IN PARTNERSHIP WITH:
CNRS

Institut polytechnique de Grenoble

Université Joseph Fourier (Grenoble)

Université Pierre Mendes-France (Grenoble)

Activity Report 2015

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
- 2.1.10. - Domain-specific languages
- 2.1.3. - Functional programming
- 2.2.1. - Static analysis
- 2.4. - 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
- 3.2.3. - Inference
- 3.2.4. - Semantic Web
- 3.2.5. - Ontologies
- 3.3.1. - On-line analytical processing
- 3.3.3. - Big data analysis
- 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
- 6.4. - Internet of things
- 6.5. - Information systems
- 7.2. - Smart travel
- 8.1.2. - Sensor networks for smart buildings
- 8.2. - Connected city
- 9.4.1. - Computer science
- 9.4.5. - Data science
- 9.7.2. - Open data
- 9.9. - Risk management

1. Members

Research Scientists
Nabil Layaida [Team leader, Inria, Senior Researcher, HdR]
Pierre Geneves [CNRS, Researcher, HdR]

Faculty Members
Nils Gesbert [INP Grenoble, Associate Professor]
Cecile Roisin [Univ. Grenoble II, Professor, HdR]

Engineers
Guillaume Dupraz-Canard [Inria]
Mathieu Razafimahazo [Univ. Grenoble I]

PhD Students
Abdullah Abbas [Univ. Grenoble I]
Damien Graux [Univ. Grenoble I]
Nicola Guido [Inria]
Louis Jachiet [ENS Paris]
Thibaud Michel [Univ. Grenoble I]

Post-Doctoral Fellow
Hyeonseung Im [Inria, until Jan 2015]

Visiting Scientist
Joel Ferreira Dos Santos [Federal Fluminense University, Brazil, until Oct 2015]

Administrative Assistant
Marion Ponsot [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 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. CSS Analyzer

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% while preserving their semantics [18].

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

5.2. Interactive eXtensible Engine (IXE)

Interactive eXtensible Engine
FUNCTIONAL DESCRIPTION

PDRTrack is a localization utility running on iOS or Android smartphones used for recording and playing data sets (accelerometer, gyroscope, barometer and magnetometer values) to study the effect of different pedometer and map matching parameters on indoor and outdoor localization accuracy. This application uses the PDR library, written in C++, which provides the user's location in real time based on the interpretation of mobile phone sensors. Three main modules have been designed to build this localization system:

- a pedometer that estimates the distance the user has walked and his speed
- 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)
- a map-matching algorithm that provides location estimates on a given OpenStreetMap description and the current user's trajectory

The PDR library is a central component of the VENTURI project. It has been used for applications such guiding a visually impaired people. Others partners have used this localisation system for retrieving a scale factor needed for the computer vision part (i.e SLAM).

GPS navigation systems, when used in an urban environment, are limited in precision and can only give instructions at the level of the street and not of the pavement or corridor. GPS is also limited to outdoor navigation and requires some transitioning system when switching to indoor navigation.

PDRTrack is embedded in IXE. IXE is an urban pedestrian navigation system based on Inertial Measurement Units (IMU) and running on mobile phones with onboard geographic data and a routing engine. IXE allows augmented reality queries on customised embedded geographical data. Queries on route nodes or POIs, on ways and relations are predefined for efficiency and quality of information. Following a web paradigm, IXE can be seen as web browser for XML documents describing navigation networks. by using the micro-format concept, one can define inside OpenStreetMap a complex format for pedestrian navigation networks allowing navigation at the level of pavements or corridors.

The big advantage of IXE is that it relies on a standard OpenStreetMap editor called JOSM to create navigation networks and augmented reality content. IXE browser reads OSM documents and produces from them visible or audible navigation information. IXE is composed of three engines, one for dead-reckoning navigation, one for interactive audio and the last one for Augmented Reality visual information.

