



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team gemo

*Management of Data and Knowledge
Distributed Over the Web*

Saclay - Île-de-France

THEME SYM

Activity
R *eport*

2008

Table of contents

1. Team	1
2. Overall Objectives	2
2.1. Introduction	2
2.2. Highlights of the year	2
3. Scientific Foundations	2
4. Application Domains	3
5. Software	3
6. New Results	4
6.1. Theoretical foundations	4
6.2. Ontology-Based Information Retrieval	4
6.2.1. Query Enrichment based on Adaptive Ontologies	4
6.2.2. Semantic Annotation	4
6.3. Peer-to-Peer Inference Systems	5
6.3.1. Consequence Finding	5
6.3.2. P2P conservative extension checking	5
6.3.3. Trust modeling	6
6.3.4. Distributed Diagnosis	6
6.3.5. Mapping distributed ontologies	6
6.4. Thematic Web Warehousing	6
6.4.1. Reference reconciliation	6
6.4.2. Mapping between ontologies	7
6.4.3. Viewpoints corroboration	7
6.5. XML query optimization	7
6.5.1. Static analysis for XML query optimization	7
6.5.2. Algebraic optimization for ActiveXML	8
6.5.3. Performance evaluation methodology	8
6.6. XML Warehousing in P2P	8
6.6.1. XML indexing: KadoP	8
6.6.2. XML views on DHTs: VIP2P	8
6.7. Monitoring and Web services	8
6.7.1. Error diagnosis and self-healing	9
6.7.2. Diagnosability	9
6.7.3. P2P Monitoring	10
7. Contracts and Grants with Industry	10
7.1. Industrial contracts	10
7.2. RNTL Project WebContent	10
7.3. WS-DIAMOND EC project	10
7.4. ANR-PREDIT DIAFORE project	11
8. Other Grants and Activities	11
8.1. National Actions	11
8.1.1. DocFlow	11
8.1.2. ACI Normes Pratiques et Régulations des Politiques Publiques	11
8.1.3. ANR JCJC WebStand	11
8.1.4. EDIFlow Digiteo project	12
8.1.5. SHIRI Digiteo project	12
8.1.6. GEONTO project	12
8.2. European Commission Financed Actions	12
8.3. Bilateral International Relations	12
8.3.1. Cooperation within Europe	12

8.3.2.	Cooperation with Senegal	13
8.3.3.	Cooperation with the Middle-East	13
8.3.4.	Cooperation with North America	13
8.3.5.	French-US team: GemSaD	13
8.4.	Visiting Professors and Students	13
9.	Dissemination	13
9.1.	Thesis	13
9.2.	Participation in Conferences	13
9.3.	Invited Presentations	14
9.4.	Scientific Animations	14
10.	Bibliography	15

1. Team

Research Scientist

Serge Abiteboul [DR-INRIA, HdR]
Ioana Manolescu [CR-INRIA, co-manager]

Faculty Member

Philippe Chatalic [Assistant Professor, Univ. Paris 11]
Philippe Dague [Professor, Univ. Paris 11, HdR]
Hélène Gagliardi [Assistant Professor, Univ. Paris 11]
François Goasdoué [Assistant Professor, Univ. Paris 11]
Nathalie Pernelle [Assistant Professor, Univ. Paris 11]
Chantal Reynaud [Professor, Univ. Paris 11, HdR]
Brigitte Safar [Assistant Professor, Univ. Paris 11]
Fatiha Saïs [Assistant Professor, Univ. Paris 11, from September]
Laurent Simon [Assistant Professor, Univ. Paris 11]
Véronique Ventos [Assistant Professor, Univ. Paris 11]

Technical Staff

Anca Ghitescu [till October]
Nobal Niraula [from December]
Mohamed Ouazara
Alin Tilea [from September]
Gabriel Vasile [till June]

PhD Student

Nada Abdallah [Allocataire MENRT, Paris 11]
Andrei Arion [Allocataire MENRT, Paris 11, till September]
Vincent Armant [Allocataire MENRT, Paris 11]
Michel Batteux [CIFRE with Sherpa Engineering, Paris 11, since November]
Pierre Bourhis [ENS Cachan, since September]
François-Elie Calvier [Grant BDI CNRS, Paris 11]
Alban Galland [X-telecom, Paris 11, from July]
Fayçal Hamdi [ANR Contract, from September]
Evgeny Kharlamov [Univ. Bolzano, 1 month]
Wael Khemiri [ANR Contract, from September]
Yingmin Li [European contract]
Gia Hien Nguyen [Grant MENRT, Grenoble 1, till November]
Bogdan Marinoiu [Grant BDI CNRS, Paris 11]
Yassine Mrabet [Digiteo contract, from September]
Nicoleta Preda [Allocataire MENRT, Paris 11, till July]
Cedric Pruski [PhD in cotutelle between Luxembourg U. and Paris 11]
Pierre Senellart [ENS Ulm, till January]
Mouhamadou Thiam [PhD in cotutelle between Gaston Berger U. and Paris 11]
Lina Ye [Allocataire MENRT, Paris 11]
Nadjet Zemirline [Contrat, Paris 11]
Spyros Zoupanos [CORDI]

Post-Doctoral Fellow

Alexander Lazovik [ERCIM Post Doc fellowship, till September]
Agnès Madalinski [Post Doc fellowship, till July]
Laura Brandán Briones [Post Doc fellowship]
Farid Nouioua [Post Doc fellowship, till August]
Gauvain Bourgne [Post Doc fellowship, from October]

Othman Nasri [Post Doc fellowship, from November]
Hassan Shraim [Post Doc fellowship, till February]
Haïfa Zargayouna [Post Doc fellowship, till August]

Visiting Scientist

Tarek Melliti [Assistant Professor, Univ. Evry Val d'Essonne, scientific advisor]
Neoklis Polyzotis [Assistant Professor, U.C. Santa Cruz, 1 month]
Victor Vianu [Professor, U.C. San Diego, 6 months, joint with Dahu team]
Yuhong Yan [Assistant Professor, U. Concordia, Montreal, 1 month]

Administrative Assistant

Marie Domingues [ITA]

2. Overall Objectives

2.1. Introduction

Gemo is a joint project with Laboratoire de Recherche en Informatique (UMR 8623 CNRS-University Paris-Sud), located in Orsay.

See <http://gemo.futurs.inria.fr/>

Information available online is more and more complex, distributed, heterogeneous, replicated, and changing. Web services, such as SOAP services, should also be viewed as information to be exploited. The goal of Gemo is to study fundamental problems that are raised by modern information and knowledge management systems, and propose novel solutions to solve these problems.

2.2. Highlights of the year

A lot of work has been devoted to the ANR Project WebContent.

Serge Abiteboul has been awarded the ERC Advanced Grant Webdam on Foundations of Web Data Management (starting December 2008).

Serge Abiteboul has been awarded the ACM PODS Alberto O. Mendelzon Test-of-Time Award 2008 for the paper "Complexity of Answering Queries Using Materialized Views" with Oliver M. Duschka.

