiled to structured d-DNNFs. In particular, we use it to recapture classical results in database theory, for factorized database representations and for MSO evaluation. This gives an independent proof of constant-delay enumeration for MSO formulae with first-order free variables on bounded-treewidth structures.

6.11. XQuery Static Type-Checking

Although XQuery is a statically typed, functional query language for XML data, some of its features such as upward and horizontal XPath axes are typed imprecisely. The main reason is that while the XQuery data model allows us to navigate upwards and between siblings from a given XML node, the type model, e.g., regular tree types, can describe only the subtree structure of the given node. Recently, Castagna et al. (2015) and Genevès and Gesbert (2015) independently propose a precise forward type inference system for XQuery using an extended type language that can describe not only a given XML node but also its context. In this work [20], as a complementary method to forward type inference systems, we propose a novel backward type inference system for XQuery, using the type language proposed by Genevès and Gesbert (2015). Our backward type inference system provides an exact typing result for XPath axes and a sound typing result for XQuery expressions.

6.12. Predicting At-Risk Patient Profiles from Big Prescription Data

In this work [16], we show how the analysis of very large amounts of drug prescription data make it possible to detect, on the day of hospital admission, patients at risk of developing complications during their hospital stay. We explore, for the first time, to which extent volume and variety of big prescription data help in constructing predictive models for the automatic detection of at-risk profiles. Our methodology is designed to validate our claims that: (1) drug prescription data on the day of admission contain rich information about the patient's situation and perspectives of evolution, and (2) the various perspectives of big medical data (such as veracity, volume, variety) help in extracting this information. We build binary classification models to identify at-risk patient profiles. We use a distributed architecture to ensure scalability of model construction with large volumes of medical records and clinical data. We report on practical experiments with real data of millions of patients and hundreds of hospitals. We demonstrate how the fine-grained analysis of such big data can improve the detection of at-risk patients, making it possible to construct more accurate predictive models that significantly benefit from volume and variety, while satisfying important criteria to be deployed in hospitals.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

Transfer contract

Partner: **Oppidoc** startup

Coordinator: Pierre Genevès

Abstract: the goal of this project is to investigate the integration of advanced static analyses in Oppidoc's flagship product, Oppidum, which is a software framework for constructing web sites with forms for the collaborative edition and publishing of structured documents.

8. Partnerships and Cooperations

8.1. Regional Initiatives

AGIR

Title: Data-CILE

Call: Appel à projet Grenoble Innovation Recherche (AGIR-Pole)

Duration: 2016-2018

Coordinator: Nabil Layaïda

Abstract: The goal of this project is to contribute to foundational and algorithmic challenges introduced by increasingly popular data-centric paradigms for programming on distributed architectures such as spark and the massive production of big linked open data. The focus of the project is on building robust and more efficient workflows of transformations of rich web data. We will investigate effective programming models and compilation techniques for producing specialised language runtimes. We will focus on high-level specifications of pipelines of data transformations and extraction for producing valuable knowledge from rich web data. We will study how to synthesise code which is correct and optimised for execution on distributed platforms. The overall expected outcome is to make the development of rich-data-intensive applications less error-prone and more efficient.

8.2. National Initiatives

8.2.1. ANR

CLEAR

Title: Compilation of intermediate Languages into Efficient big dAta Runtimes

Call: Appel à projets générique 2016 défi ‘Société de l’information et de la communication’ – JCJC

Duration: October 2016 – September 2020

Coordinator: Pierre Genevès

See also: <http://tyrex.inria.fr/clear>

Abstract: This project addresses one fundamental challenge of our time: the construction of effective programming models and compilation techniques for the correct and efficient exploitation of big and linked data. We study high-level specifications of pipelines of data transformations and extraction for producing valuable knowledge from rich and heterogeneous data. We investigate how to synthesize code which is correct and optimized for execution on distributed infrastructures.

