



Activity Report 2011

Team LEO

Distributed and heterogeneous data and
knowledge

RESEARCH CENTER
Saclay - Île-de-France

THEME
Knowledge and Data Representation
and Management

Table of contents

1. Members	1
2. Overall Objectives	2
3. Scientific Foundations	2
3.1. Efficient XML and RDF data management	2
3.2. Models for Web data management	2
3.3. Ontology-based Data and Document Management	3
3.4. Data and Knowledge Integration	3
3.5. Reasoning over Distributed Systems	3
3.6. Cloud-based Data Management	3
4. Application Domains	3
5. Software	4
5.1. AlignViz	4
5.2. AnnoViP	4
5.3. EAP Framework	4
5.4. EdiFlow	5
5.5. KD2R	5
5.6. Glucose2	5
5.7. GlucosER	5
5.8. LiquidXML	5
5.9. LN2R	5
5.10. MESAM	6
5.11. RDFViewS	6
5.12. SomeWhere	6
5.13. SpyWhere	6
5.14. TaxoMap	6
5.15. TaxoMap Framework	6
5.16. ViP2P	7
5.17. XUpOp	7
5.18. XUpIn	7
6. New Results	7
6.1. Efficient XML and RDF data management	7
6.1.1. Materialized views for XML	7
6.1.2. Type-based Update Optimization for XML	7
6.1.3. XML query-update independence	8
6.1.4. Precision and complexity of XQuery type inference	8
6.1.5. Managing temporal XML documents	8
6.1.6. Materialized view selection for RDF	9
6.1.7. Hybrid models for XML and RDF	9
6.1.8. RDF query answering	9
6.2. Models for Web data management	9
6.2.1. A rule-based language for Web data management	9
6.2.2. Web information management with access control	9
6.3. Ontology-based data and document Management	10
6.3.1. Semantic Annotation	10
6.3.2. Adaptive Ontologies for Information Retrieval	10
6.3.3. Querying ontology-based annotations	10
6.3.4. Watermarking for ontologies	10
6.3.5. Consistent query answering in DL-Lite	10
6.3.6. Module-based data management in DL-lite	11

6.4.	Data and Knowledge Integration	11
6.4.1.	Reference Reconciliation	11
6.4.2.	Context-aware Personal Information Management	12
6.4.3.	Mapping between ontologies	12
6.4.4.	Integration of Web resources	13
6.5.	Reasoning over Distributed Systems	13
6.5.1.	Distributed Diagnosis Problems	14
6.5.2.	Distributed Consequence Finding	14
6.5.3.	Towards distributed architectures for Modern SAT Solvers	14
7.	Contracts and Grants with Industry	14
8.	Partnerships and Cooperations	15
8.1.	Regional Initiatives	15
8.1.1.	DW4RDF	15
8.1.2.	EdiFlow	15
8.1.3.	Shiri	15
8.2.	National Initiatives	15
8.2.1.	ANR Codex	15
8.2.2.	ANR ConnectedClouds	15
8.2.3.	ANR DataBridges	15
8.2.4.	ANR Dataring	16
8.2.5.	ANR Geonto	16
8.2.6.	ANR PIMI	16
8.2.7.	ANR UNLOC	16
8.2.8.	Participation to evaluation committees	16
8.3.	European Initiatives	16
8.3.1.	FP7 Projects	16
8.3.1.1.	Hycon2	16
8.3.1.2.	Webdam	17
8.3.2.	Collaborations in European Programs, except FP7	17
8.3.2.1.	KIC EIT ICT Labs DataBridges	17
8.3.2.2.	KIC EIT ICT Labs ConnectedCities	18
8.3.3.	Major European Organizations with which you have followed Collaborations	18
8.4.	International Initiatives	18
9.	Dissemination	19
9.1.	Animation of the scientific community	19
9.1.1.	Participation in Editorial Boards	19
9.1.2.	Participation in Conference Organization	19
9.1.3.	Invited Presentations	21
9.2.	Teaching	21
10.	Bibliography	25

Team LEO

Keywords: Data Management, Artificial Intelligence, Semantics, Web, Cloud Computing, Distributed System

1. Members

Research Scientists

Serge Abiteboul [Senior Researcher, INRIA, until September, HdR]
Ioana Manolescu [Team Leader, Senior Researcher, INRIA, HdR]
Meghyn Bienvenu [Junior Researcher, CNRS]

Faculty Members

Nicole Bidoit [Professor, Univ. Paris 11, HdR]
Philippe Chatalic [Associate Professor, Univ. Paris 11]
Dario Colazzo [Associate Professor, Univ. Paris 11, HdR]
Philippe Dague [Professor, Univ. Paris 11, HdR]
François Goasdoué [Associate Professor, Univ. Paris 11]
Nathalie Pernelle [Associate Professor, Univ. Paris 11]
Chantal Reynaud [Professor, Univ. Paris 11, co-team leader, HdR]
Brigitte Safar [Associate Professor, Univ. Paris 11]
Fatiha Saïs [Associate Professor, Univ. Paris 11]
Laurent Simon [Associate Professor, Univ. Paris 11, HdR]

External Collaborators

Tarek Melliti [Associate Professor, Univ. Evry Val d'Essonne]
Philippe Rigaux [Professor, CNAM, HdR]
Marie-Christine Rousset [Professor, Univ. Grenoble, HdR]
Virginie Thion-Goasdoué [Associate Professor, CNAM]

Technical Staff

Andrés Aranda_Andujar [Since October]
André De_Amorim_Fonseca

PhD Students

Vincent Armant [ATER, Paris 11]
Emilien Antoine [ERC grant, Paris 11, until September]
Mohamed Amine Baazizi [Allocataire MENRT, Paris 11]
Michel Batteux [CIFRE with Sherpa Engineering, Paris 11]
Jesùs Camacho_Rodriguez [Allocataire MENRT, since October]
Julio Cesar Dos Reis [PhD joined between Luxembourg CRP H. Tudor and Paris 11]
Alban Galland [X-Telecom, Paris 11, until September]
Fayçal Hamdi [ANR grant until September, Paris 11]
Konstantinos Karanasos [ANR grant, Paris 11]
Asterios Katsifodimos [Allocataire MENRT, Paris 11]
Rania Khefifi [ANR grant, Paris 11]
Julien Leblay [Allocataire MENRT, Paris 11]
Wael Khemiri [Allocataire MENRT until July and ATER since October, Paris 11]
Noor Malla [Grant of Syrian government, Paris 11]
Yassine Mrabet [Digiteo grant until August and ATER since September, Paris 11]
Huu-Nghia Nguyen [Allocataire MENRT, Paris 11]
Alexandra Roatis [Digiteo grant, Paris 11, since October]
Marina Sahakyan [Grant of French embassy in Armenia, co-supervised between Armenia and Paris 11]

Danai Symeonidou [Allocataire MENRT since October]
Federico Ulliana [Allocataire MENRT, Paris 11]
Lina Ye [Allocataire MENRT until July and ATER since September, Paris 11]
Nadjet Zemirline [Contract until June, Paris 11]

Post-Doctoral Fellows

Nada Abdallah [Until September]
Zoi Kaoudi [ICT Labs grant since September]
George Katsirelos [ANR grant, until November]
Ioannis Katsis [ERC grant, until September]
Gianluca Quercini [INRIA Digiteo grant in October-November, Univ. Paris 11 ANR grant in December]
Fabian Suchanek [ERC grant, until September]

Administrative Assistant

Céline Halter [ITA]

2. Overall Objectives

2.1. Highlights

Our work within the ANR CODEX project has led to several important publications on: the efficient management of RDF data [21], rewriting based on XML materialized views [41] and the maintenance of such views [35], and the efficient processing of updates on XML documents through type projectors [31].

Glucose 2.0 (a SAT solver developed by L. Simon) won the first prize at the international SAT competition 2011, category Application SAT+UNSAT.

AAAI'2011 Outstanding Paper Award for “Complexity of and Algorithms for Borda Manipulation” [38], by Jessica Davies, George Katsirelos, Nina Narodytska, Toby Walsh. George Katsirelos was a PostDoc in Leo until November, 2011.

The YAGO2 demo [40] received the Best Demo Award at the 2011 WWW Conference. Fabian Suchanek was a PostDoc in Leo until September 2011.

3. Scientific Foundations

3.1. Efficient XML and RDF data management

The development of Web technologies has led to a strong increase in the number and complexity of the applications which represent their data in Web formats, among which XML (for structured documents) and RDF (for Semantic Web data) are the most prominent. Leo has carried on research on algorithms and systems for efficiently processing expressive queries on such Web data formats. We have considered the efficient management of XML and RDF data, both for query evaluation and for efficiently applying updates, possibly in concurrence with queries.

3.2. Models for Web data management

Use of the Web to share personal data is increasing rapidly with the emergence of Web 2.0 and social network applications. This creates a particularly challenging setting for data management since the data/knowledge is distributed among a large number of peers and can take many forms (e.g. localization information, ontologies, annotations, access control), and it is often necessary to combine data/knowledge from several peers in order to answer queries. We aim to develop a formal framework for this setting that will enable peers to concurrently reason about global data management activities and cooperate in solving specific tasks.

3.3. Ontology-based Data and Document Management

Ontology-based data management basically aims at enabling usual database management tasks in the presence of ontological constraints (which are typically specified using *description logics*). We are currently studying problems related to query answering in the presence of ontologies formulated in lightweight description logics. We focus mainly on the DL-lite family of description logics, which provides the foundations of the OWL2 profile QL, the W3C recommendation for managing large datasets.

Linking data initiatives aim at publishing more and more RDF data sets on the Web and setting RDF links between data items of different sources. Nevertheless, the Web remains concentrated on the interchange of more or less structured documents. Semantic annotation consists in assigning to a document or to its parts a metadata whose semantics is defined in an ontology. We are investigating ontology-based automatic document annotation approaches and we develop query engines that can be used in this setting.

3.4. Data and Knowledge Integration

A main theme of the group is integration of heterogeneous information. The current explosion of data sources available and the availability of simple technologies for exchanging data has created an important demand for interoperability among resources. It increases the need for techniques which allow their integration and thus also the integration of schemas or ontologies used to describe the data and give it meaning.

Ontologies, which define domain concepts, are essential elements in integration systems. Leo works on ontology-based environments providing access to multiple heterogeneous sources, on reconciliation techniques, and on providing full life-cycle support for ontologies.

3.5. Reasoning over Distributed Systems

In many application domains, the available data and knowledge is naturally spread among independent sources/agents. Some of them can collaborate with others, but none of them has a global view of either the total available knowledge or the topology of the network of interacting agents. We focused on an arising challenge in this context: propose distributed reasoning techniques, able to exploit all the global knowledge, despite the lack of a global view on this knowledge.

Our research has focused on consequence finding in distributed propositional-logic systems, distributed consistency-based diagnosis and distributed diagnosability analysis, consistency checking (SAT), and query answering in distributed description logic systems.

