IN PARTNERSHIP WITH:
CNRS

Institut polytechnique de Grenoble

Université Grenoble Alpes

Activity Report 2016

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

Project-Team TYREX

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

Keywords:

Computer Science and Digital Science:

2.1.1. - Semantics of programming languages
2.1.3. - Functional programming
2.1.7. - Distributed programming
2.1.10. - Domain-specific languages
2.2.1. - Static analysis
2.2.4. - Parallel architectures
2.4. - Verification, reliability, certification
3.1.1. - Modeling, representation
3.1.2. - Data management, quering and storage
3.1.3. - Distributed data
3.1.6. - Query optimization
3.1.7. - Open data
3.1.8. - Big data (production, storage, transfer)
3.2.1. - Knowledge bases
3.2.2. - Knowledge extraction, cleaning
3.3.3. - Big data analysis
3.4. - Machine learning and statistics
5.6. - Virtual reality, augmented reality
6.3.2. - Data assimilation
6.3.3. - Data processing
7.4. - Logic in Computer Science
8.1. - Knowledge
8.7. - AI algorithmics

Other Research Topics and Application Domains:

6.1. - Software industry
6.3.1. - Web
6.5. - Information systems
8.2. - Connected city
9.4.1. - Computer science
9.4.5. - Data science
9.7.2. - Open data
9.9. - Risk management
9.9.2. - Financial risks

1. Members

Research Scientists

Nabil Layaïda [Team Leader, Inria, Senior Researcher, HDR]
Pierre Genevès [CNRS, Researcher, HDR]

Faculty Members
Nils Gesbert [Grenoble INP, Associate Professor]
Cécile Roisin [Université Grenoble-Alpes, Professor, HDR]

Engineers
Thomas Calmant [Inria, from Jun 2016]
Guillaume Dupraz-Canard [Inria, until Jun 2016]
Mathieu Razafimahazo [Université Grenoble-Alpes, until Apr 2016]

PhD Students
Abdullah Abbas [Université Grenoble-Alpes]
Damien Graux [Inria]
Louis Jachiet [ENS Paris]
Thibaud Michel [Université Grenoble-Alpes]

Administrative Assistants
Marion Ponsot [Inria, until Sep 2016]
Helen Pouchot-Rouge-Blanc [Inria, since Oct 2016]

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 JSonIq 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. Benchmarks Attitude Smartphones

KEYWORDS: Experimentation - Motion analysis - Sensors - Performance analysis - 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: Thibaud Michel, Hassen Fourati, Pierre Geneves and Nabil Layaida
- Partner: GIPSA-Lab
- Contact: Pierre Genevès, Thibaud Michel
- URL: http://tyrex.inria.fr/mobile/benchmarks-attitude/

5.2. CSS Analyzer

FUNCTIONAL DESCRIPTION

This software now consists in two distinct prototypes: two static analyzers (with a different purpose) that share a common compiler for CSS. The first prototype is used for bug detection and verification of a cascading style sheet (CSS) file. It involves a compiler for CSS rules (and in particular selectors) into logical formulas, adapted for the semantics of CSS (see the initial WWW'12 paper). The second prototype performs automated refactoring for size reduction of CSS style sheets. It reuses the first compiler and the logical solver for detecting which rules can be refactored and how. It implements various optimisation techniques (like early pruning), for the purpose of dealing with large-size real CSS files. This prototype reduces the size of CSS files found in the most popular websites (such as CNN, facebook, Google Sites, Apple, etc.) by up to 30

- Participants: Pierre Geneves, Nabil Layaida and Marti Bosch Padros
- Contact: Pierre Geneves
- URL: http://tyrex.inria.fr/websolver/

5.3. RDFHive

KEYWORDS: Hadoop - RDF - SPARQL
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 RDFHive: a simple implementation of a distributed RDF datastore benefiting from Apache Hive. RDFHive is designed to leverage existing Hadoop infrastructures for evaluating SPARQL queries. RDFHive relies on a translation of SPARQL queries into SQL queries that Hive is able to evaluate.

Technically, RDFHive directly evaluates SPARQL queries i.e. there is no preprocessing step, indeed an RDF triple file is seen by Hive as a three-column table. Thus, the bash translator simply translates SPARQL queries according to this scheme. This method has two advantages: first, creating a database is very fast, second, since the upfront investment is light, RDFHive is an interesting tool to evaluate a few SPARQL queries at once.