Serge Abiteboul has been elected to the Académie des Sciences.

3. Scientific Foundations

3.1. Scientific Foundations

Keywords: *Databases, Web services, World Wide Web, XML, change control, complexity, data integration, distributed query, knowledge representation, logic, ontology, peer-to-peer (p2p), query language, query optimization, semantic integration, semi-structured data.*

A main theme of the team is the integration of information, seen as a general concept, including the discovery of meaningful information sources or services, the understanding of their content or goal, their integration and the monitoring of their evolution over time.

Gemo works on environments that are both powerful and flexible to simplify the development and deployment of applications providing fast access to meaningful data. In particular, content warehouses and mediators offering a wide access to multiple heterogeneous sources provide a good means of achieving these goals.

Gemo is a project born from the merging of INRIA-Rocquencourt project Verso, with members of the IASI group of LRI. It is located in Orsay-Saclay. A particularity of the group is to address data and knowledge management issues by combining techniques coming from artificial intelligence (such as classification) and databases (such as indexing).

The goal is to enable non-experts, such as scientists, to build *content sharing communities* in a true database fashion: declaratively. The proposed infrastructure is called a *data ring*.

4. Application Domains

4.1. Application Domains

Keywords: *Web, data warehousing, electronic commerce, enterprise portal, multimedia, search engine, telecommunications.*

Databases do not have specific application fields. As a matter of fact, most human activities lead today to some form of data management. In particular, all applications involving the processing of large amounts of data require the use of databases.

Technologies recently developed within the group focus on novel applications in the context of the Web, telecom, multimedia, enterprise portals, or information systems open to the Web. For instance, in the setting of the EDOS EC Project, we are developing some software for the P2P management the data and metadata of Mandriva Linux distribution.

5. Software

5.1. Software

Some recent software developed in Gemo:

ActiveXML: a language and system based on XML documents containing Web service calls. ActiveXML is now in Open Source within the ObjectWeb Forge.

AlignViz: a visualization tool for alignments between ontologies.

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

SHIRI-Annot: a tool to annotate semantically documents elements by exploiting both content and structure.

SHIRI-querying: a tool to approximate user's queries according to the SHIRI annotation design.

SomeWhere: a P2P infrastructure for semantic mediation.

SomeWhere+: a P2P infrastructure tolerant to inconsistency.

SpyWhere: a generator of mapping candidates for enriching peer ontologies.

KadoP: a peer-to-peer platform for warehousing of Web resources.

OptimAX: an algebraic cost-based optimizer for ActiveXML.

TARGET: a framework and a tool to search the Web by using adaptive ontologies.

TaxoMap: a prototype to automate semantic mappings between taxonomies.

XTAB2SML: an automatic ontology-based tool to enrich tables semantically.

ULoad: a tool for creating and storing XML materialized views, and using them to answer XQuery queries.

GUNSAT: a greedy local search algorithm for propositional unsatisfiability testing

LN2R: a logical and numerical tool for references reconciliation.

6. New Results

6.1. Theoretical foundations

Keywords: *Semi-structured data, automata, query languages, verification.*

Participants: Serge Abiteboul, Pierre Bourhis, Victor Vianu.

One of the reasons for the success of the relational data model was probably its clean theoretical foundations. Obtaining such a clean foundation for the semistructured data model and XML is still an on-going research task. Along this line, we have pursued our works on Active XML, a high-level specification language tailored to data-intensive, distributed, dynamic Web services. Active XML is based on XML documents with embedded function calls. The state of a document evolves depending on the result of internal function calls (local computations) or external ones (interactions with users or other services). Function calls return documents that may be active, so may activate new subtasks. The focus of the paper is on the verification of temporal properties of runs of Active XML systems, specified in a tree-pattern based temporal logic, Tree-LTL, that allows expressing a rich class of semantic properties of the application. The main results establish the boundary of decidability and the complexity of automatic verification of Tree-LTL properties [20].

6.2. Ontology-Based Information Retrieval

Keywords: *information retrieval, ontology evolution, query enrichment, resource annotation.*

Participants: Nathalie Pernelle, Cedric Pruski, Chantal Reynaud, Mouhamadou Thiam, Yassine Mrabet.

6.2.1. Query Enrichment based on Adaptive Ontologies

Web search applications still need to be more intelligent to adapt to the search domain targeted by users and to its evolution. Domain evolution is a complex problem which has rarely been integrated in Web search techniques [42]. We worked on the evolution process of ontologies in the setting of the TARGET Web search approach. We studied the adaptation of ontologies to changes through the definition of adaptation rules. Then we defined web queries enrichment rules based on evolving ontologies. In a previous work, we proposed a set of ontology-based Web query enrichment rules relying on various existing ontological relations. We now propose to adapt these rules in order to take into account evolution features and to focus on more relevant added concepts. A framework, named TARGET, usable for Web searching that adopts adaptive ontologies to model both domain concepts and user profiles has been implemented. In order to make experiments and validate our approach, we built an experimental case study which consists in the search of scientific papers published at the World Wide Web series of conference that took place over a decade [46]. This work is the core of a PhD which will be defended early next year.

6.2.2. Semantic Annotation

The SHIRI Digiteo labs project aims to build an integration system which allows the user access pertinent parts of documents related to a specific domain. This automatic and ontology-driven approach deals with heterogeneously structured documents. In the setting of this project, we have finalized the annotation model which is composed of metadata provided from the ontology and metadata that are defined specifically to take into account the less structured parts of the documents [37]. We are now validating the relevance of a set of annotation rules on a corpus made of HTML Call for Papers. We have also defined a first version of the SHIRI-querying component which exploits the annotation model. The aim of this component is to dynamically construct a set of approximated queries from a user domain query. We have defined two kinds of query approximation. The first one named *aggregative approximation* allows the querying of RDF annotations when parts of documents are not well structured. The second one named *neighboring approximation* allows reaching parts of the documents that are structurally linked even if a required semantic relation is not found. Preliminary experiments have been done on a set of publications' references.

6.3. Peer-to-Peer Inference Systems

Keywords: *RDF, distributed diagnosis, distributed reasoning, inconsistency, ontology alignment, propositional logic.*

Participants: Nada Abdallah, Vincent Armant, François-Elie Calvier, Philippe Chatalic, Philippe Dague, François Goasdoué, Gia Hien Nguyen, Chantal Reynaud, Marie-Christine Rousset, Laurent Simon, Véronique Ventos.

A P2P inference system (P2PIS) is made of autonomous agents called peers. Each peer models its application domain using a knowledge base (KB) and peers having similar interests can establish semantic correspondences between their KBs called mappings. Mappings play a central role since on the one hand they define how the KBs of some peers integrate and on the other hand they give rise to a semantic network, the decentralized KB of the P2PIS, in which it becomes possible to reason. However, reasoning in the distributed logical setting of a P2PIS is not simple. Indeed, the challenge is to design decentralized reasoning algorithms with the purpose to reduce an inference task to perform on the KB of a P2PIS to a decentralized calculus among the peers, while none of them has a comprehensive view of the global KB.