3.6. Cloud-based Data Management

We have recently started to work on the efficient management of complex Web data, in particular structured XML documents and Semantic Web data under the form of RDF, in a cloud-based data management platform. We have investigated architectures and algorithms for storing Web data in elastic cloud-based stores and building an index for such data within the efficient key-value stores provided by off-the-shelf cloud data management platform. We have devised and prototyped such platforms for both XML and RDF data, and started experimenting with them in the Amazon Web Services (AWS) platform [62]. We have recently obtained an academic grant from Amazon, to support the continuation of our work.

4. Application Domains

4.1. Application Domains

The typical IT projects to which our technologies contribute aim at efficient and flexible management of complex digital information. The form and nature of the data often varies: Web pages, Office or PDF documents, XML structured data (sometimes obtained through Web service gateways), thesauri, ontologies

etc. From such heterogeneous, complex resources, interested parties aim at building storage and processing tools, enabling the efficient storage, classification, annotation, enrichment, and fine-grained search on such data.¹ Sample real-life applications that we have already worked on in this setting are:

- Archiving filtered content from online information sources (journals, blogs, ...) with the purpose of recording their perspective on facts involving specific countries, key political actors etc. (EADS data gathering for intelligence purposes, also an application from the WebContent project)
- Building an XML data warehouse out of public e-mails exchanged in a technical standardization body (in our particular case, the W3C) in order to enable a fine-grained social network analysis to determine key players, opinion leaders etc.
- Building a complete processing chain for digital documents from the medical domain. The process may start with the digitization and text extraction from scanned documents (we do not work in this area), then continues with extraction of named entities, document annotation based on existing domain ontologies, mapping of documents to a central domain ontology, reasoning across scattered data sources for query answering, storing, indexing, and distributing the data (and query results) across distributed players.
- Data produced and made public by numerous public administration offices (in France, Europe, and the world) opens many perspectives for integrating, analyzing, and combining data sources into added-value information sources. Time is also an essential dimension here; so is data matching and reconciliation, since the same entity may be referenced from many different viewpoints and reconciliation is needed when joining data sources. Users of such applications could be public administration analyzing the impact of its policies, social scientists and journalists which already work on the data (but gather it with much difficulty) etc. This application is gathered from our collaboration with the DataPublica start-up.

5. Software

5.1. AlignViz

Name: AlignViz

Contact: Fayçal Hamdi (hamdi@lri.fr)

Other contacts: Brigitte Safar (safar@lri.fr) and Chantal Reynaud (chantal.reynaud@lri.fr)

Presentation: a visualization tool for alignments between ontologies

5.2. AnnoViP

Name: AnnoViP

Contact: Konstantinos Karanasos (konstantinos.karanasos@inria.fr)

Other contacts: Ioana Manolescu (ioana.manolescu@inria.fr) and Jesús Camacho_Rodriguez (jesus.camacho-rodriguez@inria.fr)

Presentation: a tool for editing and exploiting XML documents with annotations in a distributed P2P setting

5.3. EAP Framework

Name: EAP Framework

¹Interesting areas of content management, which we do not address, are: audio and video data, natural language processing, data mining, access control and privacy. We collaborate up with other groups specialized in these topics.

Contact: Nadjat Zémirline (nadjat.zemirline@supelec.fr)

Other contacts: Chantal Reynaud (chantal.reynaud@lri.fr)

Presentation: a prototype helping to express adaptation strategies based on the use and a semi-automatic combination of patterns

5.4. EdiFlow

Name: EdiFlow (<http://scidam.gforge.inria.fr>)

Contact: Wael Khemiri (wael.khemiri@inria.fr)

Other contacts: Ioana Manolescu (ioana.manolescu@inria.fr), Jean-Daniel Fekete (jean-daniel.fekete@inria.fr), Pierre-Luc Hémary (pierre-luc.hemery@inria.fr), Véronique Benzaken (veronique.benzaken@lri.fr)

Presentation: A platform for data-intensive visual analytics

5.5. KD2R

Name: KD2R

Contact: Danai Symeonidou (danai.symeonidou@lri.fr)

Other contacts: Nathalie Pernelle (nathalie.pernelle@lri.fr), Fatiha Saïs (fatiha.sais@lri.fr)

Presentation: a tool for OWL2 key discovery on RDF datasets

5.6. Glucose2

Name: Glucose 2.0

Contact: Laurent Simon (simon@lri.fr)

Other contacts: Gilles Audemard (audemard@cril.univ-artois.fr)

Presentation: The new version of Glucose (released in 2009), with auto-adaptive clause database management.

5.7. GlucosER

Name: GlucosER

Contact: Laurent Simon (simon@lri.fr)

Other contacts: George Katsirelos (gkatsi@gmail.com) and Gilles Audemard (audemard@cril.univ-artois.fr)

Presentation: a SAT Solver based on Glucose 1.0 with Extended Resolution.

5.8. LiquidXML

Name: Liquid XML (<http://vip2p.saclay.inria.fr/?page=liquidxml>)

Contact: Asterios Katsifodimos (asterios.katsifodimos@inria.fr)

Other contacts: Ioana Manolescu (ioana.manolescu@inria.fr)

Presentation: a prototype for automatically recommending XML materialized views in order to improve the performance of a query workload

5.9. LN2R

Name: LN2R

Contact: Fatiha Saïs (sais@lri.fr)

Other contacts: Nathalie Pernelle (pernelle@lri.fr)

Presentation: a logical and numerical tool for reference reconciliation.

5.10. MESAM

Name: MESAM

Contact: Nadjat Zémirline (nadjat.zemirline@supelec.fr)

Other contacts: Chantal Reynaud (chantal.reynaud@lri.fr)

Presentation: a plug-in for Protege 2000 to merge generic and specific models

5.11. RDFViewS

Name: RDFViewS (<http://tripleo.saclay.inria.fr/rdfvs/>)

Contact: Konstantinos Karanasos (konstantinos.karanasos@inria.fr)

Other contacts: François Goasdoué (fg@lri.fr), Julien Leblay (julien.leblay@inria.fr), and Ioana Manolescu (ioana.manolescu@inria.fr)

Presentation: a storage tuning wizard for RDF applications

5.12. SomeWhere

Name: SomeWhere

Contact: François Goasdoué (fg@lri.fr)

Other contacts: Philippe Chatalic (chatalic@lri.fr) and Laurent Simon (simon@lri.fr)

Presentation: a peer-to-peer infrastructure for propositional reasoning

5.13. SpyWhere

Name: SpyWhere

Contact: François-Elie Calvier (fcalvier@gmail.com)

Other contacts: Chantal Reynaud (chantal.reynaud@lri.fr)

Presentation: a generator of mapping candidates for enriching peer ontologies

5.14. TaxoMap

Name: TaxoMap (<http://taxomap.lri.fr>)

Contact: Fayçal Hamdi (hamdi@lri.fr)

Other contacts: Brigitte Safar (safar@lri.fr) and Chantal Reynaud (chantal.reynaud@lri.fr)

Presentation: a prototype to automate semantic mappings between taxonomies

5.15. TaxoMap Framework

Name: TaxoMap Framework

Contact: Fayçal Hamdi (hamdi@lri.fr)

Other contacts: Brigitte Safar (safar@lri.fr) and Chantal Reynaud (chantal.reynaud@lri.fr)

Presentation: an environment to specify treatments to refine mappings and to enrich ontologies

5.16. ViP2P

Name: ViP2P (views in peer-to-peer, <http://vip2p.saclay.inria.fr>)

Contact: Ioana Manolescu (ioana.manolescu@inria.fr)

Other contacts: Jesús Camacho_Rodriguez (jesus.camacho-rodriguez@inria.fr), Asterios Katsifodimos (asterios.katsifodimos@inria.fr), Konstantinos Karanasos (konstantinos.karanasos@inria.fr)

Presentation: a P2P platform for disseminating and querying XML and RDF data in large-scale distributed networks.

5.17. XUpOp

Name: XUpOp (XML Update Optimization)

Contact: Dario Colazzo (colazzo@lri.fr)

Other contacts: Nicole Bidoit (bidoit@lri.fr), Marina Sahakian (Marina.Sahakyan@lri.fr), and Mohamed Amine Baazizi (baazizi@lri.fr)

Presentation: a general purpose type based optimizer for XML updates

5.18. XUpIn

Name: XUpIn (XML Update Independence)

Contact: Federico Ulliana (Federico.Ulliana@lri.fr)

Other contacts: Dario Colazzo (colazzo@lri.fr), Nicole Bidoit (bidoit@lri.fr)

Presentation: an XML query-update independence tester

6. New Results

6.1. Efficient XML and RDF data management

Participants: Mohamed Amine Baazizi, Nicole Bidoit, Dario Colazzo, François Goasdoué, Konstantinos Karanasos, Asterios Katsifodimos, Julien Leblay, Noor Malla, Ioana Manolescu, Alexandra Roatis, Marina Sahakyan, Federico Ulliana.

6.1.1. Materialized views for XML

We have continued our work on optimizing XML queries through materialized view-based rewriting, implemented within the ViP2P system. We published in IEEE ICDE 2011 an algorithm for rewriting XQuery queries using materialized XQuery view, which improves the state of the art in terms of expressive power of the supported XQuery subset, in collaboration with V. Vassalos (AUEB, Greece) [41]. Two follow-up works concern: efficient algebraic algorithms for incrementally maintaining the materialized views when the underlying documents change, in collaboration with A. Bonifati (CNR, Italy) [35], and algorithms for automatically recommending views to materialize for a given XML query workload, with V. Vassalos.

6.1.2. Type-based Update Optimization for XML

XML projection is a well-known optimization technique for reducing memory consumption for XQuery in-memory engines in order to overcome the main-memory limitations of these systems (Galax, Saxon, QizX, and eXist). One of our main research line focuses on a schema-based projection technique for update optimization. The update language considered is XQuery Update Facility (XUF). The main idea behind this technique is: given a query q over an XML document t , instead of evaluating q over t , the query q is evaluated on a smaller document t' obtained from t by pruning out, at loading time, parts of t that are irrelevant for q . The queried document t' , a projection of the original one, is often much smaller than t due to selectivity of queries.

The scenario and type-based projection proposed for XML queries, cannot be applied directly for updates. We have proposed a new scenario which is composed of four steps:

1. from the update U and the DTD D , a type projector P is inferred;
2. the document t , valid wrt D , is projected following P in a streaming manner, at loading time;
3. U is evaluated over the projection $P(t)$ and produces a partial result $U(P(t))$;
4. the initial document t is merged with $U(P(t))$, in a streaming manner, at writing (serializing) time in order to produce the final result $U(t)$.

The scenario has been first studied and implemented for a kind of type projector which is a good compromise between simplicity and effectiveness, and corresponding results have been published in [31]. Subsequently, We have improved this technique by designing: (i) a new kind of type projector that minimizes the amount of data kept in the projection, and (ii) a new merge algorithm using the improved type projector. This analysis is complicated by the strong interconnection between the two tasks: while minimizing the projection we need to ensure a fast and correct merge process [17].

These results have also been presented in [34], providing an overview on the use of types and constraints from relational to XML data, and in the tutorial [63] focusing on schema-based techniques for safe and efficient XML processing.