- Participants: Damien Graux, Nabil Layaida and Pierre Geneves
- Contact: Pierre Genevès, Damien Graux
- URL: https://github.com/tyrex-team/rdfhive

5.4. Tree Reasoning Solver

SCIENTIFIC DESCRIPTION

The tree reasoning solver is a software tool which implements research advances in tree logics, and in the analysis of query and programming languages for the web. The core algorithm is a satisfiability solver of an expressive tree logic. The underlying logic is very expressive: it is a μ-calculus variant for finite trees, which is MSO-complete, and equipped with additional features (converse modalities, nominals, logical combinators...)

The decision procedure has an optimal worst-case complexity, and its implementation performs quite well in practice. This allows efficient reasoning with tree structures. In particular, it opens the way for solving a wide variety of problems that require reasoning with very large sets of trees.

Initially developed for the analysis of XML/XPath queries, this tool has been extended by the team to support more general query analysis, reasoning with schema constraints, with functions, and with domain specific languages such as cascading style sheets.

- Participants: Pierre Geneves, Nabil Layaida and Nils Gesbert
- Contact: Pierre Geneves
- URL: http://tyrex.inria.fr/websolver/

5.5. XQuery Type-Checker

SCIENTIFIC DESCRIPTION

This prototype implements a sound static type-system for an XQuery core. The type language supported is a large subset of RelaxNG+Schematron, which is novel in static type checking. It also supports the static typing of backward axes, which is not supported by any other system nor the XQuery recommendation. Our type checker successfully verifies complex programs for which existing type-checkers (either known from the literature or those developed in commercial software) fail by reporting false alarms. One major benefit is to allow the cost of validation to be deferred from runtime to compile-time (once only). This prototype is implemented in Scala and interacts with the solver by issuing externals calls for deciding complex subtyping relations.

- Participants: Pierre Geneves, Nabil Layaida and Nils Gesbert
- Contact: Pierre Geneves
- URL: http://tyrex.inria.fr/websolver/

5.6. 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.

- Participants: Damien Graux, Louis Jachiet, Nabil Layaida and Pierre Geneves
- Contact: Pierre Genevès, Damien Graux
- URL: https://github.com/tyrex-team/sparqlgx

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 investigated the precision of attitude estimation algorithms in the context of commodity smartphones carried by pedestrians. We considered eight typical motions (such as texting, phoning, running, etc.) with various impacts on external accelerations, as well as the presence/absence of magnetic perturbations typically found in indoor environments. We systematically analyzed, compared and evaluated eight state-of-the-art algorithms (and their variants). We precisely quantified the attitude estimation error obtained with each technique, owing to the use of a precise ground truth obtained with a motion capture system (the Inria Kinovis platform). We made our benchmark available (see Sec. 5.1 above) and payed attention to the reproducibility of results. We analyzed and discussed the obtained results and reported on lessons learned [7] [17]. We also presented a new technique which helps in improving precision by limiting the effect of magnetic perturbations with all considered algorithms.

6.2. Efficient Distributed Evaluation of SPARQL Queries

- **Context:** SPARQL is the standard query language for retrieving and manipulating data represented in the Resource Description Framework (RDF). SPARQL constitutes one key technology of the semantic web and has become very popular since it became an official W3C recommendation.

  The construction of efficient SPARQL query evaluators faces several challenges. First, RDF datasets are increasingly large, with some already containing more than a billion triples. To handle efficiently this growing amount of data, we need systems to be distributed and to scale. Furthermore, semantic data often have the characteristic of being dynamic (frequently updated). Thus being able to answer quickly after a change in the input data constitutes a very desirable property for a SPARQL evaluator.

- **Contributions:** First of all, to constitute a common basis of comparative analysis, we evaluated on the same cluster of machines various SPARQL evaluation systems from the literature [15]. These experiments led us to point several observations: (i) the solutions have very different behaviors; (ii) most of the benchmarks only use temporal metrics and forget other ones e.g. network traffic. That is why we proposed a larger set of metrics; and thanks to a new reading grid based on 5 features, we proposed new perspectives which should be considered when developing distributed SPARQL evaluators.

  Second, we developed and shared several distributed SPARQL evaluators which take into account these new considerations we introduced:

  – A SPARQL evaluator named SPARQLGX (see Sec. 5.6): an implementation of a distributed RDF datastore based on Apache Spark. SPARQLGX is designed to leverage existing Hadoop infrastructures for evaluating SPARQL queries. It relies on a translation of SPARQL queries into executable Spark code that adopts evaluation strategies according to the storage method used and statistics on data.

