



Activity Report 2013

Team LINKS

Linking Dynamic Data

RESEARCH CENTER
Lille - Nord Europe

THEME
Data and Knowledge Representation
and Processing

Table of contents

| | |
|--|----------|
| 1. Members | 1 |
| 2. Overall Objectives | 1 |
| 2.1. Presentation | 1 |
| 2.2. Highlights of the Year | 2 |
| 3. Application Domains | 2 |
| 4. Software and Platforms | 3 |
| 4.1. QuiX-Tool Suite | 3 |
| 4.2. SmartHal | 4 |
| 5. New Results | 4 |
| 5.1. Querying Heterogeneous Linked Data | 4 |
| 5.2. Managing Dynamic Linked Data | 4 |
| 5.3. Linking Data Graphs | 5 |
| 6. Bilateral Contracts and Grants with Industry | 5 |
| 7. Partnerships and Cooperations | 5 |
| 7.1. Regional Initiatives | 5 |
| 7.2. National Initiatives | 6 |
| 7.3. International Initiatives | 6 |
| 7.4. International Research Visitors | 6 |
| 7.4.1. Visits of International Scientists | 6 |
| 7.4.2. Visits to International Teams | 6 |
| 8. Dissemination | 6 |
| 8.1. Scientific Animation | 6 |
| 8.1.1. Organization of Workshops | 6 |
| 8.1.2. Invited Talks | 6 |
| 8.1.3. Program Committees | 7 |
| 8.1.4. French Scientific Responsibilities | 7 |
| 8.2. Teaching - Supervision - Juries | 7 |
| 8.2.1. Teaching | 7 |
| 8.2.2. Supervision | 8 |
| 8.2.3. Juries | 8 |
| 8.3. Popularization | 8 |
| 9. Bibliography | 8 |

Team LINKS

Keywords: Databases, Tree Automata, Logics

Creation of the Team: 2013 January 01.

1. Members

Research Scientists

Joachim Niehren [Team leader, Inria, Senior Researcher, from Jan 2013, HdR]

Pierre Bourhis [CNRS, Researcher from Oct 2013]

Faculty Members

Iovka Boneva [Université Lille 1, Associate Professor, from Jan 2013]

Angela Bonifati [Université Lille 1, Professor, from Jan 2013, HdR]

Aurélien Lemay [Université Lille 3, Associate Professor, from Jan 2013]

Slawomir Staworko [Université Lille 3, Associate Professor, from Jan 2013]

Sophie Tison [Université Lille 1, Professor, from Jan 2013, HdR]

External Collaborators

Anne-Cécile Caron [Université Lille 1, Associate Professor, from Jan 2013]

Yves Roos [Université Lille 1, Associate Professor, from Jan 2013]

Engineers

Guillaume Bagan [Université Lille 1, Engineer from June 2013]

Denis Debarbieux [Inria, granted by Region Nord Pas de Calais, Engineer from Jan 2013]

PhD Students

Adrien Boiret [Université Lille 1, PhD student, from Jan 2013]

Radu Ciucanu [Université Lille 1, PhD student, from Jan 2013]

Grégoire Laurence [Université Lille 1, PhD student, from Jan 2013 until Aug 2013]

Antoine Ndione [Université Lille 1, PhD student, from Jan 2013]

Tom Sebastian [Innovimax, Engineer from Dec 2010]

Post-Doctoral Fellows

Mathias John [Inria, PostDoc, granted by ANR LAMPADA project, from Jan 2013 until Jul 2013]

Vincent Hugot [Inria, PostDoc, from Oct 2013]

Visiting Scientist

Pavel Labath [Comenius University, Slovakia, PhD student, from Jan 2013 until Feb 2013]

Administrative Assistant

Julie Jonas [Inria, Administrative Assistant]

Other

Anthony Lick [ENS Cachan, Master student, from Jun 2013 until Jul 2013]

2. Overall Objectives

2.1. Presentation

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

The following three paragraphs summarize our main research objectives.