6.1.3. XML query-update independence

A query and an update are independent when the query result is not affected by update execution, on any possible input database. Detecting query-update independence is of crucial importance in many contexts: view maintenance, concurrency, access control policies etc. Benefits are amplified when query-update independence can be checked statically. We propose a novel schema-based approach for detecting XML query-update independence. Differently from traditional schema-based analysis for XQuery, our system infers sequence of labels, called *chains*, that are vertically navigated in each schema instance by query and update paths. More precisely, for each node that can be selected by a query/update path in a schema instance, the system infers a chain recording: a) all labels that are encountered from the root to the selected node, and b) the order in which these labels are traversed. The contextual and ordering information provided these chains is at the basis of and extremely precise static independence analysis. We have devised a quite precise chain-inference system, and devised techniques for efficient implementation of the chain-based independence analysis. Results and experiments concerning this line of research have been recently submitted to an international conference.

6.1.4. Precision and complexity of XQuery type inference

A key feature of XQuery is its type system. Any language expression is statically typed and its type is used during program type-checking. In XQuery, types of input data and functions are defined in terms of regular expression types, but it is quite easy to write queries that generate non-regular languages. As a consequence, any type system for XQuery has to rely on a *type inference* process that approximates the (possibly non-regular) output type of a query with a regular type. This approximation process, while mandatory and unavoidable, may significantly decrease the precision of the inferred types. In [37] we study and compare in terms of precision and computational complexity two main existing XQuery type systems.

6.1.5. Managing temporal XML documents

The management of temporal data is a crucial issue in many database applications. We are currently investigating efficient storage and update methods for temporal XML documents, with a focus on compactness of the representation. One of the method developed relies on the type based optimization method developed for updates [31]. First results about this line of research are included in recent publications [29], [30]. Next research activities focus on the design of expressive temporal query and update languages, and on the use of techniques proposed in [29] for query and update optimization.

6.1.6. *Materialized view selection for RDF*

Syntactically, RDF, the data format of the Semantic Web, resembles relational data. However, RDF query processing is significantly complicated by the irregular nature of RDF data and by its simplistic data model, which leads to syntactically complex queries (involving many joins over the whole triple data set). When a query workload is known, the performance of the workload can be significantly improved by materializing access support data structures such as materialized views. Our efficient algorithms for selecting a set of views to materialize in order to speed up the processing of a set of RDF queries are described in a recent publication [21]. The prototype implementing them has been demonstrated at [53].

6.1.7. *Hybrid models for XML and RDF*

We have obtained interesting results in the area of jointly managing XML and RDF data. A first direction of work in this area was to support *annotated documents*, that is, XML documents where individual nodes or fragments could be annotated with RDF triples. This model allows to capture, for instance, blog comments, user ratings on social sites etc. We have proposed a general model based on W3C standards for modeling such data [39].

6.1.8. *RDF query answering*

The current trend for efficiently querying RDF datasets consists of delegating query evaluation to a scalable RDBMS. However, RDF query answering requires in addition to handle – outside the RDBMS – the RDF semantics. We have introduced the *database (DB) fragment of RDF*, encompassing the popular Description Logic (DL) one with essential RDF features like modeling incomplete information, for which we have devised novel *saturation- and reformulation-based techniques* for answering the *Basic Graph Pattern* (BGP) queries of SPARQL. This extends the state of the art on pushing RDF query processing within RDBMSs.

6.2. Models for Web data management

Participants: Serge Abiteboul, Emilien Antoine, Meghyn Bienvenu, Alban Galland.

A book on Web Data Management and Distribution [54] was published this year.

6.2.1. *A rule-based language for Web data management*

We recently proposed [26] a Datalog-style rule-based language (called *Webdamlog*) for web data management. A novel feature of our language is *delegation*, that is, the possibility of installing a rule at another peer. In its simplest form, delegation is essentially a remote materialized view. In its general form, it allows peers to exchange rules, i.e., knowledge beyond simple facts, and thereby provides the means for a peer to delegate work to other peers.

A key contribution of our work is a study of the impact on expressiveness of delegations and explicit timestamps. We showed that both strictly augment the power of the language. In order to validate the semantics of our model, we demonstrated that under certain natural conditions, our semantics converges to the same semantics as the centralized system with the same rules.

6.2.2. *Web information management with access control*

We investigated the problem of sharing private information on the Web, where the information is hosted on different machines that may use different access control and distribution schemes. Based upon our work on Webdamlog, we introduced a distributed knowledge-base model, termed WebdamExchange, that comprises logical statements for specifying data, access control, distribution and knowledge about other peers. In a demo at ICDE [28], we showed how the model can be used in a social-network context to help users keep control on their data on the web. In particular, we demonstrated how users within very different schemes of data distribution (centralized, DHT, unstructured P2P, etc.) can still transparently collaborate while keeping a good control over their own data.

6.3. Ontology-based data and document Management

Participants: Meghyn Bienvenu, François Goasdoué, Yassine Mrabet, Nathalie Pernelle, Gianluca Quercini, Chantal Reynaud, Brigitte Safar, Fabian Suchanek.

6.3.1. Semantic Annotation

We have started a work on semantic annotation of public administration data in the setting of the project DataBridges, an ICT Labs activity. We considered public data represented in tables. The tables that we studied were tables created and published by INSEE. They are spreadsheets filled with statistics about geographic locations and are usually composed of multiple columns, of which one, that we term the subject column, contains a list of textual references to geographic entities, or toponyms, while the others contain numeric attributes. We proposed an approach and an algorithm that assigns a type, or header to the subject column of a INSEE table and identifies the geographic entities referred to by the toponyms in the column [64]. An external resource, DBpedia, is used to help to disambiguate the entities mentioned in the tables and a domain ontology ensures that the types are relative to the geographic domain. This work is continued in the setting of a post-doctoral work granted by the ANR project DataBridges. The aim of the project being to enrich a data warehouse, a first work is to automatically build an initial RDF data warehouse from data collected from the web.

6.3.2. Adaptive Ontologies for Information Retrieval

We published the approach supported by the TARGET framework for Web Information Retrieval in the International Journal of Web Portals (IJWP) [22]. This approach was the core of the PhD of Cédric Pruski defended in April 2009.

6.3.3. Querying ontology-based annotations

We have pursued our work on integrating knowledge bases and semantic annotations made on more or less structured tagged documents. We have defined an approach where RDF named graphs are used to distinguish uncertain semantic annotations from rdf triples that are provided by the populated ontology. A user domain query is then reformulated to obtain answers that are ranked according to their provenance (knowledge bases or annotations) [61].

6.3.4. Watermarking for ontologies

Ontologies are usually available under some type of license. The large ontologies of the Semantic Web, e.g., are commonly licensed under a Creative Commons License or a GNU license. These licenses require giving credit to the authors of the ontology if the ontology is ever used somewhere else. However, it can be hard to prove whether an ontology is used somewhere else, because ontologies contain world knowledge. If someone “steals” an ontology and uses it somewhere else, he can always claim that he collected the data by himself from real-world sources. To tackle this problem, we have studied approaches that watermark an ontology [43]. If a watermarked ontology is used somewhere else, the mark proves that the ontology has been stolen. Existing approaches have mainly modified the facts in the ontology to create a mark. This, however, compromises the precision of the ontology. Therefore, we have developed an approach that does not modify, but remove certain facts. Thereby, the precision of the ontology is not affected. We show that only a handful of facts have to be removed from an ontology to protect it against theft.

6.3.5. Consistent query answering in DL-Lite

An important problem which arises in ontology-based data access is how to handle inconsistencies. In the database community, the related problem of querying databases which violate integrity constraints has been extensively studied under the name of consistent query answering. The standard approach is based on the notion of a repair, which is a database which satisfies the integrity constraints and is as similar as possible to the original database. Consistent answers are defined as those answers which hold in all repairs. A similar strategy can be used for description logics by replacing the integrity constraints with the ontology. Unfortunately, recent work on consistent query answering in description logics has shown this problem to be co-NP-hard in

data complexity, even for instance queries and the simplest DL-Lite dialect. In light of this negative result, we considered the problem of identifying cases where consistent query answering is feasible, and in particular, can be done using query rewriting, with the aim of better understanding the cases in which query rewriting can be profitably used. In [51], we make some first steps towards this goal by formulating general conditions which can be used to prove that a consistent rewriting does or does not exist for a given DL-Lite TBox and instance query.

6.3.6. Module-based data management in DL-lite

The current trend for building an ontology-based data management system (DMS) is to capitalize on efforts made to design a preexisting well-established DMS (a *reference* system). The method amounts to extract from the reference DMS a piece of schema relevant to the new application needs – a module –, possibly to personalize it with extra-constraints w.r.t. the application under construction, and then to manage a dataset using the resulting schema. We have revisited the reuse of a reference ontology-based DMS in order to build a new DMS with specific needs. We go one step further by not only considering the *design* of a module-based DMS (i.e., how to extract a module from an ontological schema): we also study how a module-based DMS can benefit from the reference DMS to enhance its own data management skills. We consider the setting of the DL-Lite_A dialect of DL-Lite, which encompasses the foundations of the QL profile of OWL2 (i.e., DL-Lite_R): the W3C recommendation for managing efficiently large datasets. We introduce and study novel properties of *robustness* for modules that provide means for checking easily that a robust module-based DMS evolves *safely* w.r.t. both the schema and the data of the reference DMS. From a module *robust to consistency checking*, for any data update in a corresponding module-based DMS, we show how to query the reference DMS for checking whether the local update does not bring any inconsistency with the data and the constraints of the reference DMS. From a module *robust to query answering*, for any query asked to a module-based DMS, we show how to query the reference DMS for obtaining additional answers by also exploiting the data stored in the reference DMS.

6.4. Data and Knowledge Integration

Participants: Julio Cesar Dos Reis, Fayçal Hamdi, Rania Kheffi, Yassine Mrabet, Nathalie Pernelle, Chantal Reynaud, Fatiha Saïs, Brigitte Safar, Fabian Suchanek, Danai Symeonidou.

6.4.1. Reference Reconciliation

The reference reconciliation problem consists in deciding whether different data descriptions refer to the same real world entity (same person, same conference etc.) Some of existing approaches, such as LN2R, are declarative and knowledge-based. Different kinds of knowledge can be declared in a domain ontology, like disjointness between classes or key constraints. This knowledge can be exploited to infer reconciliation and non-reconciliation decisions.

Our reference reconciliation work pursues three directions:

- develop an automatic approach of key constraint discovery. We have proposed in [46] KD2R, a method which allows automatic discovery of key constraints associated to OWL2 classes. These keys are discovered from RDF data which can be incomplete. The proposed algorithm allows this discovery without having to scan all the data. KD2R has been tested on data sets of the international contest OAEI and obtains promising results.
- develop a reference reconciliation method for detecting redundant data in case of web data tables that are semantically annotated by an ontology. Each table cell values consists in numerical fuzzy set (NFS) or in symbolic fuzzy set (SFS). We have developed a method which uses ontology knowledge and computes similarity scores to decide the data redundancy. We have also proposed two similarity measures for numerical fuzzy set as well as symbolic fuzzy set. The proposed measures are more flexible than existing ones. This approach has been published in [36], [58]. We are working on its extension to be able to distinguish redundant data from similar ones by using provenance information.