    In [12], [11], [8], [13], we showed that SPARQLGX makes it possible to evaluate SPARQL queries on billions of triples distributed across multiple nodes, while providing attractive performance figures. We reported on experiments which show how SPARQLGX compares to related state-of-the-art implementations and we showed that our approach scales better than these systems in terms of supported dataset size. With its simple design, SPARQLGX represents an interesting alternative in several scenarios.

  – Two SPARQL direct evaluators i.e. without a preprocessing phase: SDE (stands for Sparqlgx Direct Evaluator) lays on the same strategy than SPARQLGX but the translation process is modified in order to take the orign data files as argument. RDFHive (see Sec. 5.3) evaluates translated SPARQL queries on top of Apache Hive which is a distributed relational data warehouse based on Apache Hadoop.

6.3. An Efficient Translation from a modal μ-Calculus with Converse to Tree Automata

In [16], we presented a direct translation from a sub-logic of μ-calculus to non-deterministic binary automata of finite trees. The logic is an alternation-free modal μ-calculus, restricted to finite trees and where formulae are cycle-free. This logic is expressive enough to encode significant fragments of query languages (such as Regular XPath). The size of the generated automaton (the number of transitions) is bounded by 2^n where n is the size of a Fischer-Ladner closure of the formula. This is an improvement over previous translations in 2^{n^2}. We have implemented our translation. In practice, our prototype effectively decides static analysis problems that were beyond reach, such as the XPath containment problem with DTDs of significant size.

6.4. SPARQL Query Containment with ShEx Constraints

ShEx (Shape Expressions) is a language for expressing constraints on RDF graphs. In [14], we considered the problem of SPARQL query containment in the presence of ShEx constraints. We first investigated the complexity of the problem according to the fragments considered for SPARQL queries and for ShEx constraints. In particular, we showed that the complexity of SPARQL query containment remains the same with or without ShEx constraints. We developed two radically different approaches for solving the problem and we evaluated them. The first approach relies on the joint use of a ShEx validator and a tool for checking query containment without constraints. In a second approach, we showed how the problem can be solved by a reduction to a fragment of first-order logic with two variables. This alternative approach allows to take advantage of any of the many existing FOL theorem provers in this context. We evaluated how the two approaches compare experimentally, and reported on lessons learned. To the best of our knowledge, this is the first work addressing SPARQL query containment in the presence of ShEx constraints.

6.5. XQuery Static Type-Checking

In the context of our ongoing work on XQuery static type-checking [3], we extended our type system and improved the associated software accordingly (see Sec. 5.5 and 5.4). The type language it is based on is now a subset of RelaxNG+Schematron (instead of DTDs), which is novel in the context of static typing: Schematron is normally used to validate a document after it has been generated, whereas our system is able to ensure statically that a program will always generate a valid document.

Schematron constraints present the advantage of describing some properties in a very concise way compared to schema languages based on regular tree types, e.g. it allows writing in one line that nested anchors are forbidden in HTML, a constraint which appears in the specification but not in the formal DTD schema because of the verbosity it would involve.

7. Partnerships and Cooperations

7.1. 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.

7.2. National Initiatives

7.2.1. Investissements d'avenir

Datalyse

Title: Entrepôt Intelligent pour Big Data hétérogènes. Investissements d'Avenir Développement de l'Economie Numérique.

Call: Cloud Computing, num 3 – Big Data.

Duration: May 2013 - November 2016

Coordinator: Business & Decision Eolas

Others partners: Groupement des Mousquetaires, Inria Saclay (OAK EPC), LIG (Hadas and Erods teams), LIRMM (Montpellier), LIFL (Lille).

See also: http://www.datalyse.fr/

Abstract: Project Datalyse aims at designing and deploying an infrastructure for big data storage, collection, certification, integration, categorisation, enrichment and sharing over very large heterogeneous data sets. It relies on an industrial platform, to be made available on the cloud, and focuses on three flagship applications, showcasing three uses of big data over different data sets:

- Data-Center Monitoring: The goal of this application is to provide features such as traceability, reporting, optimisation and analysis of abnormal behaviour regarding energy efficiency and security issues. The application will be built with an existing application called ScopeBR (Eolas) and will be deployed in two different green data centers, those of Eolas and GDF SUEZ.
- 'Territoire de données ouvertes et liées': This application aims at extracting and provisioning public open data collected from the city of Grenoble and its suburbs. The goal is to make public data available to third-party application developers and to federate local actors around a single platform.
- Real-time Business Intelligence for the management and processing of points of sale: this application will focus on real-time data analytics and will be deployed within 'Groupement des Mousquetaires' in support of their business intelligence platforms.