6.3.1. Consequence Finding

In the last years, we have investigated the basic AI task of consequence finding in (possibly inconsistent) propositional P2PISs. Consequence finding consists of deriving theorems of interest that are intentionally characterized within a logical theory. It proves useful in many composite AI tasks such as common sense reasoning, diagnosis, or knowledge compilation. Most of our results have been implemented in the SomeWhere platform, the scalability of which has been demonstrated on synthetic data (up to a thousand of peers). It is worth noticing that our results allow one for the very first time to perform a truly decentralized consequence finding calculus in a distributed theory of propositional logic, i.e., without having a global view of that theory.

A key point in the design of the above P2PISs is that mappings are *undirected*, i.e., any peer involved in a mapping can use it to propagate knowledge to the other peers participating in the mapping. Therefore, such systems model autonomous components that communicate through interfaces that are both input and output. We have recently proposed an alternative design of P2PISs in which mappings are *directed* [10]. A mapping is stated between two peers, but only one of them can use the mapping to propagate knowledge to the other. From a practical viewpoint, a mapping from a peer to another specifies some knowledge that the former peer has to observe and the knowledge it must notify to the latter peer if the observed knowledge holds. Such new P2PISs are of great interest in order to apply AI reasoning because they can model many real applications in which autonomous components communicate through interfaces that are either input or output, like distributed functions in Automotive Engineering, distributed control systems for industrial machinery or processes in Automation, etc. For those systems, we have proposed a new DEcentralized Consequence finding Algorithm for directed mappings (DECA_K) that computes clausal implicates of a clause submitted to a peer w.r.t. a P2PIS, including *all* the proper prime ones.

6.3.2. P2P conservative extension checking

We have pointed out that the notion of non conservative extension of a knowledge base (KB) is important to the distributed logical setting of propositional P2PIS. It is useful to a peer in order to detect/prevent that a P2PIS corrupts (part of) its knowledge or to learn more about its own application domain from the P2PIS [11], [12]. That notion is all the more important since it has connections with the privacy of a peer within a P2PIS and with the quality of service provided by a P2PIS. We have therefore studied the following tightly related problems from both the theoretical and decentralized algorithmic perspectives: (i) deciding whether a P2PIS is a conservative extension of a given peer and (ii) computing the witnesses to the corruption of a given peer's KB within a P2PIS so that we can forbid it.

6.3.3. Trust modeling

Because of the decentralized nature of P2PIS architectures, different peers may reflect different points of views on some data. The quality of semantic annotations may vary from one peer to another. User satisfaction can thus vary according to the answers returned to his queries. We proposed a Bayesian approach based on user feedbacks, in order to assess the trust that a user may have into answers depending on their semantic annotations [33], [32].

6.3.4. Distributed Diagnosis

The new research axis on consistency-based distributed diagnosis, initiated in 2007, pursued through a thesis. It is set up in the framework of propositional P2PISs. In this framework, each peer has only a local knowledge of its expected behavior (its local model unknown from other peers, e.g. for privacy reasons) and of its neighbors (its mappings), and one still wants to identify subsets of peers that, if we stop to assume they are correct, will restore consistency between the (unknown) global model (that would be the union of models of all the peers, never built in fact) and the observations. As an illustration, one may use this approach to diagnose faulty Web services. Well known algorithms exist in the centralized case, more recent ones exist in the case of a distributed model but centralized reasoning located into a supervisor, but the challenge here is to design an algorithm to achieve this diagnostic task in a completely distributed way, respecting privacy of the peers.

We developed such an algorithm, based on distributed computation of prime implicants of the global theory along a dynamically built covering tree of the peers involved. In contrast to most of previous works in model-based diagnosis in the centralized case, this algorithm computes all minimal diagnoses without needs to get any conflict first [21], [23], [22]. To the best of our knowledge, this attempt to set up the classical logic-based diagnosis theory in a purely distributed framework is completely original. We investigate now improvements of the algorithm by using jointrees. Then scalability and dynamicity (w.r.t. the observations and the structure itself of the peers network) issues will be addressed.

6.3.5. Mapping distributed ontologies

The problem of discovering mappings between distributed ontologies is addressed in the setting of the MediaD project. Previously, we proposed an approach that takes advantage of SomeRDFS reasoning [29]. Based on algorithms described in the PhD thesis of Philippe Adjiman, we implemented the peer data management system (PDMS) SomeRDFS and then a first prototype of a mapping generator, named SpyWhere. Our methods to discover mappings using query answering have then been extended to take into account various peer strategies. Filtering criteria and filtering methods have also been introduced to focus on the most relevant mapping candidates. These techniques are described in [27] and [28]. All the extensions have been implemented and experiments have been conducted using SpyWhere.

6.4. Thematic Web Warehousing

Keywords: *Warehouse, data fusion, data integration, ontology alignment.*

Participants: Serge Abiteboul, Jamal Ait Alla, H el ene Gagliardi, Alban Galland, Fay al Hamdi, Nathalie Pernelle, Chantal Reynaud, Marie-Christine Rousset, Fatiha Sa s, Brigitte Safar, Ha fa Zargayouna.

6.4.1. Reference reconciliation

The reference reconciliation problem consists in deciding whether different identifiers refer to the same data (same person, same conference, ...). The logical and numerical approach named LN2R that we have developed has been detailed in [8]. This approach allows computing a set of reference pairs that (1) refers to the same data, (2) does not refer to the same data and (3) may refer to the same data. The issue of data fusion arises once reconciliations have been determined. The objective of the fusion is to obtain a unique representation of the real world entity. An approach which deals with the uncertainty in the values associated with the attributes thanks to a formalism based on fuzzy sets [36], [35] has been developed. This work began in the LIRMM laboratory of Montpellier II University in the setting of Fatiha Sa s post-doc. It is continuing in collaboration with INRA Montpellier. The aim now is to build an ontology-driven data fusion approach.

Besides, LN2R allows inferring the semantic equivalence of heterogeneous basic values. We are studying how this learnt dictionary can be automatically refined in order to improve the reconciliation results. We plan to validate the dictionary refinement on data provided by THALES.

6.4.2. Mapping between ontologies

Our work on TaxoMap, a tool to align taxonomies, has been carried on in the setting of the WebContent and Geonto projects. Previous experiments emphasized two main difficulties: the alignment of very large ontologies and the difficulty of the evaluation process when no reference mappings are provided. As a consequence we worked on the partitioning of large scale ontologies. We proposed two methods. The aim is to have minimum blocks to align with maximal number of concepts that TaxoMap is able to handle. The originality of our methods is that they are alignment oriented, that means that the block generation is influenced by the mapping procedure [47]. Furthermore, we developed a tool, AlignViz, able to visualize OWL ontologies and RDF/XML mappings obtained as a result from an alignment tool. The aim is to help an expert to browse among the two aligned ontologies, to access a particular relation or concept, to have a global or a detailed view on an ontology and to visualize, validate or delete generated mappings [45]. At the same time, we implemented a more efficient and modular version of TaxoMap (TaxoMap 2) which has been integrated in the alignment server developed at INRIA Rhône-Alpes. The alignment server aims at storing and sharing mappings. It is based on the alignment API providing operators to handle mappings. The alignment server is a Web service in the WebContent platform. All these developments allowed us to participate for the second time in the international alignment contest OAEI 2008 [30] which consists of applying matching systems to ontology pairs and evaluating their results. Furthermore we investigated automated alignment techniques of another kind supported by a pattern and a rule-based approach well suited to align generic and specific models nociteGemoReport507. This problem has been studied in the setting of Adaptive Hypermedia (AH) in order to help AH creators' models to be reused in a platform only made up of generic components [39].