- develop a new approach which addresses the problem of resource discovery in the Linked Open Data cloud (LOD) where data described by different schemas is not always linked. We have proposed an automatic approach in [42], [58] that allows discovery of new links between data. These links can help to match schemas that are conceptually relevant with respect to a given application domain. Furthermore, these links can be exploited during the querying process in order to combine data coming from different sources. In this approach we exploit the semantic knowledge declared in different schemas in order to model: (i) the influences between concept similarities, (ii) the influences between data similarities, and (iii) the influences between data and concept similarities. The similarity scores are computed by an iterative resolution of two non linear equation systems that express the concept similarity computation and the data similarity computation.

6.4.2. Context-aware Personal Information Management

Personal information management (PIM) is the practice and analysis of the activities performed by people to acquire, organize, maintain, and retrieve information for everyday use. PIM is a growing area of interest because, everyone is looking for better use of our limited personal resources of time, money and energy. Several research on the topic is being done in different disciplines, including human-computer interaction, database management, information retrieval and artificial intelligence.

The increasingly big amount of personal information (e.g., mails, contacts, appointments) managed by a user is characterized by their heterogeneity, their dispersion and their redundancy. The general goal of this work consists in designing a system, which allows providing the end-users personal data access with services that are relevant to his/her needs, and to access personal data both by mobile devices (smartphone) and Internet-connected Personal Computers. More specifically, we focus here on the problem of defining a common meta-model for a flexible and homogeneous personal information management. The meta-model that we propose allows users creating personal information and organizing them according to different points of view (ontologies) and different contexts. Contextual queries are defined to allow users to retrieve its personal information using the geographical contexts. The semantic Web languages (OWL, RDF and SPARQL) are used to implement the approach.

6.4.3. Mapping between ontologies

We pursue our work on ontology alignment in the setting of the ANR GeOnto project by aiming to provide full life-cycle support for ontologies.

We investigated how alignment results generated by our alignment tool, TaxoMap, can be used to enrich one ontology with another. We shown that the enrichment process depends on characteristics of the ontology used for enrichment. Three enrichment contexts identified in the setting of the ANR project GeOnto have been studied and enrichment treatments performed. A first context considers ontologies of the same application domain and of a reasonable size. A second context considers small ontologies previously extracted from a generalist one. A third context considers enrichment from a huge, generalist ontology, such as Yago. Early results obtained in the setting of the ANR project GeOnto in the topographic domain have been published in [50], [25].

The module supporting our enrichment approach has been implemented in TaxoMap Framework using patterns. Initially, TaxoMap Framework was composed of our alignment tool, TaxoMap, we are working on for several years in the team and of a mapping refinement module. We extended it in order to obtain a broader framework and an interactive environment by including TaxoPart, a partitioning tool we developed to split two huge ontologies which could not be aligned into two sets of blocks of a limited size, and a module specific to ontology enrichment. Moreover, we re-implemented TaxoMap, our alignment tool, as a web service to make it easily accessible at: <http://taxomap.lri.fr:8000/axis2/services/TaxoMapService?wsdl> .

We also started a PhD work, joined with CRP Henri Tudor in Luxembourg, to investigate issues dealing with medical knowledge organizing systems evolution. We will define a formal framework to support medical knowledge organizing systems evolution in a consistent way and also to support the maintenance of mappings directly impacted by knowledge organizing systems local evolution.

On a related topic, we have developed a probabilistic framework, PARIS (Probabilistic Alignment of Relations, Instances and Schema), for matching ontologies holistically, thereby exploiting synergies between matches on the instance level and matches on the schema level [57]. The framework is parameter-free and does not require resource-specific tuning. PARIS is fully implemented and has been shown to match some of the largest ontologies on the Semantic Web with a precision of around 90%.

6.4.4. Integration of Web resources

We have pursued our work on integration of resources available on the Web in Adaptive Hypermedia Systems (AHS), allowing creators to define their own adaptation strategies based on their own domain models.

The approach is based on a set of 22 adaptation patterns, independent of any application domain and independent of any adaptation engine, published in [59], [47]. These elementary adaptation patterns are organized in a typology in order to facilitate their understanding and their use in the EAP framework to define complex strategies. In [24], we described the whole process to generate complex adaptation strategies and how the generated strategies can be integrated into existing AHSs. The results of an experiment conducted in the e-learning domain is presented. It showed that the pattern-based approach for defining adaptation strategies is more suitable than those based on "traditional" AH languages.

We also pursued our work on the integration of the EAP framework and other AHSs. Our collaboration with A. Cristea from the University of Warwick (UK) led us to a very detailed study of adaptation languages. The first flexible generic adaptation language is the LAG adaptation language. We studied the expressivity of this initial adaptation language in comparison with our newly proposed language, in the EAP framework, and the pros and cons of various decisions in terms of the ideal way of defining an adaptation language. We proposed a unified vision of adaptation and adaptation language. The unified vision is not limited to the two languages analyzed, and can be used to compare and extend other approaches in the future. Beside this theoretical qualitative study, we also made experimental evaluation and comparison of the two languages, and an article is currently being evaluated.

We have also investigated integration of Web services. The Search Computing project ("SeCo") at the Polytechnic University of Milan aims to orchestrate Web services to answer user queries. Currently, the project represents Web services by so-called Service Marts. These are frame-like representations of the services, which follow the slot-value paradigm. This representation faces several challenges if more Web services get added to the system, because it is hard to ensure that Web services added by different users can still be joined. Therefore, we have explored a more ontological representation of Web services. In our proposal [55], Web services are represented as sub-graphs of an ontology. This allows users to add new Web services that re-use the vocabulary of existing Web services.

On related topics, together with researchers from the Max-Planck Institute in Saarbrücken, we have worked on extending the YAGO ontology. YAGO already contains dozens of millions of facts. With the present work, we aim to give these facts a temporal and a spatial dimension. For every event and every entity, we want to know where and when these objects existed. For this purpose, we have developed a methodology that extracts these types of facts from Wikipedia. We have also developed a logical reasoning framework that allows propagating these time and space annotations from some facts to others. This has grown YAGO to 80 million facts in total, making it an ontology that is anchored in time and space (Best demo award at the WWW 2011 conference [40]).

6.5. Reasoning over Distributed Systems

Participants: Vincent Armant, Philippe Chatalic, Philippe Dague, François Goasdoué, George Katsirelos, Laurent Simon, Lina Ye.

6.5.1. *Distributed Diagnosis Problems*

We pursued the work on distributed algorithms for diagnosing distributed systems. The general framework is consistency-based diagnosis for propositional-logic theories in a P2P setting with privacy constraints. It boils down to distributed implicant finding and is thus in some sense dual to the problem of consequence finding described in next paragraph. Vincent Armant is finishing his PhD and has extended his previous work on more general topics, i.e. focusing on the construction of a good decomposition of the network that will ensure an efficient reasoning mechanism. An important effort has been put in the design of a real-world sized experimentation on distributed systems.

Lina Ye defended her PhD on diagnosability analysis of distributed discrete-event systems, modeled as synchronized labeled automata. The aim of diagnosability is to ensure that a given partially observable system has the property that any fault (taken from a set of faults given a priori) will be detectable and identifiable without ambiguity in a finite time after its occurrence. Distributed diagnosability analysis is optimized by abstracting necessary and sufficient information from local objects to achieve global decision. After having addressed the distribution of the system's model into local models, we focus in 2011 on the extension to systems where the observable information itself is distributed instead of centralized. Joint diagnosability definition has been provided and undecidability of deciding it has been proved in the general case where communication events are not observable, before proposing an algorithm to test its sufficient condition. In addition, decidability result and algorithm have been given when communications are observable.

Michel Batteux defended also his PhD (led in the framework of a CIFRE thesis with Sherpa Engineering) about diagnosability and diagnosis of technological systems. The work was led in the centralized case, focusing on defining, implementing, testing and validating on a real case study (a fuel cell system) an all-in-one tool to design a diagnosis system for technological systems by integrating representation of the system and its potential faults, off-line diagnosability analysis and automatic generation of the on-line embedded diagnoser .

6.5.2. *Distributed Consequence Finding*

A major reengineering of the SOMEWHERE platform, for decentralized consequence finding, has been initiated within the DISQUE project. Current efforts have focused on the rewriting of the communication layer, that now relies on the JXTA middleware. A new tool is also being developed, in order to facilitate large scale experimentations on a grid (Grid5000). This tool is designed in a fairly generic way, in order to be reusable for similar projects that require deploying sets of collaborating reasoners in a decentralized setting, and automating collaborative problem solving on various instances. We also expect this tool to be used for automating integration tests during further developments.

6.5.3. *Towards distributed architectures for Modern SAT Solvers*

If we aim at proposing a new architecture for distributed SAT Solvers, we pursued this year the improvements of Glucose, our centralized SAT solver. Glucose 2 won 3 medals at the SAT 2011 Competition, and one in the category Application SAT+UNSAT. We target to make a massively distributed version of Glucose, for very hard SAT problems.

7. Contracts and Grants with Industry

7.1. Contracts with Industry

7.1.1. *DataPublica*

Participants: Ioana Manolescu, Nathalie Pernelle, Chantal Reynaud, Fatiha Saïs, Brigitte Safar.

A collaboration has been initiated with the DataPublica start-up (F. Bancilhon, C. Frisch) and the Zenith INRIA team (ex-ATLAS) from Montpellier. DataPublica aims at drawing up a catalog of the public data sources of the French domain, and in particular those produced by public administration, mostly in Excel files. The contract with DataPublica aims at designing a semantic annotation tool for typing entities in columns of tables in the geographic domain.

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. DW4RDF

Participants: Dario Colazzo, François Goasdoué, Ioana Manolescu, Alexandra Roatis.

This Digiteo DIM LSC (*Logiciels et Systèmes Complexes*) project has started in October 2011. The aim is to design and implement data warehouse-style models and technologies for Semantic Web data (or more generally for RDF). F. Goasdoué coordinates the project, which supports the PhD scholarship of A. Roatis.

8.1.2. EdiFlow

Participants: Wael Khemiri, Ioana Manolescu.

This Digiteo DIM LSC (*Logiciels et Systèmes Complexes*) project ends this year. We have attained our goal of building an efficient and expressive data visualization tool on top of a relational database management system (RDBMS) and published our results [33]. Follow-up work is clearly possible, in particular on more expressive workflow systems and in-memory data management coupled to visual analytics platforms.

8.1.3. Shiri

Participants: Yassine Mrabet, Nathalie Pernelle, Chantal Reynaud.

This Digiteo research project, led by N. Pernelle, ends this year. It involves two partners of Digiteo, Supelec and the University of Paris-Sud. The aim is to annotate and query documents that contain both semi-structured and textual data. This project supports the PhD scholarship of Y. Mrabet.