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

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

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

2.2. Highlights of the Year

Our paper ‘A trichotomy for regular simple path queries on graphs’ has been accepted for publication in the Proceedings of the 32nd ACM SIGMOD-SIGACT-SIGART Symposium on *Principles of Database Systems*, PODS 2013, the top conference in the field of theoretical databases. The paper addresses an open problem, i.e. giving a complete classification of regular languages with respect to regular simple path queries (RSPQs), the latter being regular path queries (RPQs) with an additional constraint that prevents traversing two nodes multiple times. In particular, we have characterized the boundary between tractability and intractability, and proved a trichotomy: the evaluation of RSPQs is either AC₀, NL-complete or NP-complete in data complexity.

Pierre Bourhis has been recruited as CNRS researcher at LIFL and joined the team in October.

3. Application Domains

3.1. Context

Links are important for web users, who try to locate relevant information. They typically want to pose their queries locally and obtain the answers from both local and remote repositories. With the concept of linked data collections, today’s web users are provided with a virtual collection of data and explicit links. One of the goal of our project is to enrich the collection of data and links with more expressive mappings between local relations and external resources. The latter are not available in the current Web and would lead to better take advantage of the diversity and heterogeneity of information. The answer to a user query needs to exploit both explicit links, such as pointers to external resources or semantic correspondences to those and logical links to external repositories, represented as schema mappings. Therefore, the second goal is to evaluate local queries across such mappings and thus exploit the semantic knowledge of external resources. However, we argue that the benefits of links are not limited to casual users. In this paragraph, we briefly discuss two applications in which linked data collections need to be enriched and queried.

Collective Intelligence. Collective knowledge is a shared or group intelligence that emerges from the collaboration of individuals (from Wikipedia). There are many contexts in which such a concept is readily applicable. We advocate here one possible scenario, namely that of Business Intelligence. In the past decade, most of the enterprise data was proprietary, thus residing within the enterprise repository, along with the knowledge derived from that data. Today’s enterprises and businessmen need to face the problem of information explosion, due to the Internet’s ability to rapidly convey large amounts of information throughout the world via end-user applications and tools. Although linked data collections exist by bridging the gap between enterprise data and external resources, they are not sufficient to support the various tasks of Business Intelligence. To make a concrete example, concepts in an enterprise repository need to be matched with concepts in Wikipedia and this can be done via pointers or equalities. However, more complex logical statements (i.e. mappings) need to be conceived to map a portion of a local database to a portion of an RDF graph, such as a subgraph in Wikipedia

or in a social network, e.g. LinkedIn. Such mappings would then enrich the amount of collective knowledge shared within the enterprise and let more complex queries be evaluated. As an example, businessmen with the aid of business intelligence tools need to make complex sentimental analysis on the potential clients and for such a reason, such tools must be able to pose complex queries, that exploit the previous logical mappings to guide their analysis. Moreover, the external resources may be rapidly evolving thus leading to revisit the current state of collective intelligence.

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

4. Software and Platforms

4.1. QuiX-Tool Suite

Participants: Joachim Niehren [correspondant], Denis Debarbieux, Tom Sebastian.

The QuiX-Tool Suite provides tools to process XML streams and documents. The QuiX-Tool Suite is based on early algorithms: query answers are delivered as soon as possible and in all practical cases at the earliest time point. The QuiX-Tool Suite provides an implementation of the main XML standart over streams. XPath, XSLT, XQuery and XProc are W3C standarts while Schematron is an ISO one. The QuiX-Tool suite is developed in the Inria transfer project QuiXProc in cooperation with Innovimax. It includes among the others existing tools such as FXP and QuiXPath, along with new tools, namely X-Fun. Both, a free and a professional version are available. The ownership of QuiX-Tool Suite is shared between Inria and Innovimax. The main application of QuiX-Tool Suite is its usage in QuiXProc, an professional implementation of the W3C pipeline language XProc owned by Innovimax.

The QuiXPath language is a large fragment of XPath with full support for the XML data model. The QuiXPath library provides a compiler from QuiXPath to FXP, which is a library for querying XML streams with a fragment of temporal logic.