8.2.2. PERSYVAL-lab LabEx

Title: Mobile Augmented Reality Applications for Smart Cities

Call: Persyval Labex (“Laboratoire d’excellence”).

Duration: 2014 – 2017

Coordinators: Pierre Genevès and Nabil Layaïda

Others partners: NeCS team at GIPSA-Lab laboratory.

Abstract: The goal of this project is to increase the relevance and reliability of augmented reality (AR) applications, through three main objectives:

1. Finding and developing appropriate representations for describing the physical world (3D maps, indoor buildings, ways...), integrated advanced media types (3D, 3D audio, precisely geo-tagged pictures with lat., long. and orientation, video...)

2. Integrating the different abstraction levels of these data streams (ranging from sensors data to high level rich content such as 3D maps) and bridging the gap with Open Linked Data (the semantic World). This includes opening the way to query the environment (filtering), and adapt AR browsers to users' capabilities (e.g. blind people). The objective here is to provide an open and scalable platform for mobile-based AR systems (just like the web represents).
3. Increasing the reliability and accuracy of localization technologies. Robust and high-accuracy localization technologies play a key role in AR applications. Combined with geographical data, they can also be used to identify user-activity patterns, such as walking, running or being in an elevator. The interpretation of sensor values, coupled with different walking models, allows one to ensure the continuity of the localization, both indoor and outdoor. However, dead reckoning based on Inertial Navigation Systems (INS) or Step-and-Heading Systems (SHS) is subject to cumulative errors due to many factors (sensor drift (accelerometers, gyroscopes, etc.), missed steps, bad estimation of the length of each stride, etc.). One objective is to reduce such errors by merging and mixing these approaches with various external signals such as GPS and Wi-Fi or relying on the analyses of user trajectories with the help of a structured map of the environment. Some filtering methods (Kalman Filter, observer, etc.) will be useful to achieve this task.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organisation

9.1.1.1. Member of the Conference Program Committees

- P. Genevès was a member of the program committee of **BDA 2017**, the 33rd National Conference on Databases and Applications.
- N. Layaïda was a member of the program committee of **The Web Conference 2018 - The Web Programming, Design, Analysis, and Implementation** track.

9.1.1.2. Reviewer

- P. Genevès has been referee for the following 2017 conferences: IJCAI 2017, ICALP 2017, DocEng 2017, and BDA 2017.
- N. Gesbert has been referee for CAV 2017.
- C. Roisin has been referee for the following 2017 conferences: DocEng 2017 and SMAP 2017.

9.1.2. Journal

9.1.2.1. Member of the Editorial Boards

- P. Genevès has been referee for the **Information and Computation (I&C)** journal

9.1.3. Scientific Expertise

- P. Genevès was expert for the ANRT (CIFRE).
- C. Roisin was expert for the ANR 'Appel à Projets Génériques' 2017.

9.1.4. Research Administration

- P. Genevès is responsible for the Computer Science Specialty at the **Doctoral School MSTII (ED 217)**.

- N. Layaïda is ‘réfèrent budget’ member of the budget commission of the Inria Grenoble – Rhône-Alpes research center. The role of this commission is to allocate yearly budget (‘dotation’) to Inria project teams and services. On a yearly basis, we meet with team and service leaders individually, collect their financial needs and set their budget.
- N. Layaïda is member of the Scientific Board of Advanced Data-mining of the Persyval Labex.
- N. Layaïda is member of the Scientific Board of Digital League, the digital cluster of Auvergne-Rhône-Alpes.
- N. Layaïda is member of the experts pool (selection committee) of the minalogic competitive cluster.
- C. Roisin is member of the CNU ‘Conseil national des Universités’ section 27.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Master: P. Genevès, Semantic Web: from XML to OWL, 54hEqTD, M2 (MOSIG), University Grenoble Alpes, France

Licence : N. Gesbert, ‘Logique pour l’informatique’, 45 h eq TD, L3, Grenoble INP