8.2. National Initiatives

In France, close links exist with groups at Orsay (proofs and programs, V. Benzaken; bio-informatics, C. Froidevaux; machine learning, M. Sebag; information visualization, J.-D. Fekete), with the Cedric Group at CNAM-Paris; some INRIA groups (Dahu, L. Segoufin, at INRIA-Saclay, Zenith, P. Valduriez at Montpellier; Exmo, J. Euzenat, at INRIA Rhone-Alpes; Mostrare at INRIA-Nord-Europe); some INSERM group (Ingénierie des connaissances e-santé, J. Charlet), the BIA group at INRA (P. Buche, C. Dervin), the IRIT of the University of Toulouse (N. Aussenac), the LIRMM of the University of Montpellier (M. Chein, M-L. Mugnier), INRA/CIRAD Montpellier (P. Buche, S. Destercke and R. Thomopoulos) and INRA-AgroParisTech (J. Dibia-Barthélémy)

8.2.1. ANR Codex

Participants: Dario Colazzo, François Goasdoué, Konstantinos Karanasos, Ioana Manolescu.

The Codex ANR grant (ANR-2008-DEFIS-004) has been extended until 2012. We have finalized important parts of our work on materialized views for XML [41], [35] and for the Semantic Web [21].

8.2.2. ANR ConnectedClouds

Participants: Dario Colazzo, François Goasdoué, Ioana Manolescu, Jesús Camacho_Rodriguez.

This one-year ANR grant (2011-2012) concerns our research on cloud-based data management, in particular XML data. We have studied strategies for indexing XML content in a cloud-environment and built a prototype to test them.

8.2.3. ANR DataBridges

Participants: François Goasdoué, Ioana Manolescu, Nathalie Pernelle, Gianluca Quercini, Chantal Reynaud, Brigitte Safar, Fatiha Saïs.

This one-year ANR grant (2011-2012) is devoted to research on data integration in particular through the technologies and models of Open Data, with a particular interest in applications connected to the Digital Cities.

8.2.4. ANR Dataring

Participants: Asterios Katsifodimos, Ioana Manolescu.

The Dataring ANR grant has been extended until 2012. Within Dataring, we have proposed scalable algorithms for automatically recommending materialized XML views, in collaboration with V. Vassalos (AUEB, Greece). This has led to a submitted publication.

8.2.5. ANR Geonto

Participants: Fayçal Hamdi, Nathalie Pernelle, Chantal Reynaud, Brigitte Safar, Fatiha Saïs.

This ANR Masses de Données et de Connaissances project (2008-2011) focused on geographic data interoperability. On one hand, we aim at integrating heterogeneous geographic databases using schema matching techniques. On the other hand, we aim at querying a large collection of textual documents which are more various and for a larger readership than databases. This project is a collaboration between COGIT-IGN (Sébastien Mustière), the IC3 group at IRIT - Université Paul Sabatier (Nathalie Aussenac) and the DESI group at LI-UPPA - Université de Pau et des Pays de l'Adour (Mauro Gaio). The home page of the project could be found at: <http://geonto.lri.fr>.

8.2.6. ANR PIMI

Participants: Rania Khefifi, Fatiha Saïs.

The objectives of PIMI (Personal Information Management Through Internet) ANR project (2010-2013) are the definition of a design environment and a deployment platform for Personal Information Management System (PIMS). The future PIMS must provide the end-user personal data access with services that are relevant to his needs. Ontologies will be used to describe semantically the services and the user needs. Ontology matching techniques will be defined to compare the services and the user needs during the automatic service composition. In order to take mobility into account, the PIMS will be accessed both by mobile devices (smartphone) and Internet-connected Personal Computers. This project is a collaboration between Leo team, ForTesSE team from LRI, IRIT, IT (Institut Telecom), GENIGRAPH, Montimage, Région Midi-Pyrénées and CTIE from Luxembourg

8.2.7. ANR UNLOC

Participant: Laurent Simon.

The UNLOC project (2008-2011) finishes this year. The project aims at studying and proposing new frameworks for SAT algorithms, based on more uncompleteness. The Glucose 1.0, GlucoseER and Glucose 2.0 systems were developed within this project.

8.2.8. Participation to evaluation committees

- C. Reynaud has participated to the evaluation committee of the non thematic ANR program "Blanc" Science Informatique et Application (SIMI2) 2011 and of the thematic ANR program "Contenus numériques et interactions" (CONTINT) 2011.

8.3. European Initiatives

8.3.1. FP7 Projects

8.3.1.1. Hycon2

Title: Highly Complex and Networked Control Systems

Type: COOPERATION (ICT)

Defi: Engineering of Networked Monitoring and Control Systems

Instrument: Network of Excellence (NoE)

Duration: September 2010 - August 2014

Coordinator: CNRS (France)

Others partners: INRIA (France), ETH Zurich (Switzerland), TU Berlin (Germany), TU Delft (Netherlands) and many others

See also: <http://www.hycon2.eu>

Abstract: Hycon 2 aims at stimulating and establishing a long-term integration in the strategic field of control of complex, large-scale, and networked dynamical systems. It focuses in particular on the domains of ground and aerospace transportation, electrical power networks, process industries, and biological and medical systems.

8.3.1.2. *Webdam*

Title: Foundations of Web Data Management

Type: IDEAS

Instrument: ERC Advanced Grant (Advanced)

Duration: December 2008 - November 2013

Coordinator: INRIA (France)

Others partners: N/A

See also: <http://webdam.inria.fr>

Abstract: The goal of the Webdam project is to develop a formal model for Web data management. This model will open new horizons for the development of the Web in a well-principled way, enhancing its functionality, performance, and reliability. Specifically, the goal is to develop a universally accepted formal framework for describing complex and flexible interacting Web applications featuring notably data exchange, sharing, integration, querying and updating. We also propose to develop formal foundations that will enable peers to concurrently reason about global data management activities, cooperate in solving specific tasks and support services with desired quality of service. Although the proposal addresses fundamental issues, its goal is to serve as the basis for future software development for Web data management. *S. Abiteboul has left Leo in September 2011. From this date, WebDam is no longer formally connected to the team, although collaborations with them continue.*

8.3.2. *Collaborations in European Programs, except FP7*

8.3.2.1. *KIC EIT ICT Labs DataBridges*

Program: KIC EIT ICT Labs, "Digital Cities" Thematic Action Line

Project acronym: DataBridges

Project title: "Data Integration for Digital Cities"

Duration: January 2011 - December 2011

Coordinator: Ioana Manolescu (INRIA Saclay, Leo)

Other partners: IASI and BD teams from University of Paris 11; DFKI (Germany); TU Delft (The Netherlands); Aalto University (Finland); KTH (Sweden); Alcatel (France)

Abstract: Digital cities are information exchange marketplaces where companies, individuals, and administrations all interact through many-directional flows. Interesting recent development in this context are: the open data trend, aiming at making data freely shared by several parties, and the linked data technical initiative, leading to establishing connections across data sets toward integrating them. Our activity studies concrete and fundamental aspects connected to the creation, integration, personalization, and efficient sharing of open linked data in digital cities.

8.3.2.2. KIC EIT ICT Labs ConnectedCities

Program: KIC EIT ICT Labs, “Digital Cities” Thematic Action Line

Project acronym: ConnectedClouds

Project title: “Clouds for Connected Cities”

Duration: January 2011 - December 2011

Coordinator: Djamel Zeglache (Télécom SudParis)

Other partners: IASI and BD teams from University of Paris 11; TU Berlin (Germany) etc.

Abstract: Digital cities are information exchange marketplaces where companies, individuals, and administrations all interact through many-directional flows. The ConnectedClouds activity aims at developing tools and techniques for efficiently gathering and managing large volumes of data within cloud-based platforms [62].

8.3.3. Major European Organizations with which you have followed Collaborations

Volker Markl: Technical University of Berlin, Database and Information Systems Lab (Germany)

Efficient management of Web data in a cloud environment

Cédric Pruski: Centre de Recherche Public Henri Tudor (CRP) (Luxembourg)

Reconciliation of Dynamic Medical Knowledge Organizing Systems

Giorgio Ghelli: Università di Pisa (Italy)

Type analysis of XML queries.

Carlo Sartiani: Università della Basilicata (Italy)

Type analysis of XML queries.

Vasilis Vassalos: Athens University of Economics and Business (Greece)

Materialized XML views: rewriting and view recommendation

8.4. International Initiatives

8.4.1. Visits of International Scientists

8.4.1.1. Internships

Francesca Bugiotti (from April 2011 until August 2011)

- Subject: Cloud-based management of RDF data
- Institution: Università di Roma Tré (Italy)

Giovanna Guerrini (from June 2011 until July 2011)

- Subject: XML update optimization
- Institution: Università di Genova (Italy)

Alexandra Roatis (from Mar 2011 until Aug 2011)

- Subject: Semantic Web Data Management on ViP2P
- Institution: West Timisoara University (Romania)

Vasiliki Papavasileou (from July 2011 until August 2011)

- Subject: Efficient propagation of updates to multiple views
- Institution: University of California in San Diego (USA)

Stamatis Zampetakis (from Jan 2011 until August 2011)

- Subject: Models and Algorithms for Annotated Documents
- Institution: University of Crete (Greece)

9. Dissemination

9.1. Animation of the scientific community

9.1.1. Participation in Editorial Boards

I. Manolescu is the editor in chief of the ACM SIGMOD Record and a member of the editorial board of ACM Transactions on Internet Technology. She was on the Editorial Board of VLDB 2011.

C. Reynaud is a member of the scientific committee of the special issue of the Applied Ontology Journal - Ontologies and Terminologies: Continuum or Dichotomy

L. Simon is a member of the Editorial Board of JSAT (the Journal on Satisfiability, Boolean Modeling and Computation)

9.1.2. Participation in Conference Organization

Members of the project have been chairs of scientific events:

S. Abiteboul has been PC chair of the International Conference on Data Engineering (ICDE) 2011. He was also track chair at the International World Wide Web Conference (WWW) 2011.

I. Manolescu has been the chair of the “Experiments, Applications and Systems” track of the IEEE International Data Engineering Conference (ICDE) 2011.

F. Suchanek was co-chair of the 4th PhD Workshop on Information and Knowledge Management (PIKM 2011).

L. Simon has been the co-chair (organization and program) of SAT 2011, Ann Arbor, USA

Members of the project have participated in program committees:

Meghyn Bienvenu

- (Senior PC) Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI) 2011
- Twenty-Fifth Conference on Artificial Intelligence (AAAI) 2011
- Twenty-Fourth International Workshop on Description Logics (DL) 2011
- First International Workshop on Data, Logic and Inconsistency (DALI) 2011

Philippe Chatalic

- Sixièmes Journées Francophones sur les Modèles Formels de l’Interaction (MFI) 2011
- Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI) 2011
- Twenty-Third IEEE International Conference on Tools with Artificial Intelligence (ICTAI) 2011

Dario Colazzo

- Nineteenth Italian Symposium on Advanced Database Systems (SEBD) 2011.
- Twenty-Second Australasian Database Conference (ADC) 2011.