- Participants: Nabil Layaïda, Pierre Genevès, Thibaud Michel and Mathieu Razafimahazo
- Contact: Nabil Layaïda
- URL: http://tyrex.inria.fr/mobile/

5.3. XML Reasoning Solver

XML Reasoning Solver
FUNCTIONAL DESCRIPTION

The XML Reasoning Solver is a tool for the static analysis of queries and schemas based on our theoretical advances [12]. It allows automated verification of properties that are expressed as logical formulas over trees. A logical formula may for instance express structural constraints or navigation properties (like e.g. path existence and node selection) in finite trees.

The reasoner is built on top of a finite tree logic solver for a new modal logic equipped with recursion and backward axes. The solver is very fast in practice and uses symbolic techniques (Binary Decision Diagrams). The solver has been recently extended to support functions, parametric functions and polymorphic subtyping. One notable difficulty was to elaborate many advanced optimizations with symbolic implementation techniques. The logical solver significantly advances the state of the art. In particular, it is the first implementation that effectively solves the query containment problem for a large fragment of the XPath query language. It supports all navigation axes and regular tree constraints. Although researchers had studied XPath satisfiability before, such prior works were either unimplementable or deemed to explode even for tiny examples. As of 2014, it is still the only implementation actually capable of solving this problem in practice for real world instances.

The reasoner includes compilers and various static analyzers for web query and schema languages. This includes compilers for XPath, for XML schemas (DTDs, XML Schemas, Relax NGs) into logical formulas, parsers, benchmarks, and libraries for automated testing. Various difficulties reside in the compilation of real-world queries, including compiling XPath queries into fixed-point logics, developing specific implementation techniques in order to avoid worst case blow-ups as much as possible when e.g. supporting unordered XML attributes among (ordered) XML elements, etc. The reasoner also generates counter-examples that allow program defects to be reproduced independently from the analyzer.

The off-line version of the solver (with a native library) is fast and up-to-date with the latest advances. We developed and deployed an interactive web interface to make the solver available to the international scientific community. For this purpose, we redesigned the libraries used for the manipulation of binary decision diagrams (BDDs) so that they could used in a fully concurrent and multithreaded manner. This is in order to allow several instances of the logical solver to run concurrently for several users on a web server (GWT-based), while decreasing performance as less as possible.
The reasoner helps us to guide and validate our approach. We continue to develop, maintain and use it on an almost-daily basis.

- Participants: Pierre Genevès, Nabil Layaïda, Louis Jachiet and Nils Gesbert
- Contact: Pierre Genevès
- URL: http://tyrex.inria.fr/websolver/

5.4. XQuery type-checker

XQuery type-checker
FUNCTIONAL DESCRIPTION

This prototype implements a sound static type-system for XQuery, which, as of december 2014, is the most precise type system known for XQuery. It supports the static typing of backward axes that no other does nor is supported in the XQuery recommendation. It also includes precise typing for conditional statements which is challenging as such statement are usually sensitive to the program context. 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. This prototype is described in preprint [20].

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

5.5. claireCourseMaker

Claire CourseMaker Library
FUNCTIONAL DESCRIPTION

The goal of the ClaireCourseMaker is to provide direct and visual editing tools for structuring, annotating and timeline-based authoring of continuous content such as audio or video. I it is mainly devoted to the synchronisation and layout of pedagogical material (video, slides, chaptering, etc.) and enables the incorporation of rich media content in MOOCs. The underlying technology is based on Web standards and relies on the open source JavaScript Popcorn library and Popcorn Maker web application developed by the Mozilla Foundation. The tool is a wysiwyg web-based authoring tool which benefits from the generic features of Popcorn and offers structuring methods such chaptering and container-based synchronisation.

ClaireCourseMaker is the direct follow-up tool of the Timesheet library developed in the project. Timesheet library is a cross-browser JavaScript implementation for scheduling the dynamic behaviour of HTML5 content. It uses and provides a reference implementation for declarative synchronisation markup such as SMIL Timing and Synchronization and SMIL Timesheets.

ClaireCourseMaker has been developed in collaboration with the OpenClassrooms company in the context of the Claire project.