The X-Fun language is a functional language for defining transformations between XML data trees, while providing shredding instructions. X-Fun can be understood as an extension of Frisch's XStream language with output shredding, while pattern matching is replaced by tree navigation with XPath expressions. The QuiX-Tool suite includes QuiXSLT, which is a compiler from XSLT into a fragment of X-Fun, which can be considered as the core of XSLT. It also provides QuiXSchematron, which is a compiler from Schematron to X-Fun, and QuiXQuery, which is a compiler from XQuery to X-Fun.

See also the web page <https://project.inria.fr/quix-tool-suite/>.

- Version: QuiXPath v2.0.0
- Version: X-Fun v0.5.0
- Version: QuiXSLT v0.5.0
- Version: QuiXSchematron v1.0.0

4.2. SmartHal

Participants: Joachim Niehren [correspondant], Antoine Ndione.

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

See also the web page <http://smarthal.lille.inria.fr/>.

- Version: SmartHal v1.0.0

5. New Results

5.1. Querying Heterogeneous Linked Data

Participants: Guillaume Bagan, Iovka Boneva, Angela Bonifati, Pierre Bourhis, Radu Ciucanu, Tom Sebastian, Slawomir Staworko, Sophie Tison.

Staworko, Ciucanu and Boneva presented a new class of schemas for unordered XML trees, which are based on unordered regular expressions, also called multiplicity schemas. They show that many static analysis problems become feasible when removing disjunctions there [6].

Ciucanu and Staworko [8] investigated the case of unordered XML, where the relative order among siblings is ignored, and focused on the problem of learning schemas from examples given by the user. They considered disjunctive multiplicity schemas (DMS) and their restrictions, disjunction-free multiplicity schemas (MS). For both DMS and MS, they prove the learnable cases.

Regular path queries in graphs have found much recent interest in the context of SPARQL queries for linked open data in the RDF format. Bagan, Bonifati and Groz (former PhD student of Mostrare, now PostDoc at Tel-Aviv University) have obtained a precise characterization of those regular path queries that can be answered with polynomial data complexity [5] leading to a trichotomy (AC0, NL-complete, or else NP-complete). Thereby, they have solved an open question (raised by W. Martens in PODS'12).

XPath query evaluation over compressed trees has been studied in [12]. They focused on a fragment of XPath, which is the downward, navigational XPath and presented precise bounds on the time complexity of XPath query execution over grammar-compressed trees. In particular, they focused on counting the nodes selected by an XPath expression, extracting and materializing their pre-order numbers and serializing the obtained subtrees.

In [2], Groz, Staworko, Caron, Roos and Tison studied query rewriting with views when the classes used to define queries and views are Regular XPath and MSO. Next, they investigated problems of static analysis of security access specifications (SAS) by introducing the novel class of interval-bounded SAS and they defined three different manners to compare views (i.e. queries), with a security point of view. Finally, they provided a systematic study of the complexity for deciding these three comparisons.

5.2. Managing Dynamic Linked Data

Participants: Angela Bonifati, Denis Debarbieux, Joachim Niehren, Tom Sebastian.

Bonifati, Goodfellow (former PhD student at the University of Strathclyde, UK), Manolescu and Sileo (former PhD student at the University of Basilicata, Italy, directed by Bonifati) studied XML view maintenance in the presence of updates [1]. Their approach relies on algebraic operators for propagating source updates to the target XML view, e.g. in a typical scenario of GAV (global-as-view) schema mappings. Their algebraic approach is set-oriented as opposed to tuple-oriented methods presented in the literature. Moreover, it leverages structural identifiers and structural join algorithms. As such, it proved to be more efficient than existing methods for updating materialized XML views.

Debarbieux, Gauwin (former PhD student in the team, now Assistant Professor at the University of Bordeaux), Niehren, Sebastian and Zergaoui (CEO at Innovimax) focused on using early nested word automata in order to approximate earliest query answering algorithms for nested word automata in a highly efficient manner [9]. This approximation can be made tight in practice for automata obtained from XPath expressions. An XPath streaming algorithm based on early nested word automata has been implemented in the FXP tool. FXP outperforms most previous tools in efficiency, while covering more queries of the XPathMark benchmark.