6.4.3. Viewpoints corroboration

We consider a set of views stating possibly conflicting facts. Negative facts in the views may come, e.g., from functional dependencies in the underlying database schema. We want to predict the truth values of the facts. Beyond simple methods such as voting (typically rather accurate), we explore techniques based on "corroboration", i.e., taking into account trust in the views. We introduce three fixpoint algorithms corresponding to different levels of complexity of an underlying probabilistic model. They all estimate both truth values of facts and trust in the views. We present experimental studies on synthetic and real-world data. This analysis illustrates how and in which context these methods improve corroboration results over voting methods. We believe that corroboration can serve in a wide range of applications such as source selection in the semantic Web, data quality assessment or semantic annotation cleaning in social networks. This work sets the bases for a wide range of techniques for solving these more complex problems.

6.5. XML query optimization

Keywords: *Query Optimization, Semi-structured Data.*

Participants: Serge Abiteboul, Andrei Arion, Ioana Manolescu, Spyros Zoupanos.

6.5.1. Static analysis for XML query optimization

XML query optimization can benefit from knowledge about the structure of the document(s) on which the queries are asked. Such knowledge can be expressed by an a priori schema, such as a DTD or XML Schema. An alternative is to extract from the data an a posteriori schema, encapsulating the characteristics of the actual data set. We considered the usage of Dataguides (introduced in 1997 as a basis for indexing in early semistructured data stores) as support for query optimization in the context of modern XML query processing systems. We demonstrate their usefulness via several new query analysis and simplification techniques, as well as their feasibility by implementing an open-source library, XSum, to be used in conjunction with an XML store [3].

6.5.2. Algebraic optimization for ActiveXML

The ActiveXML language (AXML in short) allows describing complex distributed data manipulation tasks. Each such task could be executed in many ways producing the same results but with very different performance. We have completed the implementation of an algebraic optimizer, named OptimAX, devised to function in conjunction with an ActiveXML peer. OptimAX can be seen as an AXML rewriting engine, transforming one document into another whose materialization produces the same results but potentially with better performance. OptimAX has lead to several publications, focusing on its design and performance [18], [17], respectively, on its applicability in the WebContent RNTL project which is one of its main applications [13].

6.5.3. Performance evaluation methodology

Performance evaluation is a natural component in many data-oriented works such as those carried on in Gemo. More generally, the last years have seen the emergence of intense discussion within the international scientific data management community as to what is the standard of proof for performance claims made in research papers, and how best to foster rapid prototyping and improvement of research platforms by sharing or at least facilitating the process of experimenting with somebody else's code. Gemo has been at the forefront of this effort, since four team members (I. Manolescu, A. Arion, P. Senellart and S. Zoupanos) have been involved in the SIGMOD 2008 repeatability assessment effort, whereas code provided for 80 submissions to the top conference ACM SIGMOD has been tested to check that similar results to those reported in the paper can be obtained. Insights obtained from this effort have been made available to the community via a joint paper [4] and a tutorial [31].

6.6. XML Warehousing in P2P

Keywords: *P2P, Warehouse, XML.*

Participants: Serge Abiteboul, Ioana Manolescu, Nicoleta Preda, Melanie Weis.

6.6.1. XML indexing: KadoP

We have worked on the optimization of KADOP, a peer-to-peer platform for building and managing warehouses of Web resources. KADOP relies on a Distributed Hash Table implementation (namely, FreePastry) to keep the network of peers connected, and to build a shared global resource index, and on the ActiveXML platform to store, query, and maintain the index. Furthermore, KADOP is able to process simple queries carrying over resources distributed in the whole network. A main goal is to be able to index not only extensional XML data but also intensional one and in particular Web services.

A recent development of the system includes two techniques meant to handle efficiently long posting lists exchanged during query processing. The first technique relies on a distributed search structure that parallelizes the transfer of long posting lists, while the second enables to reduce the transferred lists at the expense of some precision. These techniques are described in [16].

6.6.2. XML views on DHTs: VIP2P

The VIP2P project (standing for views in p2p) has started in 2008, building on the knowledge we acquired while developing KadoP. VIP2P is a peer-to-peer platform whereas peers independently publish (share) XML documents that others may query. Moreover, each peer may choose to materialize a given set of views, described by XML tree patterns. A query posed in the network of peers is thus first rewritten with the help of the views available in the network, then executed by exploiting these views.

6.7. Monitoring and Web services

Keywords: *Web services, diagnosability, formal models, model checking, model-based diagnosis, monitoring, self-healing.*

Participants: Serge Abiteboul, Gauvain Bourgne, Laura Brandán Briones, Pierre Bourhis, Philippe Dague, Alexander Lazovik, Yingmin Li, Agnès Madalinski, Bogdan Marinoiu, Tarek Melliti, Farid Nouioua, Lina Ye.

6.7.1. Error diagnosis and self-healing

In the framework of the EC project WS-DIAMOND devoted to self-healibility of Web services, our contribution was about modeling and decentralized (with a coordinator) diagnosis of conversationally complex Web services. We had developed in 2007 a BPEL2PN model generator, which takes as input a BPEL service package and produces as output a Petri net model of its control and data workflows, enriched by data dependencies. This model can then be exploited for several purposes (simulation, test, conformity to specifications, diagnosis, etc.). We then developed a diagnostic model compiler, which generates from this Petri net a diagnostic knowledge base as a set of causal rules expressed as logical Horn clauses [38]. We finally implemented an observation and data log platform, which stores in databases input, output and error messages of basic Web services, and a decentralized diagnostic algorithm, which exploits both the local diagnostic rules and the observations stored during an execution process to compute minimal diagnoses. The diagnostic architecture is made up of local diagnosers (provided to each BPEL service) that perform local diagnosis and communicate with a coordinator that merges these local diagnoses to build global ones, with the only knowledge of shared variables between individual Web services. Both local diagnosers and coordinator are implemented as Web services.