- Participants: Cécile Roisin, Nabil Layaïda and Nicolas Hairon
- Contact: Cécile Roisin
- URL: https://github.com/NicolasHairon/popcorn.webmaker.org

6. New Results

6.1. Expressive Logical Combinators

A popular technique for the analysis of web query languages relies on the translation of queries into logical formulas. These formulas are then solved for satisfiability using an off-the-shelf satisfiability solver. A critical aspect in this approach is the size of the obtained logical formula, since it constitutes a factor that affects the combined complexity of the global approach. In this work [21], we present logical combinators whose benefit is to provide an exponential gain in succinctness in terms of the size of the logical representation. This opens the way for solving a wide range of problems such as satisfiability and containment for expressive query languages in exponential-time, even though their direct formulation into the underlying logic results in an exponential blowup of the formula size, yielding an incorrectly presumed two-exponential time complexity. We illustrate this from a practical point of view on a few examples such as numerical occurrence constraints and tree frontier properties which are concrete problems found with semi-structured data [21].

6.2. Behavioural Types

Behavioural type systems ensure more than the usual safety guarantees of static analysis. They are based on the idea of "types-as-processes", providing dedicated type algebras for particular properties, ranging from protocol compatibility to race-freedom, lock-freedom, or even responsiveness.

Two successful, although rather different, approaches, are session types and process types. The former allows to specify and verify (distributed) communication protocols using specific type (proof) systems; the latter allows to infer from a system specification a process abstraction on which it is simpler to verify properties, using a generic type (proof) system. What is the relationship between these approaches? Can the generic one subsume the specific one? At what price? And can the former be used as a compiler for the latter?

In [15], we showed how communication protocols can be integrated into an object-oriented type system supporting *non-uniform objects*, i. e. objects where the sequences of method calls are restricted, such as a `File` where `read()` cannot be called after `close()`. In such a system, communication protocols can be enforced by giving appropriate non-uniform types to the socket objects. We defined a sound and complete type checking algorithm for a small distributed class-based object-oriented language with structural subtyping. Static typing guarantees that both sequences of messages on channels, and sequences of method calls on objects, conform to type-theoretic specifications, thus ensuring type-safety.

6.3. SPARQL Queries

Static analysis is a core task in query optimization and knowledge base verification. In [14], [24], we study static analysis techniques for SPARQL, the standard language for querying Semantic Web data. We are interested in developing techniques through reductions to the validity problem in logic

In [22], we investigate techniques for detecting SPARQL query update independence. A query is independent of an update when the execution of the update does not affect the result of the query. Determining independence is especially useful in the context of huge RDF repositories, where it permits to avoid expensive yet useless re-evaluation of queries. While this problem has been intensively studied for fragments of relational calculus, very few works exist for the standard query language for the semantic web. We report on our investigations on how a notion of independence can be defined in the SPARQL context.

6.4. Semantic Subtyping

In a programming language, subtyping represents a notion of safe substitutability (it is always safe to replace a value of some type with a value of a subtype). There are several ways such a relation can be formally defined. Semantic subtyping consists of giving a set-theoretic denotation to types and using set inclusion to define subtyping. Works by Benzaken, Castagna, Frisch and Xu have described how to define such relations

for complex type algebras comprising recursive, product, function, intersection, union, and complement types together with type variables. In [17], we showed how to formalise such a relation in logic and decide it in EXPTIME, answering an open question, and discussed experiments made with the full implementation of the system in our solver (5.3).

6.5. Spatio-temporal validation of multimedia documents

A multimedia document authoring system should provide analysis and validation tools that help authors find and correct mistakes before document deployment. Although very useful, multimedia validation tools are not often provided. Spatial validation of multimedia documents may be performed over the initial position of media items before presentation starts. However, such an approach does not lead to good results when media item placement changes over time. Some document authoring languages allow the definition of spatio-temporal relationships among media items and they can be moved or resized during runtime. Current validation approaches do not verify dynamic spatio-temporal relationships. In [19], we present a novel approach for spatio-temporal validation of multimedia documents. We model the document state, extending the Simple Hypermedia Model (SHM), comprising media item positioning during the whole document presentation. Mappings between document states represent time lapse or user interaction. We also define a set of atomic formulas upon which the author's expectations related to the spatio-temporal layout can be described and analyzed.