5.3. Linking Data Graphs

Participants: Angela Bonifati, Radu Ciucanu, Joachim Niehren, Aurélien Lemay, Grégoire Laurence, Antoine Ndione, Slawomir Staworko.

In [7], Bonifati, Ciucanu and Staworko investigate the problem of inferring arbitrary n-ary join predicates across two relations via user interactions. The relations can be found on the Web, thus they lack integrity constraints. In such a scenario, the user is asked to label as positive or negative a few tuples depending on whether she would like them in the join result or not. Deciding whether the remaining tuples are uninformative, i.e. do not allow to infer the query goal, can be done in polynomial time.

The PhD thesis of Ndione focuses on probabilistic algorithms to decide approximate membership of words in a language by using property testing. In [3], Ndione, Lemay and Niehren presented an algorithm that tests the membership modulo the edit distance. Their algorithm runs in polynomial time, as opposed to other property testing algorithms, leveraging the Hamming distance or the edit distance with moves, that are exponential.

In [11], Laurence, Lemay, Niehren, Staworko and Tommasi (project leader of the Magnet team) studied the problem of learning sequential top-down tree-to-word transducers (STWs). They present a Myhill-Nerode characterization of the corresponding class of sequential tree-to-word transformations (STW). Next, they investigate what learning of STWs means, identify fundamental obstacles, and propose a learning model with abstain. Finally, they present a polynomial learning algorithm.

In [4], Niehren, Champavère (former PhD student in the team), Gilleron and Lemay addressed the problem of learnability of regular queries in unranked trees. The idea is that tree pruning strategies and the schemas (DTD in the specific case) can guide the learning process and lead to a class of queries that are learnable according to those. The obtained learning algorithm adds pruning heuristics to the traditional learning algorithm based on tree automata and exploiting positive and negative examples.

6. Bilateral Contracts and Grants with Industry

6.1. Bilateral Contracts with Industry

Innovimax, Cifre and Engineer (2010-2014) The PhD thesis of Tom SEBASTIAN within the QUIXPROC project is supervised by J. NIEHREN in cooperation with M. ZERGAOUI the head of the INNOVIMAX company. The software development in this context is supported by D. DEBARBIEUX, a senior engineer co-funded by INNOVIMAX and INRIA.

7. Partnerships and Cooperations

7.1. Regional Initiatives

7.1.1. FUI Région AAP 14 Hermes (2013-2015)

Participants: Angela Bonifati [correspondent], Joachim Niehren, Iovka Boneva, Denis Debarbieux.

The Hermes project on “Relation Client Personnalisée et Contextualisée” is coordinated by Bonifati from Links. Our partners are the Université Lille 1, Logos Keyneosoft, Cylande, Norsys, Numisight, Leroy Merlin, Kiabi and Auchan.

The project addresses the problem of enriching the client communication within the marketing process. Starting from heterogeneous data sources (connected devices, social networks and traditional marketing channels), one has to extract the necessary information at hand. The data sources can be seen in a streaming fashion as they produce continuous data.

7.2. National Initiatives

7.2.1. Competitvity Clusters

We participate to the following <http://www.picom.fr/> (Pôle de compétitivité PICOM - regional research cluster on commerce industries). In particular, the Hermes project has been conceived within the cluster.

7.3. International Initiatives

7.3.1. Inria International Partners

7.3.1.1. Declared Inria International Partners

We have submitted a proposal for an Inria North-European Lab Lille-Oxford, which has been accepted. The main people involved are Joachim Niehren (leader), Pierre Bourhis and Angela Bonifati, but the cooperation is equally relevant for Iovka Boneva, Aurélien Lemay, Slawek Staworko, Sophie Tison, Radu Ciucanu (PhD student). The Oxford database group (<http://www.cs.ox.ac.uk/isg/db>) is one of the top database groups world wide. The main persons involved will be Michael Benedikt (leader), Dan Olteanu, Andreas Pieris (postdoc). Further promising cooperation opportunities are to be explored with members of Georg Gottlob’s ERC project DiaDem (<http://www.cs.ox.ac.uk/projects/DIADEM/index.html>) on semantics-based information ex- traction.