Many researchers have worked on Web services modeling and, at a preliminary level, decentralized diagnosis architecture with a supervisor. What is new in our work is the enrichment of a model by causal information necessary and sufficient to achieve diagnosis, and the automatic generation of both this model and the derived diagnostic rules from a BPEL code. We conducted an analogue research with labelled transition systems, instead of Petri nets, models [9]. Work in progress concerns modeling directly semantic faults and their propagation inside the Petri net model itself. For this, faults inside places and transitions are introduced and their propagation is represented by using colored tokens (e.g., colors represent normal, faulty and unknown status of data values) and a color propagation function. Such an ECPN (Enriched Colored Petri Net) model will allow effective matrix calculation for diagnosis and investigation of decentralized diagnosis for cooperating ECPNs.

6.7.2. Diagnosability

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 in a bounded time after its occurrence. We began to study this topic one year ago and we lead several coordinated actions in parallel in the framework of discrete-event systems. In the DIAFORE project, we are formalizing diagnosability analysis, currently expressed in terms of automata in the literature, in terms of Input-Output Symbolic Transition Systems and studying with CEA the use of its Agatha tool of symbolic verification of formal specifications for checking diagnosability. Other work concerned the link between observability and diagnosability, i.e. defining optimal observability (e.g. sensors placement, messages filtering) for a system to be diagnosable [25], [26], and the link between diagnosability and additional probabilistic knowledge (events, in particular faults occurrences, probabilities) exploited through Markov chains to quantify the degree of non diagnosability instead getting just a binary answer [34].

All this work is conducted first in a centralized framework, but our objective is to tackle then distributed diagnosability checking for systems where the global model is not known, and to apply it in particular to conversational Web services as studied in the WS-DIAMOND project. It is why we are also investigating systems modeled by Petri nets in order to use the computation of minimal prefixes of Petri nets unfoldings for verifying diagnosability, and then to study distributed computation of these prefixes. In the same way, distributed diagnosability of a given pattern (automaton) is studied for distributed systems modeled as communicating automata.

As there exists obvious relationships between diagnosability analysis and verification and model checking and also, we guess, with testing methods, we began a collaboration with colleagues of the ForTesSE group of LRI, in the framework of their WebMov project dedicated to Web services modeling and verification, in order to investigate in this common application field the complementarity and combination of these methods.

6.7.3. P2P Monitoring

We have worked on the conception and implementation of tools for monitoring Peer to Peer Systems.

A system named P2PMonitor has been developed for this purpose. It is a P2P system itself, with peers exchanging messages by Web service calls.

We focused on a problem closely related to monitoring: view maintenance over active documents. Indeed, the monitoring problem can be seen as aggregating streams into an active document and incrementally evaluating a tree-pattern query over this active document.

We have developed algorithmic datalog-based foundations for such an incremental query processing [14]. We have also addressed interesting issues that appeared in this context, like query satisfiability over active documents and stream relevance for considered queries.

A paper presenting a demonstration scenario for the monitoring system that integrates components for view maintenance over active documents has been published in [19].

7. Contracts and Grants with Industry

7.1. Industrial contracts

Gemo has had technical meetings in 2008 with many industrial partners, in particular France Telecom R&D, Exalead and Mandrakesoft, as well as national organizations, in particular, Institut National de Recherche en Agronomie.

7.2. RNTL Project WebContent

The WebContent project (<http://www.webcontent.fr>) completed its second year in July 2008. The goal of WebContent is to build a flexible and generic platform for content management and to integrate Semantic Web technologies in order to show their effectiveness on real applications with strong economic or societal stakes. Gemo activity in WebContent this year has been manifold. In the architecture group (Lot 0), we have secured an agreement on the usage of Web services as means of interconnecting the project components. In the peer-to-peer group (Lot 5), we have completed a total rehaul of the AXML peer developed by Gemo. E. Tarozza has proposed and implemented a new peer, more robust, and more modular; for instance, the XML storage services provided by the AXML peer have been isolated as a separated component and delegated to the eXist system. M. Ouazara is currently moving this storage component to MonetDB, the system which was retained by the WebContent consortium. At the same time, we have worked on extending the KadoP system to support XML namespaces. The development of OptimAX, the algebraic optimizer for AXML, also contributes to WebContent. Finally, we are currently extending the SomeRDFS prototype with the ability to translate from SPARQL to XQuery, as required by the WebContent integrated platform.

In the semantic enrichment of ontologies group (Lot3) we developed a more efficient and modular version (TaxoMap2). We worked on integrating this version in the alignment server which is now a web service in the WebContent platform. Furthermore, we worked on the partitioning of large scale ontologies and on the visualization of OWL ontologies and RDF/XML mappings in order to help to validate generated mappings. Both works have been carried out in response to difficulties highlighted by previous experiments.

7.3. WS-DIAMOND EC project

WS-DIAMOND (“Web Services - DIAGnosability, MONitoring and DiAGnosis”) was a FP6 European project (FET Open Strep) which started on Sept. 1, 2005 and terminated on June 31, 2008. EU funding for University Paris-Sud was 188 kEuros. The project was coordinated by the University of Turin, and involved the Polytechnic University of Milan, the Vrije University of Amsterdam, the University of Vienna, the University of Klagenfurt, and from France the LAAS-CNRS, the University of Rennes 1, and the University of Paris-Sud.

Participants from Gemo in 2008 were Philippe Dague (site leader for U. Paris-Sud), Tarek Melliti (assistant professor at U. of Evry from Sept. 2006), Yingmin Li (Ph.D. student from Oct. 2006), Lina Ye (Ph.D. student from end of Sept. 2007), and Laura Brandán Briones (post-doc from May 2007). The final review held in mid September 2008 with very good appreciation from the three experts.

7.4. ANR-PREDIT DIAFORE project

DIAFORE (“DIAGnostic de FONctions REparties”) is an ANR PREDIT (“Programme de recherche et d’innovation dans les transports terrestres”) project which started on Jan. 1, 2006 and will end on Aug. 31, 2009. Funding for University Paris-Sud is 481 kEuros. The project is coordinated by CEA LIST and involves Renault Trucks (Volvo group), Serma Ingénierie, UTC and University Paris-Sud. Two objectives are pursued in parallel by Gemo inside this project: embedded diagnosis of distributed functions inside a vehicle (with the LFSE group of CEA) and diagnosability analysis (with the LISE group of CEA). Participants from Gemo in 2008 were Philippe Dague (site leader for U. Paris-Sud), Hassan Shraim (post-doc from July 2007, till Feb. 2008) on diagnosis, and Farid Nouioua (post-doc from Sept. 2007, till Aug. 2008), Agnès Madalinski (post-doc from Jan. to June 2008), Laura Brandán Briones (post-doc, from July 2008) and Gauvain Bourgne (post-doc from Oct. 2008) on diagnosability analysis of discrete-event systems.

8. Other Grants and Activities

8.1. National Actions

In France, close links exist with groups at Orsay (databases, V. Benzaken and N. Bidoit; bio-informatics, C. Froidevaux; machine learning, M. Sebag; information visualization, J.-D. Fekete), with the Cedric Group at CNAM-Paris; some INRIA groups (Atlas, P. Valduriez and DistribCom, A. Benveniste, at INRIA-Bretagne; Exmo, J. Euzenat, at INRIA Rhone-Alpes; Mostrare at INRIA Futurs Lille); the BIA group at INRA (P. Buche, C. Dervin), the GRIMM of the University of Toulouse Le Mirail (O. Haemmerlé), the LIRIS of the University of Lyon 1 (M. Hacid), the LIRMM of the University of Montpellier (M. Chein, M-L. Mugnier), the LI of the University of Tours (G. Venturini), and the UMPA at École normale supérieure de Lyon (Y. Ollivier).