7.2.2. 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.

7.2.3. 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.

7.3. European Initiatives

7.3.1. Collaborations in European Programs, Except FP7 & H2020

Program: COST

Project acronym: BETTY

Project title: Behavioural Types for Reliable Large-Scale Software Systems

Duration: October 2012 – October 2016

Coordinator: Professor Simon Gay, University of Glasgow, UK

Other partners: Bosnia and Herzegovina, Croatia, Cyprus, Denmark, Estonia, fYR Macedonia, Germany, Greece, Ireland, Italy, Lithuania, Malta, Netherlands, Norway, Poland, Portugal, Romania, Serbia, Spain, Sweden, United Kingdom

Abstract: Modern society is increasingly dependent on large-scale software systems that are distributed, collaborative and communication-centred. Correctness and reliability of such systems depend on compatibility between components and services that are newly developed or may already exist. The consequences of failure are severe, including security breaches and unavailability of essential services. Current software development technology is not well suited to producing these large-scale systems, because of the lack of high-level structuring abstractions for complex communication behaviour.

This Action will use behavioural type theory as the basis for new foundations, programming languages, and software development methods for communication-intensive distributed systems. Behavioural type theory encompasses concepts such as interfaces, communication protocols, contracts, and choreography. As a unifying structural principle it will transform the theory and practice of distributed software development.

The significance of behavioural types has been recognised world-wide during the last five years. European researchers are internationally leading. There is an urgent need for European co-ordination to avoid duplication of effort, facilitate interactions among research groups, and ensure that the field proceeds efficiently from academic research to industrial practice. This Action will provide the co-ordination layer and leverage the efforts of European researchers, to increase the competitiveness of the European software industry.

See also: http://behavioural-types.eu

7.4. International Research Visitors

7.4.1. Internships

Jakob Zietsch from Technische Universität München visited the team from March to July to work on geolocalization with smartphones based on fingerprinting.

8. Dissemination

8.1. Promoting Scientific Activities

8.1.1. Scientific Events Organisation

8.1.1.1. Member of the Organizing Committees

- C. Roisin is a member of the steering committee of the ACM Symposium on Document Engineering. (until Sept. 2016)

8.1.2. Scientific Events Selection

8.1.2.1. Member of the Conference Program Committees

- P. Genevès has been external review committee member for the 21st ACM SIGPLAN International Conference on Functional Programming (ICFP'16).
- P. Genevès has been program committee member for the 16th ACM Symposium on Document Engineering (DocEng'16).
- C. Roisin has been program committee member for the 16th ACM Symposium on Document Engineering (DocEng'16).

8.1.2.2. Reviewer

- P. Genevès has been a referee for the '32ème Conférence sur la Gestion de Données - Principes, Technologies et Applications' (BDA 2016)

8.1.3. Journal

8.1.3.1. Reviewer - Reviewing Activities

- N. Gesbert has been a referee for Logical Methods in Computer Science (LMCS).
- C. Roisin has been a referee for Journal of Visual Languages & Computing, Elsevier (JVLC).

8.1.4. Leadership within the Scientific Community

- C. Roisin is member of the section 27 of the CNU (Conseil National des Universités).

8.1.5. Scientific Expertise

Oppidoc

Title: Choice of Methods and Algorithms for XQuery Static Analyses in the Oppidum Framework

Duration: November - December 2016

Coordinator: Pierre Genevès

Abstract: The Oppidoc startup develops 'Oppidum': an XQuery web application framework which simplifies the development of XML-REST-XQuery applications (XRX) with the full XML technology stack (XQuery, XSLT, native XML database). It relies on a RESTful approach and on a well defined application model using concepts (routes, conventions, pipelines) popularized in other frameworks such as Ruby On Rails, Orbeon Forms and more recently Express on nodejs. Our collaboration concerns a study about the introduction of advanced static analyses techniques in the Oppidum development process.

8.1.6. Research Administration

- P. Genevès is co-responsible of the doctoral school of Grenoble University for Computer Science (around 400 PhD students).
- P. Genevès is a permanent member of the committee in charge of hiring research engineers at Inria Grenoble - Rhône-Alpes research center.
- N. Layaïda is 'référent 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 experts pool (selection committee) of the minalogic competitive cluster.
- N. Layaïda is a permanent member of the jury in charge of evaluation harmonisation of the Master of Science in Informatics at Grenoble.