7.4. International Research Visitors

7.4.1. Visits of International Scientists

Tova Milo (Tel-Aviv University, Israel) visited the team in Febraury 2013 for one day.

Amr El Abbadi (UCSB, Usa) visited the team in March 2013 for one day.

Jan van den Bussche (Hasselt University, Belgium) visited the team in November 2013 for two days.

7.4.2. Visits to International Teams

Pierre Bourhis visited the Oxford database group (<http://www.cs.ox.ac.uk/isg/db>) for three weeks in October, November and December 2013.

8. Dissemination

8.1. Scientific Animation

8.1.1. Organization of Workshops

A two-days workshop on Data-centric Web Services has been organized by Bourhis and Niehren in October 2003. The workshop gathered together researchers from Inria Lille, Inria Rennes, Inria Saclay and Université Marne La Vallée.

8.1.2. Invited Talks

P. BOURHIS was invited to give a talk at the Journées Complexité et Modèles Finis 2013 - 60 ans d’Etienne Grandjean, on “Containment of monadic datalog for general structures”.

S. TISON, J. NIEHREN, A. LEMAY and A. BOIRET were invited in a Dagstuhl seminar on Tree Transducers and Formal Methods.

8.1.3. Program Committees

J. NIEHREN is member of the editorial board of FUNDAMENTA INFORMATICA. He was in the program committees of STACS (Symposium on Theoretical Aspects of Computer Science) 2014 and TTATT (International Workshop on Trends in Tree Automata and Tree Transducers) 2013.

I. BONEVA was member of the program committee of ADBIS (East-European Conference on Advances in Databases and Information Systems) 2013. She was external reviewer for CIIA 2013 and TOIT.

A. BONIFATI was Chair of WebDB (International Workshop on the Web and Databases) 2013. She was also member of the program committees of SIGMOD (International Conference on Management of Data) 2013, VLDB (International Conference on Very Large Databases) 2013 and EDBT (International Conference on Extending Database Technology) 2013. She was a member of the program committee of BDA (Journées Bases de Données Avancées) 2013. She is a member of the program committee of ICDE (International Conference on Data Engineering) 2014, VLDB 2014, SIGMOD 2014, EDBT 2014, and ICDT (International Conference on Database Theory) 2015.

P. BOURHIS was member of the program committee of BDA (Journées Bases de Données Avancées) 2013 and of the CIKM PhD Workshop (PIKM) 2013. He was external reviewer for PODS 2013, TKDE and Information Systems.

S. STAWORKO was member of the program committee of ICDT 2013 and WebDB 2013 (International Workshop on the Web and Databases).

S. TISON is member of the steering committee of RTA (International Conference on Rewriting Techniques and Applications) and STACS (Annual Symposium on Theoretical Aspects of Computer Science). She is member of the editorial committee of RAIRO-ITA. She is member of the program committee of Highlights of Logic, Games and Automata 2013 and LATA (International Conference on Language and Automata Theory and Applications) 2014.

8.1.4. French Scientific Responsibilities

A. BONIFATI, serves in the vivier of Université Lille 1.

J. NIEHREN serves in the vivier of Université Lille 1.

S. TISON is head of the computer science lab in Lille (LIFL). She is elected member of the “Comité National de la Recherche Scientifique (CoNRS)” (Section 6) and of the IUF senior committee 2013. She was president of the scientific and technical council of the cluster “Pôle de Compétitivité Industries du commerce” and she is member of the administrative board of the cluster Pictanovo. She serves in the vivier of Université Lille 1, where she was member of the selection committee for professor positions. She was also member of the selection committee for professor positions in Marseille.