6.6. XQuery and Static Typing

XQuery is a functional language dedicated to XML data querying and manipulation. As opposed to other W3C-standardized languages for XML (e.g. XSLT), it has been intended to feature strong static typing. Currently, however, some expressions of the language cannot be statically typed with any precision.

In [20], we argue that this is due to a discrepancy between the semantics of the language and its type algebra. We discuss how to handle this discrepancy by improving the type system. We describe a logic-based language of extended types able to represent inner tree nodes and show how it can dramatically increase the precision of typing for navigation expressions. We describe how inclusion between these extended types and the classical regular tree types can be decided, allowing a hybrid system combining both type languages. The result is a net increase in precision of typing.

In a previous work, we aimed at bridging the gap between path-based XML processing languages like XQuery and pattern-based such languages like CDuce. We extend the language CDuce into a succinct core λ-calculus that captures XQuery 3.0. The extensions we consider essentially allow CDuce to implement XPath-like navigational expressions by pattern matching and precisely type them. The elaboration of XQuery 3.0 into the extended CDuce provides a formal semantics and a sound static type system for XQuery 3.0 programs.

6.7. Efficiently Deciding μ-calculus with Converse over Finite Trees

In [16], we present a sound and complete satisfiability-testing algorithm and its effective implementation for an alternation-free modal μ-calculus with converse, where formulas are cycle-free and are interpreted over finite ordered trees. The time complexity of the satisfiability-testing algorithm is $2^O(n)$ in terms of formula size n. The algorithm is implemented using symbolic techniques (BDD). We present crucial implementation techniques and heuristics that we used to make the algorithm as fast as possible in practice. Our implementation is detailed in 5.3.

6.8. Reasoning with Style

The Cascading Style Sheets (CSS) language constitutes a key component of web applications. It offers a series of sophisticated features to stylize web pages. Its apparent simplicity and power are however counterbalanced by the difficulty of debugging and maintaining style sheets, tasks for which developers still lack appropriate tools. In particular, significant portions of CSS code become either useless or redundant,

and tend to accumulate over time. The situation becomes even worse as more complex features are added to the CSS language (e.g. CSS3 powerful selectors). A direct consequence is a waste of CPU that is required to display web pages, as well as the significant amount of useless traffic at web scale. Style sheets are designed to operate on a set of documents (possibly generated). However, existing techniques consist in syntax validators, optimizers and runtime debuggers that operate in one particular document instance. As such, they do not provide guarantees concerning all web pages in CSS refactoring, such as preservation of the formatting. This is partly because they are essentially syntactic and do not take advantage of CSS semantics to detect redundancies. In [18], we propose a set of automated refactoring techniques aimed at removing redundant and inaccessible declarations and rules, without affecting the layout of any document to which the style sheet is applied. We implemented a prototype that has been extensively tested with popular web sites (such as Google Sites, CNN, Apple, etc.). We show that significant size reduction can be obtained while preserving the code readability and improving maintainability.

6.9. A Comparative Analysis of Attitude Estimation

We investigate the precision of attitude estimation techniques in the context of pedestrian dead-reckoning with commodity smartphones. We propose a comparative analysis of state-of-the-art algorithms for attitude estimation in this setting. We provide an experimental setup with a precise ground truth obtained with a motion capture system. We precisely quantify the error in attitude estimation obtained with each technique. We discuss the obtained results and analyse advantages and limitations of current technology for further PDR research.

7. Partnerships and Cooperations

7.1. National Initiatives

7.1.1. 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.

7.1.2. ANR

Typex

Title: Typeful certified XML: integrating language, logic, and data-oriented best practices

Call: Programme Blanc

Duration: January 2012 - December 2015

Coordinator: PPS (CNRS - Paris 7 Diderot)

Others partners: LRI (Orsay)

See also: http://typex.lri.fr

Abstract: The highly ambitious and final goal of this project is to produce a new generation of XML programming languages stemming from the synergy of integrating three approaches into a unique framework:

- a logical approach based on solvers
- a programming language (PL) approach
- a data-oriented approach

These languages will feature precise and polymorphic type systems that merge PL typing techniques with logical-solver-based type inference. They will be implemented efficiently using the latest research on tree automata and formally certified using modern theorem prover technology. They will offer the capacity to specify and formally verify invariants, business rules, and data integrity, and will have a direct and immediate impact on standardization processes.