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

Master : P. Genevès, 'Semantic web: from XML to OWL', 36 h, M2, Univ. Grenoble-Alpes

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, 5 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 (15 apprentices).

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

8.2.2. Supervision

PhD : D. Graux, On the Efficient Distributed Evaluation of SPARQL Queries, Université Grenoble-Alpes, 15 December 2016, N. Layaïda and P. Genevès

PhD in progress : A. Abbas, Web query rewriting for heterogeneous data sources, since October 2014, N. Layaïda and P. Genevès

PhD in progress : T. Michel, Mobile Augmented Reality Applications for Smart Cities, since October 2014, P. Genevès, N. Layaïda and H. Fourati

PhD in progress : L. Jachiet, Reasoning with NoSQL Data Flows in Massively Parallel Systems, since October 2014, N. Layaïda and P. Genevès

8.2.3. Juries

- C. Roisin has been referee and jury member of the Mira Sarkis PhD thesis, Telecom ParisTech (oct 2016).

9. 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] D. GRAUX. *On the Efficient Distributed Evaluation of SPARQL Queries*, Université Grenoble Alpes, December 2016, https://hal.inria.fr/tel-01405319

Articles in International Peer-Reviewed Journals

[9] D. ANCONA, V. BONO, M. BRAVETTI, J. CAMPOS, G. CASTAGNA, P.-M. DENIÉLOU, S. J. GAY, N. GESBERT, E. GIACHINO, R. HU, E. B. JOHNSEN, F. MARTINS, V. MASCARDI, F. MONTESI, R. NEYKOVA, N. NG, L. PADOVANI, V. T. VASCONCELOS, N. YOSHIDA. *Behavioral Types in Programming Languages*, in "Foundations and Trends in Programming Languages", July 2016, vol. 3, n^{o} 2-3, pp. 95-230 [*DOI :* 10.1561/2500000031], https://hal.inria.fr/hal-01348054

International Conferences with Proceedings

[10] D. GRAUX, P. GENEVÈS, N. LAYAÏDA. *Smart Trip Alternatives for the Curious*, in "15th International Semantic Web Conference (ISWC 2016 demo paper)", Kobe, Japan, October 2016, https://hal.inria.fr/hal-01342030

[11] D. GRAUX, L. JACHIET, P. GENEVÈS, N. LAYAÏDA. *SPARQLGX in Action: Efficient Distributed Evaluation of SPARQL with Apache Spark*, in "15th International Semantic Web Conference (ISWC 2016 demo paper)", Kobe, Japan, October 2016, https://hal.inria.fr/hal-01358125

[12] D. GRAUX, L. JACHIET, P. GENEVÈS, N. LAYAÏDA. *SPARQLGX: Efficient Distributed Evaluation of SPARQL with Apache Spark*, in "The 15th International Semantic Web Conference", Kobe, Japan, October 2016 [*DOI :* 10.1007/978-3-319-46547-0_9], https://hal.inria.fr/hal-01344915

National Conferences with Proceedings

[13] D. GRAUX, L. JACHIET, P. GENEVÈS, N. LAYAÏDA. *SPARQLGX : Une Solution Distribuée pour RDF Traduisant SPARQL vers Spark*, in "BDA 2016 - 32ème Conférence sur la Gestion de Données - Principes, Technologies et Applications", Poitiers, France, November 2016, https://hal.inria.fr/hal-01412035

Other Publications

[14] A. ABBAS, P. GENEVE`S, C. ROISIN, N. LAYAÏDA. *SPARQL Query Containment with ShEx Constraints*, October 2016, Submitted, https://hal.inria.fr/hal-01414509

[15] D. GRAUX, L. JACHIET, P. GENEVE`S, N. LAYAÏDA. *A Multi-Criteria Experimental Ranking of Distributed SPARQL Evaluators*, October 2016, Submitted, https://hal.inria.fr/hal-01381781

[16] L. JACHIET, P. GENEVÈS, N. LAYAÏDA. *An efficient translation from a modal μ-calculus with converse to tree automata*, September 2016, Submitted, https://hal.inria.fr/hal-01117830

[17] T. MICHEL, P. GENEVE`S, H. FOURATI, N. LAYAÏDA. *On Attitude Estimation with Smartphones*, September 2016, Accepted for the International Conference on Pervasive Computing and Communications (PerCom 2017), Mar 2017, Kona, United States, https://hal.inria.fr/hal-01376745