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

Master (Mocad): Information extraction, 18h, M2, Université Lille 1, France by J. NIEHREN

Master (Mocad): Information extraction, 18h, M2, Université Lille 1, France by A. BONIFATI

License: Introduction to Databases, 54h, L3, Université Lille 1, France by A. BONIFATI

Master: XML technologies: Sax and DOM, 18h, M1, Université Lille 3, France by D. DEBARBIEUX

Licence: Propositional Logic, 18h, L2, Université Lille 3, France by S. STAWORKO

Master: Modelization XML, 24h, M1, Université Lille 3, France by S. STAWORKO

Licence: Artificial Intelligence and Logic, 63h, L3, Université Lille 3, France by S. STAWORKO

Master: Introduction to XML, 40h, M1, Université Lille 3, France by S. STAWORKO

Master : Advanced algorithms and complexity, M1, 57h, Université Lille 1, by S. TISON

Master: XML Technologies, 16h, M2, Université Lille 3, France by A. LEMAY

8.2.2. Supervision

PhD in Progress: R. CIUCANU, Cross-model queries and schema mappings: evaluation and learning. Since Sept. 2012. Supervised by Bonifati.

PhD in Progress: G. LAURENCE, Learning XML transformations for data exchange on the web. Since Sept. 2008. Submitted in January 2014. Supervised by Tommasi, Niehren, Staworko and Lemay.

PhD in Progress: A. M. NDIONE, Probabilistic algorithms for tree automata and transducers. Since Sept. 2010. Submitted in January 2014. Supervised by Niehren and Lemay.

PhD in Progress: T. SEBASTIAN, Streaming algorithms for XSLT. Since May 2011. Supervised by Niehren.

PhD in Progress: A. BOIRET, Top-down tree transformations with look-ahead : foundations and learning. Since Sept. 2011. Supervised by Niehren and Lemay.

PhD in Progress: K. BATMANOV submitted in January 2014. Supervised by Niehren

PhD in Progress: G. MADELAINE. Co-supervised by Niehren (2012-2015).

8.2.3. Juries

A. BONIFATI was in the PhD committee of Federico Cavalieri at Università di Genova, Italy.

J. NIEHREN was the supervisor of the HdR of Cédric Lhoussaine at Université de Lille 1 in December 2013.

J. NIEHREN was in the HdR committee of Pierre Senellart at Paris Telecom in July 2013.

J. NIEHREN was the president of the PhD committee of Evgenia Kopylova from the Bonsai group at the University of Lille 1 in Dec 2013.

S. STAWORKO was in the PhD committee of Alexandre Decan at Université de Mons, Belgium.

S. TISON was in the HdR committee of Sébastien Picault and in the PhD committee of Julie Jacques, at the Université Lille 1. She was in the PhD committees of Vincent Hugot and Elena Tushkanova at Université de Franche-Compté. She was in the HdR committee of Nabil Layaida in Grenoble.

8.3. Popularization

Within the Initiative ‘Chercheurs à l’École’ during the ‘Fête de la science 2013’ in October, Iovka Boneva has done presentations for high-school students in two lycées in Villeneuve d’Ascq and Marcq en Baroeul.

9. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journals

- [1] A. BONIFATI, M. GOODFELLOW, I. MANOLESCU, D. SILEO. *Algebraic incremental maintenance of XML views*, in "ACM Transactions on Database Systems", April 2013, vol. 38, n° 3, pp. 1-45, <http://hal.inria.fr/hal-00816483>
- [2] B. GROZ, S. STAWORKO, A.-C. CARON, Y. ROOS, S. TISON. *Static analysis of xml security views and query rewriting*, in "Information and Computation", December 2013, <http://hal.inria.fr/hal-00796297>

- [3] A. NDIONE, A. LEMAY, J. NIEHREN. *Approximate Membership for Regular Languages modulo the Edit Distance*, in "Theoretical Computer Science", February 2013, <http://hal.inria.fr/hal-00801970>
- [4] J. NIEHREN, J. CHAMPAVÈRE, R. GILLERON, A. LEMAY. *Query Induction with Schema-Guided Pruning Strategies*, in "Journal of Machine Learning Research", April 2013, vol. 14, pp. 927–964, <http://hal.inria.fr/inria-00607121>