8.1.1. DocFlow

DocFlow is a research project supported by the ANR Masses de données (2007-2009) with the Distribcom team at INRIA-Rennes (Albert Benveniste) and the Méthodes Formelles group at Labri-Bordeaux (Anca Muscholl). The topic is the analysis, monitoring, and optimization of Web documents and services. It builds on Active XML, a formalism for data exchange across peers developed by Gemo. The project aims at achieving a convergence of data and workflow management over the Web through the concept of active peer-to-peer documents.

8.1.2. ACI Normes Pratiques et Régulations des Politiques Publiques

This ACI started in 2005 and is projected to last three years. This ACI is a collaboration between Benjamin Nguyen (University of Versailles), and François-Xavier Dudouet (CNRS, Laboratoire IRISES). The project has completed this year, but the work carried on has been merged (and continues through) the WebStand project (see below).

8.1.3. ANR JCJC WebStand

The objective of this ANR, that started in 2006, is to analyze the problems surrounding the use of semi-structured databases in social sciences. This ANR regroups both computer science and sociology laboratories. The contract has been extended until July 2009. A platform has been developed, based on which our social science partners have been able to conduct some studies of the W3C and telecommunication standardization activities.

8.1.4. EDIFlow Digiteo project

This project has been funded starting from December 2008. It is a collaboration between data management and information visualisation researchers (with V. Benzaken and J.-D. Fekete). The purpose is to study models and build a corresponding platform for efficient data-intensive workflow systems; the main class of applications envisioned concerns scientific data management systems.

8.1.5. SHIRI Digiteo project

SHIRI is a research project funded by the Ile de France region as a Digiteo project which started on Oct. 1st 2007 and will last until Sept. 30th, 2011. It involves two partners of Digiteo, Supelec and the University of Paris-Sud. The aim of SHIRI is to design an annotation system to improve the relevance of the search on the Web when resources contain both semi-structured and textual data.

8.1.6. GEONTO project

The objective of this ANR MDCO project, that started in 2008, is to make data in the geographic domain inter-operate. We focus on two main goals. 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 just mentioned before. It is projected to last 3 years. 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 LIUPPA - 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. European Commission Financed Actions

In Europe, close links exist with University of Dortmund (T. Schwentick), University of Madrid (A. Gomez-Perez), University of Manchester (I. Horrocks), University of Rome (M. Lenzerini).

8.2.1. Webdam ERC Grant

The Webdam grant (S. Abiteboul) started in December 2008. The goal 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.

8.3. Bilateral International Relations

8.3.1. Cooperation within Europe

Procope

Gemo has a PHC-Procope project with the database group of Thomas Schwentick at Dortmund University, Germany. The project will end in 2008. Its goal is to work on verification and queries in the presence of data values. It produced already several joint papers between the two groups.

TARGET

Gemo started a cooperation with the Luxembourg University in November 2005 which leads to a PhD in co-tutelle with Paris-Sud university. The PhD project is TARGET for opTimal Adaptive infoRmation manaGemEnT over the web. It aims at improving web information retrieval by integrating web data evolution, users knowledge evolution and search domain evolution. The PhD student is Cedric Pruski.

University of Oxford

Gemo continued its collaboration with Georg Gottlob from University of Oxford. S. Abiteboul was Visiting Fellow at Sainte-Anne's College in winter 2008.

8.3.2. Cooperation with Senegal

Gemo started a cooperation with the Gaston Berger University in december 2006 which leads to a PhD in co-tutelle with Paris-Sud university. The subject of the thesis is the integration of semi-structured data for information retrieval. The PhD student is Mouhamadou Thiam.

8.3.3. Cooperation with the Middle-East

Close links exist with University of Tel-Aviv (T. Milo).

8.3.4. Cooperation with North America

Close links also exist with UC Santa Cruz (N. Polyzotis), U. of Rutgers (A. Borgida), Google Research (O. Benjelloun),

8.3.5. French-US team: GemSaD

Since 2003, Gemo and the data management group at the University of California at San Diego (V. Vianu, A. Deutch, Y. Papakonstantinou) form an associated team funded by INRIA International. This association is completed in 2008. Victor Vianu spent 6 months in Gemo and Dahu. The home page of GemSaD can be found at <http://www-rocq.inria.fr/~segoufin/GEMSAD/>. GemSad is also partially supported by the National Science Foundation.

8.4. Visiting Professors and Students

This year the following professors visited Verso:

- Tova Milo, professor at the University of Tel-Aviv (in February)
- Neoklis Polyzotis, professor at the University of Southern California (in September)
- Victor Vianu, professor, UC San Diego (July to December); the visit was joint with the Dahu INRIA group in Cachan (Luc Segoufin).
- Yuhong Yan, associate professor, U. Concordia of Montreal (in December)

An engineer student came for internship in the group: Jamal Ait Alla [Mohammadia Graduate School, Morocco, 3 months].

9. Dissemination

9.1. Thesis

The following PhD thesis were defended in 2008:

- Gia-Hien Nguyen, Reliability of answers in logical peer-to-peer networks.
- Nicoleta Preda, P2P indexing.

9.2. Participation in Conferences

S. Abiteboul is technical program chair of the Very Large Database Conference, Lyon, 2009.

I. Manolescu has been the ACM SIGMOD 2008 Experiment Repeatability chair, the demonstration chair of the EDBT 2008 conference, co-chair of the database track of the ACM CIKM 2008 conference, and general chair of the PLAN-X 2008 workshop, in conjunction with the ACM POPL conference.

Members of the project have participated in program committees:

S. Abiteboul

- ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS), Vancouver, Canada (2008)
- World Wide Web Conference, WWW2008, Beijing

P. Chatalic

- Journées Francophones de Programmation par Contraintes (JFPC 2008)

Ph. Dague

- 18th European Conference on Artificial Intelligence (ECAI, 2008), area chair
- 19th International Workshop on Principles of Diagnosis (DX 2008)
- 22th International Workshop on Qualitative Reasoning (QR 2008)
- 16ème Congrès francophone Reconnaissance des Formes et Intelligence Artificielle (RFIA 2008)
- Journées Nationales de l'Intelligence Artificielle Fondamentale (IAF 2008)

I. Manolescu

- Very Large Databases Conference (VLDB) 2008
- International Conference on Data Engineering (ICDE) 2008
- Journées en Bases de Données Avancées (BDA) 2008

F. Goasdoué

- 16èmes congrès francophone Reconnaissance des Formes et Intelligence Artificielle (RFIA08)

C. Reynaud

- 16th International Conference on Knowledge Engineering and Knowledge Management - Knowledge Patterns (EKAW 2008)
- International Conference on Web and Information Technologies (ICWIT 08)
- 16èmes congrès francophone Reconnaissance des Formes et Intelligence Artificielle, member of the editorial board (RFIA08)
- Conférence Extraction et Gestion des Connaissances (EGC08)
- 18èmes Journées Francophones d'Ingenierie des connaissances (IC08)
- 2èmes Journées Francophones sur les ontologies (JFO2008)
- Atelier Modélisation des connaissances (EGC08)
- Atelier Mesures de similarité sémantique (EGC08)

B. Safar

- 2èmes Journées Francophones sur les ontologies (JFO2008)

9.3. Invited Presentations

Seminary at Stanford I. Manolescu was invited at the Stanford University database group seminary in january 2008, with a talk on “XML Access Modules: Towards Physical Independence in XML Databases”.