Licence : N. Gesbert, ‘Bases de la programmation impérative’, 33 h eq TD, L3, Grenoble INP

Licence : N. Gesbert, academic tutorship of an apprentice, 10 h eq TD, L3, Grenoble INP

Master : N. Gesbert, ‘Fondements logiques pour l’informatique’, 12 h eq TD, M1, Grenoble INP

Master : N. Gesbert, ‘Construction d’applications Web’, 22 h 30 eq TD, M1, Grenoble INP

Master : N. Gesbert, ‘Analyse, conception et validation de logiciels’, 41 h 15 eq TD, M1, Grenoble INP

Licence : C. Roisin, ‘Programmation C’, 12h eq TD, L2, IUT2, Univ. Grenoble-Alpes

Licence : C. Roisin, ‘Architecture des réseaux’, 112h eq TD, L1, IUT2, Univ. Grenoble-Alpes

Licence : C. Roisin, ‘Services réseaux’, 22h eq TD, L2, IUT2, Univ. Grenoble-Alpes

Licence : C. Roisin, ‘Introduction système Linux’, 21h eq TD, L1, IUT2, Univ. Grenoble-Alpes

Licence : C. Roisin, ‘Système et réseaux’, 14h eq TD, L3, IUT2, Univ. Grenoble-Alpes

Licence : C. Roisin, academic tutorship of four apprentices, 20h eq TD, L3, IUT2, Univ. Grenoble-Alpes

Licence : C. Roisin, academic tutorship of 18 students, 13h eq TD, L1, IUT2, Univ. Grenoble-Alpes

N. Gesbert is responsible of the L3-level course ‘logique pour l’informatique’ (25 apprentices) and of the M1-level course ‘construction d’applications Web’ (72 students).

P. Genevès is co-responsible of the Master-level course ‘Semantic Web: from XML to OWL’ in the Mosig, Univ. Grenoble Alpes.

C. Roisin is responsible of the Licence Professionnelle en Alternance ‘Administration et Sécurité des Systèmes et des Réseaux’, L3, IUT2, Univ. Grenoble-Alpes (20 apprentices).

C. Roisin is responsible of the L1-level course ‘Architecture des réseaux’ (150 students).

9.2.2. Supervision

PhD: Abdullah Abbas, Static Analysis of SPARQL Queries with ShEx Schema Constraints, University Grenoble Alpes, defended on November 6th 2017, supervised by Pierre Genevès and Cécile Roisin.

PhD: Thibaud Michel, On Mobile Augmented Reality Applications based on Geolocation, University Grenoble Alpes, defended on November 10th 2017, supervised by Pierre Genevès and Nabil Layaïda and Hassen Fourati.

PhD in progress: Louis Jachiet, Foundations for the Analysis and Distributed Evaluation of SPARQL Queries, started in Sept. 2014, supervised by Pierre Genevès and Nabil Layaïda.

PhD in progress: Raouf Kerkouche, Optimized Predictive Analytics with Big Medical Data, started in Oct. 2017, supervised by Pierre Genevès and Claude Castelluccia.

PhD in progress: Fateh Boulmaiz, Extensions of the SPARQLGX System, started in Nov. 2017, supervised by Pierre Genevès and Nabil Layaïda.

9.2.3. Juries

- P. Genevès was member of the jury for the PhD defense of Carlyna Bondiombouy entitled “Traitement de requêtes dans les systèmes multistores” and defended on July 12th, 2017 in Montpellier.
- N. Layaïda was member and president of the Jury for the Habilitation thesis from University Grenoble-Alpes of Jérôme Malick entitled “the Variational-analysis look at combinatorial optimization and other selected topics in optimization” defended on January 26th, 2017 in Grenoble.
- N. Layaïda was member and president of the Jury for the PhD thesis from University Grenoble-Alpes of Thomas Capelle entitled “Development of optimization methods for land-use and transportation models” defended on April 3rd, 2017 in Grenoble.