7.1.3. Transfer Contracts with Startups

Oppidoc

Title: Study of Potential Benefits of Introducing Static Analyses in the Oppidum Development Process

Duration: November - December 2015

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.

7.2. International Research Visitors

7.2.1. Internships

Martì Bosch Padros from Universitat Politècnica de Catalunya (UPC) Spain spent six months in the team to work on Automated Refactoring for Size Reduction of CSS Style Sheets.

Joel Ferreira Dos Santos from Universidade Federal Fluminense, UFF, Brasil spent a one year sandwich PhD in the team to work on the formal verification of multimedia presentations.

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.

8.1.2. Scientific events selection

8.1.2.1. Chair of conference program committees

P. Genevès is the PC Chair of the ACM Symposium on Document Engineering 2015.

8.1.2.2. Member of the conference program committees

P. Genevès is member of the external review committee for the 42nd ACM Symposium on Principles of Programming Languages (POPL'15).

C. Roisin and N. Layaïda are program committee members for the 15th ACM Symposium on Document Engineering (DocEng'15).

8.1.2.3. Reviewer - Reviewing activities

P. Genevès has been a referee for the ACM Transactions on Database Systems (TODS).

P. Genevès has been a referee for the International Colloquium on Automata, Languages and Programming (ICALP).

P. Genevès has been a referee for the International Conference for Human-Computer Interaction (CHI).

P. Genevès has been a referee for the International Conference on Relational and Algebraic Methods in Computer Science (RAMiCS).

N. Gesbert has been a referee for Logical Methods in Computer Science (LMCS).

8.1.3. Research administration

C. Roisin is a board member of the University Pierre Mendes-France, University of Grenoble, in charge of IT systems and TICE.

C. Roisin is president of the disciplinary commission of the University Pierre Mendes-France, University of Grenoble.

N. Layaïda is a permanent member of the pool of experts (project selection committee) of the Minalogic competitive cluster.

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.

P. Genevès is a permanent member of the committee in charge of hiring research engineers at Inria Grenoble - Rhône-Alpes research center, since 2007.

P. Genevès is co-responsible of the doctoral school of Grenoble University for Computer Science (around 400 PhD students), since July 2015.

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

P. Genevès: 18h (foundations for XML: logics, tree automata and grammars) in the Master of Science in Informatics at Grenoble (Université Joseph Fourier); 3h (tree automata and grammars) in the Master at Ecole Polytechnique Fédérale de Lausanne; 3h (XML Essentials) in the Master at Ecole Nationale Supérieure d'Informatique et de Mathématiques appliquées de Grenoble.

N. Layaïda: 9h of Master lectures on SMIL and Multimedia Principles at Grenoble Universities (Master of Science in Informatics at Grenoble, M2R SIGAL: UE IST, final year).

C. Roisin: 192h at License level on computer network and web technologies; responsible of a L3 degree in web technologies with 50 students per year involved in a dual education system.
C. Roisin is the contact person of the Grenoble Inria Centre for the doctoral schools of University of Grenoble (mainly the MSTII Doctoral School), until June 2015.

N. Gesbert: 223h at License and Master levels in an engineering school (Ensimag), including: 60h on algorithmics and 45h on logic at license level, and 45h on software analysis, conception and validation, 45h on logic and 22h30 on Web application development at master level, plus a few hours of tutoring.

8.2.2. Supervision

PhD defence: N. Guido, On the Static Analysis for SPARQL Queries using Modal Logic, 3 rd October 2015, C. Roisin and P. Genevès

PhD in progress: D. Graux, Large scale evaluation of semantic web queries, since November 2013, 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

P. Genevès is co-responsible of a master-level university course entitled “Semantic Web: from XML to OWL”, given each year in the second year of the Master of Science in Informatics at Grenoble.

N. Layaïda is a permanent member of the jury in charge of evaluation harmonisation of the Master of Science in Informatics at Grenoble.