International Conferences with Proceedings

- [5] G. BAGAN, A. BONIFATI, B. GROZ. *A Trichotomy for Regular Simple Path Queries on Graphs*, in "ACM PODS", New York, United States, June 2013, pp. 261-272, PODS13 conference, <http://hal.inria.fr/hal-00806448>
- [6] I. BONEVA, R. CIUCANU, S. STAWORKO. *Simple Schemas for Unordered XML*, in "16th International Workshop on the Web and Databases (WebDB)", New York, United States, June 2013, <http://hal.inria.fr/hal-00824459>
- [7] A. BONIFATI, R. CIUCANU, S. STAWORKO. *Interactive Inference of Join Queries*, in "17th International Conference on Extending Database Technology (EDBT)", Athens, Greece, January 2014, <http://hal.inria.fr/hal-00875680>
- [8] R. CIUCANU, S. STAWORKO. *Learning Schemas for Unordered XML*, in "14th International Symposium on Database Programming Languages (DBPL)", Riva del Garda, Trento, Italy, July 2013, <http://hal.inria.fr/hal-00846809>
- [9] D. DEBARBIEUX, O. GAUWIN, J. NIEHREN, T. SEBASTIAN, M. ZERGAOUI. *Early Nested Word Automata for XPath Query Answering on XML Streams*, in "18th International Conference on Implementation and Application of Automata", Halifax, Canada, March 2013, pp. 292-305, <http://hal.inria.fr/hal-00676178>
- [10] M. JOHN, M. NEBUT, J. NIEHREN. *Knockout Prediction for Reaction Networks with Partial Kinetic Information*, in "14th International Conference on Verification, Model Checking, and Abstract Interpretation", Rome, Italy, Lecture Notes in Computer Science, Springer, January 2013, vol. 7737, pp. 355-374, <http://hal.inria.fr/hal-00692499>
- [11] G. LAURENCE, A. LEMAY, J. NIEHREN, S. STAWORKO, M. TOMMASI. *Learning Sequential Tree-to-Word Transducers*, in "8th International Conference on Language and Automata Theory and Applications", Madrid, Spain, Springer, March 2014, <http://hal.inria.fr/hal-00912969>
- [12] T. SEBASTIAN, S. MANETH. *XPath Node Selection over Grammar-Compressed Trees*, in "TTATT 2013", Hanoi, Viet Nam, November 2013 [DOI : 10.4204/EPTCS.134.5], <http://hal.inria.fr/hal-00878748>

Research Reports

- [13] P. LABATH, J. NIEHREN. , *A Functional Language for Hyperstreaming XSLT*, March 2013, <http://hal.inria.fr/hal-00806343>
- [14] A. LICK, J. NIEHREN. , *Early = Earliest?*, October 2013, <http://hal.inria.fr/hal-00873742>
- [15] A. NDIONE, J. NIEHREN, A. LEMAY. , *Sublinear DTD Validity*, March 2013, <http://hal.inria.fr/hal-00803696>

- [16] M. SCHMIDT-SCHAUSS, D. SABEL, J. NIEHREN, J. SCHWINGHAMMER. , *Observational Program Calculi and the Correctness of Translations*, May 2013, <http://hal.inria.fr/hal-00824349>

Other Publications

- [17] R. CIUCANU. , *Learning Queries for Relational, Semi-structured, and Graph Databases*, June 2013, SIGMOD/PODS PhD Symposium, <http://hal.inria.fr/hal-00839904>
- [18] E. FILIOT, F. JACQUEMARD, S. TISON. *Tree Automata with Constraints: a brief survey*, in "Tree Transducers and Formal Methods (Dagstuhl Seminar 13192)", Wadern, Germany, May 2013, pp. 1-18, Tree Transducers and Formal Methods (Dagstuhl Seminar 13192), <http://hal.inria.fr/hal-00840959>