9.4. Scientific Animations

Editors

S. Abiteboul is a member of the steering committee of Proceedings of the VLDB Endowment (PVLDB) Journal, a journal that started in 2008.

F. Goasdoué

- Guest editor of a special issue of *Technique et Science Informatiques (TSI)* on the Semantic Web, Hermès-Lavoisier.
- Member of the reading committee of the book *Semantic Web Methodologies for E-Business Applications: Ontologies, Processes and Management Practices*, Idea Group Publishing (scheduled for publication in 2008).

I. Manolescu

- Guest editor of a special issue of the Elsevier *Journal of Information Systems on Performance Evaluation in Database Systems*.

C. Reynaud

- *Journal Electronique d'IA de l'AFIA (JEDAI)*.
- *Revue Information - Interaction - Intelligence (RI3)*.

M-C. Rousset

- *Interstices* (revue electronique de vulgarisation sur la recherche en informatique): <http://interstices.info/>
- *AI Communications (AICOM)*
- *Electronic Transactions on Artificial Intelligence (ETAI)* (for the areas: Concept-based Knowledge Representation and Semantic Web).
- *Revue Information - Interaction - Intelligence (I3)*

L. Simon

- Member of the Editorial Board of *JSAT* (the *Journal on Satisfiability, Boolean Modeling and Computation*)
- Guest Editor of a Special Issue of *JSAT* on SAT 2006 Competitions and Evaluations.

9.4.1. Presentations to a larger public

Serge Abiteboul:

- Article: *La gestion de données pair-à-pair*, La Recherche, 2008
- Article: *Le Web malin de demain*, in *La science c'est fun!*, Okapi et Bayard Jeunesse, 2008
- Presentation: *Le Web Nouvelle génération*, Conference at CNAM organized by La Recherche and CNAM., 2008
- Debate: *Les réseaux sociaux*, in *Les Entretiens du Nouveau Monde Industriel*, Centre Georges Pompidou, 2008.
- Article: *Le Web de demain*, in *Journal du Net* 2008
- Interview podcast: *Le Web de demain*, in *Vnunet*, 2008

Ioana Manolescu:

- Article: *Et demain...*, 01 Informatique, mai 2008

10. Bibliography

Year Publications

Doctoral Dissertations and Habilitation Theses

- [1] N. PREDA. *Efficient Web resource management in structured peer-to-peer networks*, Ph. D. Thesis, Université Paris-Sud 11, may 2008.

Articles in International Peer-Reviewed Journal

- [2] S. ABITEBOUL, O. BENJELLOUN, T. MILO. *The Active XML project: an overview*, in "VLDB Journal", 2008.
- [3] A. ARION, A. BONIFATI, I. MANOLESCU, A. PUGLIESE. *Path Summaries and Path Partitioning in Modern XML Databases*, in "World Wide Web", vol. 11, n^o 1, 2008, p. 117-151.
- [4] I. MANOLESCU, L. AFANASIEV, A. ARION, J.-P. DITTRICH, S. MANEGOLD, N. POLYZOTIS, K. SCHNAIT-TER, P. SENELLART, S. ZOUPANOS, D. SHASHA. *The repeatability experiment of SIGMOD 2008*, in "SIGMOD Record", vol. 37, n^o 1, 2008, p. 39-45.
- [5] I. MANOLESCU, P. BONNET. *Foreword*, in "Inf. Syst.", vol. 33, n^o 2, 2008, p. 153-154.
- [6] P. MICHIELS, I. MANOLESCU, C. MIACHON. *Toward microbenchmarking XQuery*, in "Inf. Syst.", vol. 33, n^o 2, 2008, p. 182-202.
- [7] C. REYNAUD, B. SAFAR. *Construction automatique d'adaptateurs guidée par une ontologie pour l'intégration de sources et de données XML*, in "Revue TSI Technique et Science Informatiques", 2008.
- [8] F. SAIS, N. PERNELLE, M.-C. ROUSSET. *Combining a Logical and a Numerical Method for Data Reconciliation*, in "Journal of Data Semantics (JODS)", vol. 12, 2008.
- [9] Y. YAN, P. DAGUE, Y. PENCOLÉ, M.-O. CORDIER. *A Model-based Approach for Diagnosing Faults in Web Service Processes*, in "International Journal of Web Services Research JWRS", vol. 6(1), 2009.

International Peer-Reviewed Conference/Proceedings

- [10] N. ABDALLAH, F. GOASDOUÉ. *Calcul de conséquences dans un système d'inférence pair-à-pair propositionnel (revisité)*, in "Reconnaissance des Formes et Intelligence Artificielle (RFIA)", 2008.
- [11] N. ABDALLAH, F. GOASDOUÉ. *Calcul de conséquences pour le test d'extension conservative dans un système pair-à-pair*, in "Journées Francophones de Programmation par Contraintes", 2008.
- [12] N. ABDALLAH, F. GOASDOUÉ. *Systèmes pair-à-pair sémantiques et extension (non) conservative d'une base de connaissances*, in "Bases de données avancées (BDA)", 2008.
- [13] S. ABITEBOUL, T. ALLARD, P. CHATALIC, G. GARDARIN, A. GHITESCU, F. GOASDOUÉ, I. MANOLESCU, B. NGUYEN, M. OUAZARA, A. SOMANI, N. TRAVERS, G. VASILE, S. ZOUPANOS. *WebContent: Efficient P2P Warehousing of Web Data*, in "Very Large Databases Conference", 2008.
- [14] S. ABITEBOUL, P. BOURHIS, B. MARINOIU. *Efficient Maintenance Techniques for Views over Active Documents*, in "In Conference on Eextending Data Base Technology", ACM, 2009.
- [15] S. ABITEBOUL, O. GREENSPAN, T. MILO. *Modeling the Mashup Space*, in "International Workshop on Web Information and Data Management", 2008.
- [16] S. ABITEBOUL, I. MANOLESCU, N. POLYZOTIS, N. PREDA, C. SUN. *XML processing in DHT networks*, in "ICDE", 2008, p. 606-615.