Philippe Dague

- Twenty-Fifth International Workshop on Qualitative Reasoning (QR) 2011
- Twenty-Second International Workshop on Principles of Diagnosis (DX) 2011
- Dix-Huitième Congrès Francophone Reconnaissance des Formes et Intelligence Artificielle (RFIA) 2012

François Goasdoué

- Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI) 2011
- Twenty-Third IEEE International Conference on Tools with Artificial Intelligence (ICTAI) 2011

Ioana Manolescu

- Twentieth ACM Conference on Information and Knowledge Management (CIKM) 2011
- 27èmes Journées de Bases de Données Avancées (BDA) 2011
- ICDE 2011 Workshop on Data Management through its Life Cycle (DaLi) 2011
- Thirteenth Asia-Pacific Web Conference (APWeb) 2011

Nathalie Pernelle

- International Conference on Advances in Information Mining and Management (IMM) 2011
- Evaluation des méthodes d'Extraction des Connaissances dans les Données (EvalECD) workshop in collaboration with EGC 2011
- Qualité des Données et des Connaissances (QDC) workshop in collaboration with EGC 2011
- Web social workshop in collaboration with EGC 2011

Chantal Reynaud

- 8th Extended Semantic Web Conference - Ontologies Track (ESWC) 2011
- 1st International Conference on Model and data Engineering (MEDI) 2011
- Conférence Extraction et Gestion des Connaissances (EGC) 2011
- 22èmes Journées Francophones d'Ingénierie des Connaissances (IC) 2011
- 29ème Congrès INFORMATIQUE des ORGANISATIONS et SYSTÈMES d'INFORMATION et de DÉCISION (IN-FORSID) 2011
- Journées Internationales Francophones sur les Ontologies (JFO) 2011
- Evaluation des méthodes d'Extraction des Connaissances dans les Données (EvalECD) workshop in collaboration with EGC 2011
- Qualité des Données et des Connaissances (QDC) workshop in collaboration with EGC 2011

Brigitte Safar

- Journées Internationales Francophones sur les Ontologies (JFO) 2011

Fatiha Saïs

- International Workshop on Resource Discovery (RED) collocated with the international conference ESWC 2011.
- Conférence Francophone sur Extraction et Gestion des Connaissances (EGC) 2011
- Conférence Maghrébine sur Extraction et Gestion des Connaissances (EGC-M) 2011
- Social Web workshop collocated with EGC 2011 Conference.
- Optimization and Information Systems (COSI) 2011
- Qualité et Robustesse dans le Web de Données workshop collocated with the national conference IC 2011.
- Cotation des Informations : Théorie et Applications workshop collocated with the national conference IC 2011.

- Organization of the three editions of Evaluation des méthodes d'Extraction des Connaissances dans les Données (EvalECD) workshop in collaboration with EGC 2009, 2010 and 2011.

Laurent Simon

- (Senior PC) Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI) 2011
- Twenty-Fifth Conference on Artificial Intelligence (AAAI) 2011

Members of the project participate in steering committees:

Meghyn Bienvenu

- International Workshop on Description Logics (DL), elected 2009-2012

Ioana Manolescu

- International Conference on Extending Database Technologies (EDBT), since 2010
- *Bases de Données Avancées* (BDA) 2011

Laurent Simon

- International Conference on Theory and Application of Satisfiability Testing since 2011

9.1.3. Invited Presentations

M. Bienvenu gave an invited presentation at *Journées Nationales de l'IA Fondamentale (IAF)*, 2011.

I. Manolescu gave an invited presentation at *International Workshop on Data Engineering Meets the Semantic Web (DeSWeb)* in conjunction with IEEE ICDE 2011, and a keynote talk at the Digiteo Forum 2011.

C. Reynaud gave an invited presentation at *Journée Web et Médiation*, Rhône-Alpes Web Intelligence Project, April 2011.

L. Simon gave an invited presentation at NII, Japan, 2011.

9.2. Teaching

Serge Abiteboul

Licence (ou équivalent): Bases de données, L3, ENS Cachan, France

Master (ou équivalent): Web data management, 36h éq TD, M2, Université Paris Sud, France

Meghyn Bienvenu

Master (ou équivalent): Modèles de raisonnement distribué, 9h éq TD, M2, Université Paris Sud, France

Philippe Chatalic

Licence et Master: Introduction à l'intelligence artificielle, 66h éq TD, L3&M1, Université Paris Sud, France

Master (équivalent) : Résolution de contraintes, 27h éq TD, M2, Université Paris Sud, France

Master: Programmation par contraintes, 91h éq TD, M1, Université Paris Sud, France

Master (ou équivalent): Modèles de raisonnement distribué, 5h éq TD, M2, Université Paris Sud, France

Dario Colazzo

Master (ou équivalent): Bases de données avancées: implementation, 21h éq TD, Formation en alternance par l'apprentissage, POLYTECH Paris-Sud, France.

Licence (ou équivalent): Bases de données, 27h éq TD, DEUST, Université Paris Sud, France.

Master (ou équivalent): Données semi-structurées et XML : langages et optimisation. M2, 16h éq TD, Université Paris Sud, France.

Philippe Dague

Licence (ou équivalent): Langages formels et analyse syntaxique, 25,5h éq TD, L3, Université Paris-Sud, France

Master (ou équivalent): Automates et langages, 37,5h éq TD, Polytech Paris-Sud, France

Master (ou équivalent): Modèles de raisonnement distribué, 4,5h éq TD, M2, Université Paris-Sud, France

Master (ou équivalent): Test et diagnosticabilité à base de modèles, 6h éq TD, M2, Université Paris-Sud, France

François Goasdoué

Licence (ou équivalent): Bases de données, 63h éq TD, L3, Université Paris Sud, France

Master (ou équivalent): Web sémantique, 74h éq TD, M2, Université Paris Sud, France

Master (ou équivalent): Données et connaissances sur le Web, 8h éq TD, M2, Université Paris Sud, France

Master (ou équivalent): Modèles de raisonnement distribué, 3h éq TD, M2, Université Paris Sud, France

Master (ou équivalent): Gestion sémantique de données, 9h éq TD, M2, Université Paris Dauphine, France

Master (ou équivalent): Logiques de description pour le Web sémantique, 30h éq TD, M2, Université Libanaise de Beyrouth, Liban

Ioana Manolescu

Master (ou équivalent): Données semi-structurées et XML: langages et optimisation. M2, 18h éq TD, Université Paris Sud, France

Master (ou équivalent): Services Web. M2, 18h éq TD, Université Paris Dauphine, France.

Nathalie Pernelle

Licence (ou équivalent): Bases de données appliquées , 20h éq TD, L3 (Licence Pro), IUT Sceaux, France

Licence (ou équivalent): Analyse et conception de Systèmes d'Information, 75h éq TD, 2ème année DUT GEA, IUT Sceaux, France

Licence (ou équivalent): Fonctionnement Informatisé de l'Entreprise, 9h éq TD, 2ème année DUT GEA, IUT Sceaux, France

Licence (ou équivalent): Jeu d'entreprise, 16h éq TD, 1ère année DUT GEA, IUT Sceaux, France

Licence (ou équivalent): Systèmes d'Information, 15h éq TD, L3 DUGFM, IUT Sceaux, France

Licence (ou équivalent): Projet Professionnel Personnalisé, 17h éq TD, L1 and L2, IUT Sceaux, France

Master (ou équivalent): Construction et Alignement d'ontologies pour le Web Sémantique, 9h éq TD, M2, Université Paris Dauphine, France

Master (ou équivalent) : Réconciliation de réf' erences, 3h éq TD, M2, Université Paris Sud, France

Chantal Reynaud

- Licence (ou équivalent): Projet tutoré, 8h TD, 1ère année DUT Informatique, IUT Orsay, France
- Licence (ou équivalent): Analyse et Conception de Systèmes d'Information, 13h CM et 105h TD, 2ème année DUT Informatique, IUT Orsay, France
- Licence (ou équivalent): Analyse et Conception de Systèmes d'Information, 20h TD, Licence professionnelle, IUT Orsay, France
- Licence (ou équivalent): Synthèse, 18h TD, 2ème année DUT Informatique, IUT Orsay, France
- Master (ou équivalent): Données et connaissances sur le Web, 8h éq TD, M2, Université Paris Sud, France
- Master (ou équivalent): Intégration de données et Web sémantique, 3h CM, M2, Université Paris Sud, France

Fatima Saïs

- Licence (ou équivalent): Projet tutoré, 8h TD, 1ère année DUT Informatique, IUT Orsay, France
- Licence (ou équivalent): Analyse et Conception de Systèmes d'Information, 63h TD, 2ème année DUT Informatique, IUT Orsay, France
- Licence (ou équivalent): Analyse et Conception de Systèmes d'Information, 42h TD, 1ère année DUT Informatique (DUCI), IUT Orsay, France
- Licence (ou équivalent): Systèmes et Réseaux, 22h TP, 1ère année DUT Informatique, IUT Orsay, France
- Licence (ou équivalent): Programmation Web, 45h TD, 2ème année DUT Informatique, IUT Orsay, France
- Licence (ou équivalent): Bases de Données Réparties, 18h TD, Licence professionnelle, IUT Orsay, France
- Licence (ou équivalent): Synthèse, 18h TD, 2ème année DUT Informatique, IUT Orsay, France
- Master (ou équivalent): Web Sémantique, 6h CM, M2, Université Paris-Dauphine, France
- Master (ou équivalent): Contraintes et Résolution, 16h CM, 16 TD et 8 TP, Ingénieurs-apprentis 5ème année, PolyTech Paris-Sud, France

Laurent Simon

- Master : Intelligence Artificielle et Robotique, 20h éq TD, M1, Polytech Paris Sud
- Licence-Master : Introduction à l'Intelligence Artificielle, 45h éq TD, L3-M1, Université Paris Sud
- Master : Intelligence Artificielle pour le temps réel et les jeux vidéos, 28h éq TD, M2, Polytech Paris Sud
- Licence : Introduction à l'informatique, 55h éq TD, L1, Université Paris Sud
- Master : Modèles de raisonnement distribué, 9h éq TD, M2, Université Paris Sud
- Master : Contraintes, 9h éq TD, M2, ENS Lyon
- Master : Résolution de Contraintes, 24h éq TD, M2, Polytech Paris Sud

Fabian Suchanek

- Master (ou équivalent): Modélisation cognitive et informatique intelligente, 7.5h éq TD, Télécom ParisTech, France
- Master (ou équivalent): Ingénierie des services Web, 14h éq TD, Télécom ParisTech, France
- Master (ou équivalent): Programmation logique et connaissances, 4.5h éq TD, Télécom ParisTech, France
- Master (ou équivalent): Web search, 7.5h éq TD, Télécom ParisTech, France

PhD & HdR:

HdR: Dario Colazzo, “Schemas for safe and efficient XML processing”, defended on September 8 [13].

PhD: Michel Batteux, “Diagnosticabilité et diagnostic de systèmes technologiques pilotés”, Université Paris-Sud, defended on December 13, 2011, supervised by P. Dague, N. Rapin (CEA) and P. Fiani (Sherpa Engineering) .

PhD: Alban Galland, “Distributed data management with access control : social Networks and Data of the Web”, defended on September 28, supervised by Serge Abiteboul [14].