P. Genevès is co-responsible of the doctoral school of Grenoble University for Computer Science (around 400 PhD students), since July 2015.

N. Layaïda is member of the Scientific Board of Advanced Data-mining of the Parsyval Labex.

N. Layaïda is member of the experts pool (selection committee) of the minalogic competitive cluster.

8.3. Popularization

M. Razafimahazo developed a mobile application for the pupils of the école de la paix, Grenoble. The application uses Augmented Reality for a cultural heritage tour of Grenoble City with content produced by the pupils.

M. Razafimahazo animated a brainstorming during CitizenTIC around OpenStreetMap crowd-sourcing: “Les cartes changent la donne autour des parcours accessibles dans l’agglomération Grenobloise”.

N. Layaïda animated several workshops on Open Linked Data for Grenoble City and Grenoble Agglomeration (La métro).

TyReX received two pupils from secondary school, Jeremy Andreoletti and Hugo Kersaudy, for a one week initiation to research. We also hosted several pupils from primary school for one day for a presentation and demos.

9. Bibliography

Major publications by the team in recent years

[1] M. BOSCH, P. GENEVÈS, N. LAYAÏDA. *Automated Refactoring for Size Reduction of CSS Style Sheets*, in "Proceedings of the 2014 ACM symposium on Document engineering", Fort Collins, Denver, United States, September 2014 [*DOI :* 10.1145/2644866.2644885], https://hal.inria.fr/hal-01081876

[2] G. CASTAGNA, N. GESBERT, L. PADOVANI. *A theory of contracts for Web services*, in "ACM Transactions on Programming Languages and Systems (TOPLAS)", 2009, vol. 31, n^{o} 5, pp. 1-61 [*DOI :* 10.1145/1538917.1538920], http://hal.archives-ouvertes.fr/hal-00434459

[3] G. CASTAGNA, H. IM, K. NGUYỄN, V. BENZAKEN. *A Core Calculus for XQuery 3.0: Combining Navigational and Pattern Matching Approaches*, September 2014, working paper or preprint, https://hal.inria.fr/hal-01104872

[4] F. CAZENAVE, V. QUINT, C. ROISIN. *Timesheets.js: When SMIL Meets HTML5 and CSS3*, in "DocEng 2011: Proceedings of the Eleventh ACM Symposium on Document Engineering", Mountain View, États-Unis, September 2011, 10 p. , http://hal.archives-ouvertes.fr/hal-00619382

[5] M. W. CHEKOL, J. EUZENAT, P. GENEVÈS, N. LAYAÏDA. *SPARQL Query Containment under RDFS Entailment Regime*, in "6th International Joint Conference on Automated Reasoning (IJCAR)", Manchester, United Kingdom, B. GRAMLICH, D. MILLER, U. SATTLER (editors), Springer, June 2012, pp. 134-148, http://hal.inria.fr/hal-00749087

[6] M. W. CHEKOL, J. EUZENAT, P. GENEVÈS, N. LAYAÏDA. *SPARQL Query Containment Under SHI Axioms*, in "26th AAAI Conference on Artificial Intelligence", Toronto, Canada, AAAI Press, July 2012, pp. 10-16, http://hal.inria.fr/hal-00749080

[7] S. J. GAY, N. GESBERT, A. RAVARA. *Session Types as Generic Process Types*, in "Proceedings of the Combined 21st International Workshop on Expressiveness in Concurrency and 11th Workshop on Structural Operational Semantics (EXPRESS/SOS)", Rome, Italy, Electronic Proceedings in Theoretical Computer Science, September 2014, vol. 160, pp. 94 - 110 [*DOI :* 10.4204/EPTCS.160.9], https://hal.inria.fr/hal-01102349

[8] S. J. GAY, N. GESBERT, A. RAVARA, V. T. VASCONCELOS. *Modular Session Types for Objects*, May 2011, http://hal.archives-ouvertes.fr/hal-00700635