10. Bibliography

Major publications by the team in recent years

- [1] M. BOSCH, P. GENEVÈS, N. LAYAÏDA. *Reasoning with Style*, in "International Joint Conference On Artificial Intelligence (IJCAI 2015)", Buenos Aires, Argentina, July 2015, <https://hal.inria.fr/hal-01149248>
- [2] S. J. GAY, N. GESBERT, A. RAVARA, V. T. VASCONCELOS. *Modular session types for objects*, in "Logical Methods in Computer Science", December 2015, vol. 4, n^o 12, 76 p. [DOI : 10.2168/LMCS-11(4:12)2015], <https://hal.archives-ouvertes.fr/hal-00700635>
- [3] P. GENEVÈS, N. GESBERT. *XQuery and Static Typing: Tackling the Problem of Backward Axes*, in "ICFP (International Conference on Functional Programming)", Vancouver, Canada, ACM SIGPLAN, August 2015 [DOI : 10.1145/2784731.2784746], <https://hal.inria.fr/hal-01082635>
- [4] P. GENEVÈS, N. LAYAÏDA, A. SCHMITT, N. GESBERT. *Efficiently Deciding μ -calculus with Converse over Finite Trees*, in "ACM Transactions on Computational Logic", March 2015, vol. 16, n^o 2, 41 p. [DOI : 10.1145/2724712], <https://hal.inria.fr/hal-00868722>
- [5] P. GENEVÈS, A. SCHMITT. *Expressive Logical Combinators for Free*, in "International Joint Conference on Artificial Intelligence (IJCAI 2015)", Buenos Aires, Argentina, July 2015, <https://hal.inria.fr/hal-00868724>
- [6] N. GESBERT, P. GENEVÈS, N. LAYAÏDA. *A Logical Approach To Deciding Semantic Subtyping*, in "ACM Transactions on Programming Languages and Systems (TOPLAS)", 2015, vol. 38, n^o 1, 31 p. [DOI : 10.1145/2812805], <https://hal.inria.fr/hal-00848023>
- [7] T. MICHEL, H. FOURATI, P. GENEVÈS, N. LAYAÏDA. *A Comparative Analysis of Attitude Estimation for Pedestrian Navigation with Smartphones*, in "Indoor Positioning and Indoor Navigation", Banff, Canada, October 2015, vol. 2015 International Conference on Indoor Positioning and Indoor Navigation, 10 p. , <https://hal.inria.fr/hal-01194811>

Publications of the year

Doctoral Dissertations and Habilitation Theses

- [8] A. ABBAS. *Static Analysis of Semantic Web Queries with ShEx Schema Constraints*, Université Grenoble - Alpes, November 2017, <https://hal.inria.fr/tel-01673074>
- [9] T. MICHEL. *On Mobile Augmented Reality Applications based on Geolocation*, Université Grenoble Alpes, November 2017, <https://hal.inria.fr/tel-01651589>