PhD: Fayçal Hamdi, “Améliorer l’interopérabilité sémantique: Applicabilité et utilité de l’alignement d’ontologies”, defended on December 2, supervised by Chantal Reynaud and Brigitte Safar [15]

PhD: Wael Khemiri, “Data-intensive interactive workflows for visual analytics”, Université Paris-Sud, defended on December 12, 2011, supervised by V. Benzaken (Proval), J.-D. Fekete (INRIA Aviz) and I. Manolescu [16]

PhD: Marina Sahakyan, “ Main Memory XML Update Optimization: algorithms and experiments”, defended on November 17, 2011, supervised by Nicole Bidoit and Dario Colazzo .

PhD: Nadjet Zemirline, “Assisting in the reuse of Existing Materials to Build Adaptive Hypermedia”, Université Paris-Sud, July 12th 2011, Chantal Reynaud and Yolaine Bourda (Supelec) [19]

PhD in progress: Amine Baazizi, “Schemas for efficient updates and maintenance of temporal XML data”, October 2008, Nicole Bidoit and Dario Colazzo.

PhD in progress: Jesús Camacho-Rodríguez, “Efficient Cloud-based Management of XML data”, October 2011, Dario Colazzo and Ioana Manolescu

PhD in progress: Julio Cesar Dos Reis, “Reconciliation of Dynamic Medical Knowledge Organizing Systems”, October 2011, Chantal Reynaud and Cédric Pruski (Institut H. Tudor - Luxembourg)

PhD in progress: Konstantinos Karanasos, “View-based query processing for XML and RDF data”, January 2009, Ioana Manolescu and François Goasdoué

PhD in progress: Asterios Katsifodimos, “Efficient view-based techniques for XML query processing”, October 2009, Ioana Manolescu

PhD in progress: Rania Khefifi, “Context-aware Personal Information Management”, October 2010, Fatiha Saïs and Pascal Poizat (Evry-Val-d’Essonne University & LRI)

PhD in progress: Julien Leblay, “Query processing for XML documents with RDF annotations”, October 2010, François Goasdoué and Ioana Manolescu

PhD in progress: Noor Malla, “Projecting and Partitioning XML Data”, December 2008, Nicole Bidoit and Dario Colazzo.

PhD in progress: Yassine Mrabet, “ Approche hybride d’intégration pour la réconciliation et la RI dans les ressources hétérogènes”, October 2008, Chantal Reynaud, Nathalie Pernelle and Nacéra Bennacer (Supelec)

PhD in progress: Alexandra Roatis, “OnLine Analytical Processing for RDF warehouses”, September 2011, Dario Colazzo, François Goasdoué and Ioana Manolescu

PhD in progress: Danai Symeonidou, “Découverte automatique de contraintes pour la réconciliation de références”, October 2011, Fatiha Saïs and Nathalie Pernelle

PhD in progress: Federico Ulliana, “Type Based Access Analysis for XML Queries and Updates”, October 2009, Nicole Bidoit and Dario Colazzo.

PhD in progress: Vincent Armant, “Diagnostic Distribué”, October 2007, Philippe Dague and Laurent Simon.

10. Bibliography

Major publications by the team in recent years

- [1] S. ABITEBOUL, M. BIENVENU, A. GALLAND, E. ANTOINE. *A rule-based language for Web data management*, in "Principles of Database Systems", Athens, Greece, June 2011, <http://hal.inria.fr/inria-00582891/en>.
- [2] S. ABITEBOUL, I. MANOLESCU, P. RIGAUX, M.-C. ROUSSET, P. SENELLART. *Web Data Management and Distribution*, Cambridge University Press, 2011, <http://hal.inria.fr/inria-00542796/en>.
- [3] A. BAAZIZI, N. BIDOIT, D. COLAZZO, N. MALLA, M. SAHAKYAN. *Projection for XML Update Optimization.*, in "International Conference on Extending Database Technology (EDBT)", Uppsala, Sweden, February 2011, <http://hal.inria.fr/inria-00626225/en>.
- [4] J. DAVIES, N. NARODYTSKA, G. KATSIRELOS, T. WALSH. *Complexity of and Algorithms for Borda Manipulation*, in "AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE", San-Francisco, États-Unis, August 2011, <http://hal.inria.fr/hal-00642093/en/>.
- [5] F. GOASDOUÉ, K. KARANASOS, J. LEBLAY, I. MANOLESCU. *View Selection in Semantic Web Databases*, in "Proceedings of the VLDB Endowment (PVLDB)", October 2011, vol. 5, n^o 2, <http://hal.inria.fr/inria-00625090/en>.
- [6] F. HAMDI, C. REYNAUD, B. SAFAR. *Pattern-based Mapping Refinement*, in "International Conference on Knowledge Engineering and Knowledge Management (EKAW)", Lisbon Portugal, 10 2010.
- [7] F. HAMDI, B. SAFAR, C. REYNAUD, H. ZARGAYOUNA. *Alignment-based Partitioning of Large-scale Ontologies*, in "Advances in Knowledge Discovery And Management", F. GUILLET, G. RITSCHARD, D. ZIGHED, H. BRIAND (editors), Springer, 2010, vol. 292, p. 251-269, <http://dx.doi.org/10.1007/978-3-642-00580-0>.
- [8] A. MADALINSKI, F. NOUIOUA, P. DAGUE. *Diagnosability verification with Petri net unfoldings*, in "International Journal of Knowledge-Based and Intelligent Engineering Systems, KES", 2010, vol. 14, n^o 2, p. 49-55, <http://hal.inria.fr/inria-00540650/en>.
- [9] I. MANOLESCU, K. KARANASOS, V. VASSALOS, S. ZOUPANOS. *Efficient XQuery Rewriting using Multiple Views*, in "International Conference on Data Engineering (ICDE)", Hannover, Germany, 2011, <http://hal.inria.fr/inria-00544366/en>.
- [10] Y. MRABET, N. BENNACER, N. PERNELLE, M. THIAM. *Supporting Semantic Search on Heterogeneous Semi-structured Documents*, in "22nd International Conference on Advanced Information Systems Engineering (CAISE)", Tunisie Hammamet, LNCS, Springer-Verlag, Jun 2010, p. 224-229, <http://hal.inria.fr/hal-00502286/en>.

Publications of the year

Doctoral Dissertations and Habilitation Theses

- [11] M. BATTEUX. *Diagnosticabilité et diagnostic de systèmes technologiques pilotés*, Université Paris Sud-11, December 2011.

- [12] M. BATTEUX. *Diagnosticabilité et diagnostic de systèmes technologiques pilotés*, Université Paris Sud - Paris XI, December 2011, <http://hal.inria.fr/tel-00643460/en>.
- [13] D. COLAZZO. *Schemas for safe and efficient XML processing*, Université Paris Sud-11, September 2011, Habilitation à Diriger des Recherches, <http://hal.inria.fr/tel-00626227/en>.
- [14] A. GALLAND. *Distributed data management with access control : Social Networks and Data of the Web*, Université Paris Sud-11, September 2011, <http://hal.inria.fr/tel-00640725/en>.
- [15] F. HAMDI. *Améliorer l'interopérabilité sémantique : Applicabilité et utilité de l'alignement d'ontologies*, Université Paris Sud-11, December 2011.
- [16] W. KHEMIRI. *Data-intensive interactive workflows for visual analytics*, Université Paris Sud-11, December 2011.
- [17] M. SAHAKYAN. *Main Memory XML Update Optimization: algorithms and experiments.*, Université Paris Sud-11, November 2011.
- [18] L. YE. *Optimized Diagnosability of Distributed Discrete Event Systems Through Abstraction*, Université Paris Sud-11, July 2011, <http://tel.archives-ouvertes.fr/tel-00635695/en/>.
- [19] N. ZEMIRLINE. *Aide à la Création d'Hypermédia Adaptatifs par Réutilisation des Modèles des Créateurs*, Université Paris Sud - Paris XI, July 2011, <http://tel.archives-ouvertes.fr/tel-00641569/en/>.

Articles in International Peer-Reviewed Journal

- [20] M. BIENVENU, C. FRITZ, S. MCILRAITH. *Specifying and computing preferred plans*, in "Artificial Intelligence", May 2011, vol. 175, n^o 7-8, p. 1308-1345 [DOI : 10.1016/J.ARTINT.2010.11.021], <http://hal.inria.fr/hal-00624481/en>.
- [21] F. GOASDOUÉ, K. KARANASOS, J. LEBLAY, I. MANOLESCU. *View Selection in Semantic Web Databases*, in "Proceedings of the VLDB Endowment (PVLDB)", October 2011, vol. 5, n^o 2, <http://hal.inria.fr/inria-00625090/en>.
- [22] N. GUELFY, C. PRUSKI, C. REYNAUD. *Adaptive Ontology-based Web Information Retrieval: The TARGET Framework*, in "International Journal of Web Portals", 2011, vol. 3, n^o 3, p. 41-58 [DOI : 10.4018/IJWP.2011070104], <http://hal.inria.fr/inria-00624102/en>.
- [23] B. NGUYEN, A. VION, F.-X. DUDOUET, D. COLAZZO, I. MANOLESCU, P. SENELLART. *XML content warehousing: Improving sociological studies of mailing lists and web data*, in "Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique", October 2011, vol. 112, n^o 1, p. 5-31 [DOI : 10.1177/0759106311417540], <http://hal.inria.fr/hal-00616613/en>.
- [24] N. ZEMIRLINE, Y. BOURDA, C. REYNAUD. *Expressing Adaptation Strategies using Adaptation Patterns*, in "IEEE Transactions on Learning Technologies", 2011, p. 14 pages - (accepté, à paraître) [DOI : 10.1109/TLT.2011.15], <http://hal.inria.fr/hal-00583450/en>.

Articles in National Peer-Reviewed Journal

- [25] M. SÉBASTIEN, A. NATHALIE, N. AUSSÉNAC- GILLES, M.-N. BESSAGNET, M. KAMEL, E. KERGOISIEN, C. REYNAUD, B. SAFAR, C. SALLABERRY. *Analyses linguistiques et techniques d'alignement pour créer et enrichir une ontologie topographique*, in "Revue internationale de Géomatique", 2011, <http://hal.inria.fr/inria-00595964/en>.