[9] S. J. GAY, V. T. VASCONCELOS, A. RAVARA, N. GESBERT, A. Z. CALDEIRA. *Modular Session Types for Distributed Object-oriented Programming*, in "Proceedings of the 37th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages", New York, NY, USA, POPL '10, ACM, 2010, pp. 299–312, http://doi.acm.org/10.1145/1706299.1706335

[10] P. GENEVÈS, N. LAYAÏDA, V. QUINT. *Impact of XML Schema Evolution*, in "ACM Transactions on Internet Technology", July 2011, vol. 11, n^{o} 1 [*DOI :* 10.1145/1993083.1993087], http://hal.inria.fr/inria-00619225

[11] P. GENEVÈS, N. LAYAÏDA, V. QUINT. *On the Analysis of Cascading Style Sheets*, in "WWW '12 - 21st international conference on World Wide Web - 2012", Lyon, France, April 2012, pp. 809-818 [*DOI :* 10.1145/2187836.2187946], http://hal.inria.fr/hal-00690899

[12] P. GENEVÈS, N. LAYAÏDA, A. SCHMITT. *Efficient Static Analysis of XML Paths and Types*, in "Proceedings of the 2007 ACM SIGPLAN conference on Programming language design and implementation", San Diego, États-Unis, 2007, pp. 342–351 [*DOI :* 10.1145/1250734.1250773], http://hal.archives-ouvertes.fr/hal-00189123

[13] N. GESBERT, P. GENEVÈS, N. LAYAÏDA. *Parametric Polymorphism and Semantic Subtyping: the Logical Connection*, in "International conference on functional programming", Tokyo, Japon, ACM SIGPLAN, September 2011, pp. 107-116 [*DOI :* 10.1145/2034773.2034789], http://hal.inria.fr/inria-00585686

Publications of the year

Doctoral Dissertations and Habilitation Theses

[14] N. GUIDO. *On the Static Analysis for SPARQL Queries using Modal Logic*, Université de Grenoble, December 2015, https://hal.inria.fr/tel-01250984

Articles in International Peer-Reviewed Journals

[15] S. J. GAY, N. GESBERT, A. RAVARA, V. T. VASCONCELOS. *Modular session types for objects*, in "Logical Methods in Computer Science (LMCS)", December 2015, vol. 4, n^{o} 12, 76 p. [*DOI :* 10.2168/LMCS-11(4:12)2015], https://hal.archives-ouvertes.fr/hal-00700635

[16] 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

[17] N. GESBERT, P. GENEVÈS, N. LAYAÏDA. *A Logical Approach To Deciding Semantic Subtyping*, in "ACM Trans. On Programming Languages and Systems (TOPLAS)", 2015, vol. 38, n^{o} 1, 31 p. [*DOI :* 10.1145/2812805], https://hal.inria.fr/hal-00848023

International Conferences with Proceedings

[18] 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

[19] J. A. FERREIRA DOS SANTOS, C. BRAGA, D. C. MUCHALUAT-SAADE, C. ROISIN, N. LAYAÏDA. *Spatio-temporal Validation of Multimedia Documents*, in "DocEng'15", Lausanne, Switzerland, Document Engineering 2015, ACM, September 2015 [*DOI :* 10.1145/2682571.2797060], http://hal.univ-grenoble-alpes.fr/hal-01211807

[20] 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

[21] 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

[22] N. GUIDO, P. GENEVÈS, N. LAYAÏDA, C. ROISIN. *On Query-Update Independence for SPARQL*, in "CIKM'15", Melbourne, Australia, ACM (editor), ACM, October 2015, pp. 1675-1678 [*DOI :* 10.1145/2806416.2806586], http://hal.univ-grenoble-alpes.fr/hal-01211811

[23] 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

Conferences without Proceedings

[24] N. GUIDO. *On the Static Analysis for SPARQL Queries Using Modal Logic*, in "IJCAI 2015", Buenos Aires, Argentina, July 2015, https://hal.archives-ouvertes.fr/hal-01215680

Other Publications

[25] L. JACHIET, P. GENEVÈS, N. LAYAÏDA. *An Efficient Translation from a Two-Way μ-Calculus of Finite Trees into Tree Automata*, February 2015, working paper or preprint, https://hal.inria.fr/hal-01117830