International Conferences with Proceedings

- [10] A. ABBAS, P. GENEVÈS, C. ROISIN, N. LAYAÏDA. *SPARQL Query Containment with ShEx Constraints*, in "ADBIS 2017 - 21st European Conference on Advances in Databases and Information Systems", Nicosia, Cyprus, M. KIRIKOVA, K. NØRVÅG, G. A. PAPAPOPOULOS (editors), LNCS - Lecture Notes in Computer Science, Springer, September 2017, vol. 10509, Submitted, <https://hal.inria.fr/hal-01414509>
- [11] A. ABBAS, P. GENEVÈS, C. ROISIN, N. LAYAÏDA. *Optimising SPARQL Query Evaluation in the Presence of ShEx Constraints*, in "BDA 2017 - 33ème conférence sur la « Gestion de Données — Principes, Technologies et Applications »", Nancy, France, November 2017, pp. 1-12, <https://hal.inria.fr/hal-01673067>
- [12] A. AMARILLI, P. BOURHIS, L. JACHET, S. MENGEL. *A Circuit-Based Approach to Efficient Enumeration*, in "ICALP 2017 - 44th International Colloquium on Automata, Languages, and Programming", Varsovie, Poland, I. CHATZIGIANNAKIS, P. INDYK, A. MUSCHOLL (editors), July 2017, pp. 1-15 [DOI : 10.4230/LIPICS.ICALP.2017.111], <https://hal.inria.fr/hal-01639179>
- [13] D. GRAUX, L. JACHET, P. GENEVÈS, N. LAYAÏDA. *Une classification expérimentale multi-critère des évaluateurs SPARQL répartis*, in "BDA 2017 - 33ème conférence sur la « Gestion de Données — Principes, Technologies et Applications »", Nancy, France, November 2017, pp. 1-2, <https://hal.inria.fr/hal-01673114>
- [14] T. MICHEL, P. GENEVÈS, H. FOURATI, N. LAYAÏDA. *On Attitude Estimation with Smartphones*, in "IEEE International Conference on Pervasive Computing and Communications", Kona, United States, March 2017, Accepted for the International Conference on Pervasive Computing and Communications (PerCom 2017), Mar 2017, Kona, United States, <https://hal.inria.fr/hal-01376745>

Other Publications

- [15] A. ABBAS, P. GENEVÈS, C. ROISIN, N. LAYAÏDA. *Selectivity Estimation for SPARQL Triple Patterns with Shape Expressions*, December 2017, working paper or preprint, <https://hal.inria.fr/hal-01673013>
- [16] P. GENEVÈS, T. CALMANT, N. LAYAÏDA, M. LEPELLEY, S. ARTEMOVA, J.-L. BOSSON. *Predicting At-Risk Patient Profiles from Big Prescription Data*, December 2017, working paper or preprint, <https://hal.inria.fr/hal-01517087>
- [17] D. GRAUX, P. GENEVÈS, N. LAYAÏDA. *HAP: Building Pipelines with Heterogeneous Data and Hive*, January 2017, working paper or preprint, <https://hal.inria.fr/hal-01436850>
- [18] D. GRAUX, P. GENEVÈS, N. LAYAÏDA. *SPARUB: SPARQL UPDATE Benchmark*, May 2017, working paper or preprint, <https://hal.inria.fr/hal-01523496>

-
- [19] D. GRAUX, L. JACHET, P. GENEVÈS, N. LAYAÏDA. *The SPARQLGX System for Distributed Evaluation of SPARQL Queries*, October 2017, working paper or preprint, <https://hal.inria.fr/hal-01621480>
- [20] H. IM, P. GENEVÈS, N. GESBERT, N. LAYAÏDA. *Backward Type Inference for XML Queries*, March 2017, working paper or preprint, <https://hal.inria.fr/hal-01497857>
- [21] L. JACHET, P. GENEVÈS, N. GESBERT, N. LAYAÏDA. *Extending the SPARQL Algebra for the optimization of Property Paths*, December 2017, working paper or preprint, <https://hal.inria.fr/hal-01673025>
- [22] L. JACHET, P. GENEVÈS, N. LAYAÏDA, N. GESBERT. *Extending the SPARQL Algebra for the optimization of Property Paths*, November 2017, pp. 1-2, BDA 2017 - 33ème conférence sur la « Gestion de Données », Poster, <https://hal.archives-ouvertes.fr/hal-01647638>
- [23] L. JACHET, P. GENEVÈS, N. LAYAÏDA. *Optimizing sparql query evaluation with a worst-case cardinality estimation based on statistics on the data*, May 2017, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01524387>
- [24] T. MICHEL, P. GENEVÈS, H. FOURATI, N. LAYAÏDA. *Attitude Estimation with Smartphones*, November 2017, working paper or preprint, <https://hal.inria.fr/hal-01650142>