-
- [17] S. ABITEBOUL, I. MANOLESCU, S. ZOUPANOS. *OptimAX: efficient support for data-intensive mash-ups*, in "ICDE", 2008, p. 1564-1567.
- [18] S. ABITEBOUL, I. MANOLESCU, S. ZOUPANOS. *OptimAX: Optimizing Distributed ActiveXML Applications*, in "International Conference on Web Engineering", july 2008.
- [19] S. ABITEBOUL, B. MARINOIU, P. BOURHIS. *Distributed Monitoring of Peer to Peer Systems (demo)*, in "International Conference on Data Engineering", 2008.
- [20] S. ABITEBOUL, L. SEGOUFIN, V. VIANU. *Static analysis of active XML systems*, in "PODS", 2008, p. 221-230.
- [21] V. ARMANT, P. DAGUE, L. SIMON. *Diagnostic Distribué à base de modèles sans calcul préalable des conflits*, in "4èmes Journées Francophones de Programmation par Contraintes", 2008, p. 41–48.
- [22] V. ARMANT, P. DAGUE, L. SIMON. *Distributed Consistency-Based Diagnosis*, in "Proc. of the 16th International Conference on Logic for Programming, Artificial Intelligence and Reasoning LPAR-08", 2008.
- [23] V. ARMANT, P. DAGUE, L. SIMON. *Distributed Consistency-Based Diagnosis without Conflicts*, in "Proc. of the 19th International Workshop on Principles of Diagnosis DX-08", 2008.
- [24] G. AUDEMARD, L. SIMON. *Experimenting with Small Changes in Conflict-Driven Clause Learning Algorithms*, in "Proc of International Conference on Principles and Practice of Constraint Programming", 2008.
- [25] L. BRANDÁN-BRIONES, A. LAZOVIK, P. DAGUE. *Optimal Observability for Diagnosability*, in "Proc. of the 19th International Workshop on Principles of Diagnosis DX-08", 2008.
- [26] L. BRANDÁN-BRIONES, A. LAZOVIK, P. DAGUE. *Optimizing the system observability level for diagnosability*, in "Proc. of the 3rd International Symposium on Leveraging Applications of Formal Methods, Verification and Validation ISoLA-08", 2008.
- [27] F.-E. CALVIER, C. REYNAUD. *Guiding Ontology Matching process with reasoning in a PDMS*, in "16th International Conference on Knowledge Engineering and Knowledge management - Knowledge pattern, Acitrezza, catania, Italy, EKAW 2008", 2008.
- [28] F.-E. CALVIER, C. REYNAUD. *Ontology Matching Supported by Query Answering in a P2P System*, in "The 7th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE 2008)", Springer Verlag, 2008.
- [29] F.-E. CALVIER, C. REYNAUD. *Une aide à la découverte de mappings dans SomeRDFS*, in "8èmes Journées Francophones Extraction et Gestion des Connaissances (EGC2008)", 2008.
- [30] F. HAMDY, H. ZARGAYOUNA, B. SAFAR, C. REYNAUD. *TaxoMap in the OAEI 2008 alignment contest*, in "Ontology Alignment Evaluation Initiative (OAEI) 2008 Campaign - Workshop on Ontology Matching", 2008.
- [31] I. MANOLESCU, S. MANEGOLD. *Performance Evaluation in Database Research: Principles and Experience (tutorial)*, in "International Conference on Data Engineering", 2008.

- [32] G.-H. NGUYEN, P. CHATALIC, M.-C. ROUSSET. *A probabilistic trust model for semantic peer to peer systems*, in "International Workshop on Data Management in Peer-to-Peer Systems (DaMaP'2008)", A. DOUCET, S. GANÇARSKI, E. PACITTI (editors), ACM International Conference Proceeding Series, ACM, march 2008, p. 59-65.
- [33] G.-H. NGUYEN, P. CHATALIC, M.-C. ROUSSET. *A probabilistic trust model for semantic peer-to-peer systems*, in "Proc. of the European Conference on Artificial Intelligence (ECAI 2008)", July 2008, p. 881-882.
- [34] F. NOUIOUA, P. DAGUE. *A Probabilistic Analysis of Diagnosability in Discrete Event Systems*, in "Proc. of the 18th European Conference on Artificial Intelligence (ECAI-08)", July 2008.
- [35] F. SAIS, R. THOMOPOULOS. *Reference Fusion and Flexible Querying*, in "The 7th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE 2008)", Springer Verlag, 2008.
- [36] F. SAIS, R. THOMOPOULOS. *Une méthode flexible de fusion de références*, in "4èmes Journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA2008)", june 2008.
- [37] M. THIAM, N. BENNACER, N. PERNELLE. *Contextual and Metadata based Approach for the Semantic Annotation of Heterogeneous Documents*, in "workshop SEMMA of ESWC", 2008.
- [38] L. YE, P. DAGUE. *Decentralized Diagnosis for BPEL Web Services*, in "Proc. of the 4th International Conference on Web Information Systems and Technologies (WEBIST'08)", May 2008, p. 283-287.
- [39] N. ZEMIRLINE, Y. BOURDA, C. REYNAUD, F. POPINEAU. *Assisting in Reuse of Adaptive Hypermedia Creator's models*, in "5th Int. conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH2008)", 2008.
- [40] N. ZEMIRLINE, C. REYNAUD, Y. BOURDA, F. POPINEAU. *A Pattern and Rule-based Approach for reusing Hypermedia Creator's Models*, in "16th International Conference on Knowledge Engineering and Knowledge management - Knowledge pattern, Acitrezza, catania, Italy, EKAW 2008", 2008.

Scientific Books (or Scientific Book chapters)

- [41] G. DEQUEN, L. SIMON. *Algorithmes de Recherche Systématique*, in "SAT : Progrès et Défis", Hermes, 2008.
- [42] N. GUEIFI, C. PRUSKI, C. REYNAUD. *Towards the Adaptive Web using Metadata Evolution*, in "handbook of research on Web Information Systems Quality", 2008.
- [43] A. LÉGER, J. HEINECKE, L. J. NIXON, P. SHVAIKO, J. CHARLET, P. HOBSON, F. GOASDOUÉ. *Semantic Web take-off in an Industry Perspective*, in "Semantic Web for Business: Cases and Applications", Idea Group Inc., 2008.
- [44] P. SENELLART, V. D. BLONDEL. *Automatic discovery of similar words*, in "Survey of Text Mining: Clustering, Classification and Retrieval", M. W. BERRY, M. CASTELLANOS (editors), Second Edition, Springer-Verlag, January 2008.

Research Reports

- [45] J. A. ALLA. *Outil de visualisation d'ontologies et de validation d'alignements*, Technical report, Computer Science Engineer report, Mohammadia School, Rabat, 2008.
- [46] N. GUEFI, C. PRUSKI, C. REYNAUD. *Experimental Assessment of the TARGET Adaptive Ontology-based Web Search Framework*, Technical report, LRI - University Paris-Sud, Orsay, 2008.
- [47] F. HAMDI. *Partitionnement d'ontologies pour l'alignement*, Technical report, Computer Science Master report, University Paris-Sud, Orsay, 2008.