International Peer-Reviewed Conference/Proceedings

- [26] S. ABITEBOUL, M. BIENVENU, A. GALLAND, E. ANTOINE. *A rule-based language for Web data management*, in "Principles of Database Systems", Athens, Greece, June 2011, <http://hal.inria.fr/inria-00582891/en>.
- [27] S. ABITEBOUL, M. BIENVENU, A. GALLAND, M.-C. ROUSSET. *Distributed Datalog Revisited*, in "Datalog 2.0 Workshop", Oxford, United Kingdom, 2011, <http://hal.inria.fr/inria-00540814/en>.
- [28] E. ANTOINE, A. GALLAND, K. LYNGBAEK, A. MARIAN, N. POLYZOTIS. *[Demo] Social Networking on top of the WebdamExchange System*, in "International Conference on Data Engineering", Hannover, Germany, April 2011, <http://hal.inria.fr/inria-00536361/en>.
- [29] A. BAAZIZI, N. BIDOIT, D. COLAZZO. *Efficient Encoding of Temporal XML documents*, in "International Symposium on Temporal Representation and Reasoning (TIME)", Lubeck, Germany, September 2011, <http://hal.inria.fr/inria-00626226/en>.
- [30] A. BAAZIZI, N. BIDOIT, D. COLAZZO. *Maintenance efficace de documents XML volumineux*, in "27èmes journées Bases de Données Avancées", Rabat, Maroc, October 2011, <http://hal.inria.fr/hal-00641602/en/>.
- [31] A. BAAZIZI, N. BIDOIT, D. COLAZZO, N. MALLA, M. SAHAKYAN. *Projection for XML Update Optimization.*, in "International Conference on Extending Database Technology (EDBT)", Uppsala, Sweden, February 2011, <http://hal.inria.fr/inria-00626225/en>.
- [32] M. BATTEUX, P. DAGUE, N. RAPIN, P. FIANI. *Diagnosability study of technological systems*, in "24th International Conference on Industrial, Engineering and other Applications of Applied Intelligent Systems IEA/AIE 2011", Syracuse, United States, June 2011, vol. 6703, <http://hal.inria.fr/hal-00643664/en>.
- [33] V. BENZAKEN, J.-D. FEKETE, P.-L. HMERY, W. KHEMIRI, I. MANOLESCU. *EdiFlow: data-intensive interactive workflows for visual analytics*, in "International conference on Data Engineering", Hannover, Germany, IEEE, April 2011, <http://hal.inria.fr/inria-00532552/en>.
- [34] N. BIDOIT. *Types and constraints: from relational to XML data*, in "Semantics in Data and Knowledge Bases - 4th International Workshops, SDKB 2010, Revised Selected Papers.", Bordeaux, France, May 2011, <http://hal.inria.fr/hal-00641909/en/>.
- [35] A. BONIFATI, M. GOODFELLOW, I. MANOLESCU, D. SILEO. *Algebraic incremental maintenance of XML views*, in "EDBT: 14th International Conference on Extending Database Technology", Uppsala, Sweden, A. AILAMAKI, S. AMER-YAHIA, J. M. PATEL, T. RISCH, P. SENELLART, J. STOYANOVICH (editors), March 2011, <http://hal.inria.fr/inria-00624986/en>.
- [36] P. BUCHE, J. DIBIE-BARTHÉLEMY, R. KHEFIFI, F. SAÏS. *An Ontology-Based Method for Duplicate Detection in Web Data Tables*, in "DEXA'11: 22nd DEXA Conference", Toulouse, France, A. HAMEURLAIN, S. W. LIDDLE, K.-D. SCHEWE, X. ZHOU (editors), LNCS, Springer-Verlag, August 2011, p. 511-525, <http://hal.inria.fr/lirmm-00611944/en>.

- [37] D. COLAZZO, C. SARTIANI. *Precision and Complexity of XQuery Type Inference.*, in "ACM-SIGPLAN Symposium on Principles and Practice of Declarative Programming (PPDP)", Odense, Denmark, July 2011, <http://hal.inria.fr/inria-00626217/en>.
- [38] J. DAVIES, N. NARODYTSKA, G. KATSIRELOS, T. WALSH. *Complexity of and Algorithms for Borda Manipulation*, in "AAAI conference on artificial intelligence", San-Francisco, États-Unis, August 2011, <http://hal.inria.fr/hal-00642093/en/>.
- [39] F. GOASDOUÉ, K. KARANASOS, Y. KATSIKIS, J. LEBLAY, I. MANOLESCU, S. ZAMPETAKIS. *Growing Triples on Trees: an XML-RDF Hybrid Model for Annotated Documents*, in "First International Workshop on Searching and Integrating New Web Data Sources", Seattle, United States, M. BRAMBILLA, F. CASATI, S. CERI (editors), September 2011, <http://hal.inria.fr/inria-00624996/en>.
- [40] J. HOFFART, F. SUCHANEK, K. BERBERICH, E. LEWIS-KELHAM, G. DE MELO, G. WEIKUM. *YAGO2: Exploring and Querying World Knowledge in Time, Space, Context, and Many Languages*, in "WWW - Demos", Hyderabad, India, 2011, <http://hal.inria.fr/inria-00591780/en>.
- [41] I. MANOLESCU, K. KARANASOS, V. VASSALOS, S. ZOUPANOS. *Efficient XQuery Rewriting using Multiple Views*, in "International Conference on Data Engineering (ICDE)", Hannover, Germany, 2011, <http://hal.inria.fr/inria-00544366/en>.
- [42] N. PERNELLE, F. SAÏS. *LDM: Link Discovery Method for new Resource Integration*, in "Fourth International Workshop on Resource Discovery", heraklion, Greece, Z. LACROIX, E. RUCKHAUS, M.-E. VIDAL (editors), CEUR-WS, May 2011, vol. vol 737, p. pp 94-108, <http://hal.inria.fr/inria-00625689/en>.
- [43] F. SUCHANEK, D. GROSS-AMBLARD, S. ABITEBOUL. *Watermarking for Ontologies*, in "ISWC 2011: Proceedings of the 10th International Semantic Web Conference", 2011.
- [44] F. SUCHANEK, M. THEOBALD, G. WEIKUM, H. LAUW, R. SCHENKEL. *Semantic Knowledge Bases from Web Sources*, in "IJCAI", Barcelona, Spain, SIG, 2011, Tutorial, <http://hal.inria.fr/inria-00539623/en>.
- [45] F. SUCHANEK, A. VARDE, R. NAYAK, P. SENELLART. *The Hidden Web, XML and Semantic Web: A Scientific Data Management Perspective*, in "EDBT - Tutorial", Uppsala, Sweden, 2011, <http://hal.inria.fr/inria-00591776/en>.
- [46] D. SYMEONIDOU, N. PERNELLE, F. SAÏS. *KD2R : a Key Discovery method for semantic Reference Reconciliation*, in "7th International IFIP Workshop on Semantic Web & Web Semantics (SWWS 2011)", Heraklion, Greece, LNCS, Springer Verlag, October 2011, (10 pages, à paraître), <http://hal.inria.fr/inria-00626199/en>.
- [47] N. ZEMIRLINE, Y. BOURDA, C. REYNAUD. *A set of adaptation patterns for expressing adaptive navigation in Adaptive Hypermedia*, in "International Workshop on Dynamic and Adaptive Hypertext - Generic Frameworks, Approaches and Techniques (DAH'11), 22nd ACM Conference on Hypertext and Hypermedia", Eindhoven, Netherlands, 2011, 12 pages, <http://hal.inria.fr/hal-00627465/en>.
- [48] Z. ZHU, C. M. LI, F. MANYA, L. SIMON. *Minimum Satisfiability and its Applications*, in "International Joint Conference on Artificial Intelligence", Barcelona, Espagne, July 2011, <http://hal.inria.fr/hal-00642014/en/>.

National Peer-Reviewed Conference/Proceedings

- [49] M. BATTEUX, P. FIANI, N. RAPIN, P. DAGUE. *Génération du comportement observable d'un système pour l'étude de la diagnosticabilité de défauts*, in "QUALITA 2011", Angers, France, March 2011, <http://hal.inria.fr/hal-00643674/en>.
- [50] F. HAMDY, B. SAFAR, C. REYNAUD. *Utiliser des résultats d'alignement pour enrichir une ontologie*, in "11ème Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances (EGC'2011)", Brest, France, D. A. ZHIGED, G. VENTURINI (editors), RNTI, Hermann, January 2011, vol. n°E.20, p. 407-412, <http://hal.inria.fr/inria-00564843/en>.

Workshops without Proceedings

- [51] M. BIENVENU. *First-Order Expressibility Results for Queries over Inconsistent DL-Lite Knowledge Bases*, in "24th International Workshop on Description Logics (DL 2011)", 2011, <http://hal.archives-ouvertes.fr/hal-00640838/en/>.
- [52] F. GOASDOUÉ, K. KARANASOS, Y. KATSIKIS, J. LEBLAY, I. MANOLESCU, S. ZAMPETAKIS. *Growing Triples on Trees: an XML-RDF Hybrid Model for Annotated Documents*, in "Journées de Bases de Données Avancées", Rabat, Morocco, October 2011, <http://hal.inria.fr/inria-00624991/en>.
- [53] F. GOASDOUÉ, K. KARANASOS, J. LEBLAY, I. MANOLESCU. *RDFViewS: A Storage Tuning Wizard for RDF Applications*, in "Journées de Bases de Données Avancées", Rabat, Morocco, October 2011, <http://hal.inria.fr/inria-00624998/en>.

Scientific Books (or Scientific Book chapters)

- [54] S. ABITEBOUL, I. MANOLESCU, P. RIGAUX, M.-C. ROUSSET, P. SENELLART. *Web Data Management and Distribution*, Cambridge University Press, 2011, <http://hal.inria.fr/inria-00542796/en>.
- [55] F. SUCHANEK, A. BOZZON, E. DELLA VALLE, A. CAMPI, S. RONCHI. *Towards an Ontological Representation of Services in Search Computing*, in "New Trends in Search Computing", S. CERI, M. BRAMBILLA (editors), Springer, 2011, <http://hal.inria.fr/inria-00591790/en>.

Research Reports

- [56] F. GOASDOUÉ, K. KARANASOS, J. LEBLAY, I. MANOLESCU. *View Selection in Semantic Web Databases*, INRIA, September 2011, n° RR-7738, <http://hal.inria.fr/inria-00623942/en>.
- [57] F. SUCHANEK, S. ABITEBOUL, P. SENELLART. *Ontology Alignment at the Instance and Schema Level*, INRIA, August 2011, n° RT-0408, Technical Report at INRIA RT-0408, <http://hal.inria.fr/inria-00596210/en>.
- [58] D. SYMEONIDOU. *KD2R: a Key Discovery method for semantic Reference Reconciliation in OWL*, Université Paris-Sud, September 2011, <http://hal.inria.fr/hal-00641858/en/>.
- [59] N. ZEMIRLINE, Y. BOURDA, C. REYNAUD. *A typology of adaptation patterns for expressing adaptive navigation in Adaptive Hypermedia*, Université Paris-Sud, 2011, 1541, <http://hal.inria.fr/hal-00627457/en>.

- [60] N. ZEMIRLINE, Y. BOURDA. *Expressing Adaptation Strategies Using Adaptation Patterns*, Université Paris Sud-11, 2011, n^o 1540, <http://hal.inria.fr/hal-00627080/en>.

Other Publications

- [61] N. BENNACER, Y. MRABET, N. PERNELLE, M. THIAM. *Shiri: système hybride d'intégration et de recherche d'information*, 2011, poster DIGITEO forum 2011 (best poster award).
- [62] J. CAMACHO-RODRÍGUEZ, D. COLAZZO, I. MANOLESCU. *Building Large XML Stores in the Amazon Cloud*, 2011, Technical report.
- [63] D. COLAZZO, G. GHELLI, C. SARTIANI. *Schemas for Safe and Efficient XML Processing*, April 2011, Tutorial - IEEE International Conference on Data Engineering (ICDE), <http://hal.inria.fr/inria-00626218/en>.
- [64] C. MOREL. *Reconnaissance de types de données dans des tableaux de l'INSEE*, Université Paris Sud-11, Lille, September 2011, <http://hal.inria.fr/hal-00641101/en/>.