INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Team Dahu

Verification in Database

Saclay - Île-de-France

THEME SYM

Table of contents

1. Team **1**
2. Overall Objectives **1**
3. Scientific Foundations **1**
4. Application Domains **2**
5. New Results **2**
5.1. Logics for data words and data trees 2
5.2. Extended automata models 3
5.3. Automata theory 3
6. Other Grants and Activities **4**
6.1. National collaborations 4
6.2. International collaborations 5
6.2.1. Cooperation within Europe 5
6.2.2. Cooperation with Tunisia 5
6.2.3. Cooperation with North America 5
6.3. Visiting Professors and Students 5
7. Dissemination **5**
7.1. Participation in conferences organisation 5
7.2. Participation to symposia, seminars, invitations 6
7.3. Scientific Animations 6
7.4. Teaching 6
8. Bibliography **6**

Dahu is a common project with LSV and ENS de Cachan. The team was created on January the 1st, 2008.

1. Team

Research Scientist

Stéphane Demri [Research Director (DR) CNRS, HdR]
Florent Jacquemard [Research assistant (CR), Inria]
Luc Segoufin [Team leader, Research Director (DR), Inria, HdR]

Faculty Member

Cristina Sirangelo [ENS Cachan, since September]

PhD Student

Claire David [Until November 1st]
Diego Figueira [Cordi]
Thomas Place [Allocation couplée]

Visiting Scientist

Victor Vianu [U.C. San Diego, USA, from September 1st till December 31]

Administrative Assistant

Marie Dominguès [Secretary (SAR) Inria]

Other

Serge Abiteboul [Scientific Advisor, HdR]

2. Overall Objectives

2.1. Overall Objectives

For more information see http://www.lsv.ens-cachan.fr/axes/DAHU/dahu.php.

The need to access and exchange data on the Web has led to database management systems (DBMS) that are increasingly distributed and autonomous. Data extraction and querying on the Web is harder than in classical DBMS, because such data is heterogeneous, redundant, inconsistent and subject to frequent modifications. DBMS thus need to be able to detect errors, to analyze them and to correct them. Moreover, increasingly complex Web applications and services rely on DBMS, and their reliability is crucial. This creates a need for tools for specifying DBMS in a high-level manner that is easier to understand, while also facilitating verification of critical properties.

The study of such specification and verification techniques is the main goal of Dahu.

3. Scientific Foundations

3.1. Scientific Foundations

Keywords: *Databases*, *XML*, *specification*, *verification*.

Dahu has strong connections with the Gemo project.

Dahu aims at developing mechanisms for high-level specifications of systems built around DBMS, that are easy to understand while also facilitating verification of critical properties. This requires developing tools that are suitable for reasoning about systems that manipulate data. Some tools for specifying and reasoning about data have already been studied independently by the database community and by the verification community, with various motivations. However, this work is still in its infancy and needs to be further developed and unified.

Most current proposals for reasoning about DBMS over XML documents are based on tree automata, taking advantage of the tree structure of XML documents. For this reason, the Dahu team is studying a variety of tree automata. This ranges from restrictions of "classical" tree automata in order to understand their expressive power, to extensions of tree automata in order to understand how to incorporate the manipulation of data.

Moreover, Dahu is also interested in logical frameworks that explicitly refer to data. Such logical frameworks can be used as high level declarative languages for specifying integrity constraints, format change during data exchange, web service functionalities and so on. Moreover, the same logical frameworks can be used to express the critical properties we wish to verify.

In order to achieve its goals, Dahu brings together world-class expertise in both databases and verification.

4. Application Domains

4.1. Application Domains

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

Databases are pervasive across many application fields. Indeed, most human activities today require some form of data management. In particular, all applications involving the processing of large amounts of data require the use of a database. Increasingly complex Web applications and services also rely on DBMS, and their correctness and robustness is crucial.

We believe that the automated solutions that Dahu aims to develop for verifying such systems will be useful in this context.

5. New Results

5.1. Logics for data words and data trees

Keywords: *Data words*, *data trees*, *logics*.

Participants: Claire David, Stéphane Demri, Diego Figueira, Luc Segoufin, Victor Vianu.

Our main objective is to provide tools for specifying and verifying systems with data. This means finding a suitable logical framework for specifying such systems. A logical framework is suitable if it is expressible enough for modeling classical database operations. Of course, for the logical framework to be useful, it must come with techniques and tools for reasoning about it, in particular it should be decidable.

Most of our new results in this direction concerns data words and data trees. Those are words and trees where each position contains a data value together with the classical label. Data words and data trees can model many systems with data with a focus on one variable flow. Data trees can also model XML data.

We have studied several logical frameworks that explicitly refer to data values. For instance, in the journal version [12], an extension of LTL with registers is introduced. This logic can express properties over data words and it has the ability to store a datum and to test it later against the current datum. Decidability and complexity results are shown. For instance, in the finitary case, the logic restricted to one register has a decidable (but non primitive recursive) satisfiability problem. Unfortunately many obvious extensions of this setting are shown to be undecidable (infinitary case, adding past-time operators, two registers). Nonemptiness problems for classes of register automata are also considered in [12].

In [20], we present complexity results related to the model-checking problem for LTL with registers over one-counter automata. This makes a difference with [12] that is mainly focused on satisfiability problems. We consider several classes of one-counter automata and several syntactic fragments. As in [12], the logic has the ability to store a counter value and to test it later against the current counter value. We show that model checking LTL with registers over deterministic one-counter automata is PSPACE-complete with infinite accepting runs. By contrast, we prove that model checking LTL with registers over nondeterministic one-counter automata is undecidable in the infinitary and finitary cases even if only one register is used. This makes a difference with the facts that several verification problems for one-counter automata are known to be decidable with relatively low complexity, and that finitary satisfiability for LTL with a unique register is decidable [12].

Another logical framework was studied in [19]. Here the properties are expressed by the mean of Boolean combination of data tree patterns. It allows to express properties over data trees. In general it is undecidable whether a Boolean combinations of data tree patterns is consistent. However decidability can be achieved by restricting negation appropriately.

A logical framework that extends modal logic is proposed in [15]. This model contains features for explicitly memorizing some information for later comparison. Its expressive power together with its decidability is studied in [15].

Finally we have also considered the specification of the evolution of data trees with time in [14]. This work builds on the data tree pattern paradigm of [19] and introduces an extension of LTL with data tree patterns called tree-LTL. Decidability of tree-LTL is obtained in a reasonably expressive scenario.

5.2. Extended automata models

Keywords: *Automata*, *constraints*, *memory*, *registers*.

Participants: Stéphane Demri, Florent Jacquemard.

In many cases verification is achieved by translating logical formula into automata and then checking for emptiness of automata. In the presence of data, the design of appropriate classes of automata with optimal complexities remains an on-going task.

For this reason we have studied several extensions of the classical model of finite automata with features that could be used for manipulating data. This is done either by using registers or memory explicitly in the model or by restricting the transitions of the automata with constraints that can involve data comparisons. Several models have been considered.

The first one extends alternating automata with a register. This model can accept languages over data words. The extension of LTL with one register can be translated into this model of automata. It is shown to be decidable in [12].

The second model consists of a tree automata computing with an auxiliary memory with a tree structure (instead of a stack e.g. for pushdown automata). In order to have good closure properties a visibility condition is enforced. Finally, different kinds of tests between memories and constraints about subtree isomorphisms are added. This model is shown decidable in [11].

Several classes of tree automata with constraint are studied in [13]. They combine tree automata testing subtree isomorphisms and automata computing modulo equational theories. We prove the decidability of several properties for these extended tree automata, using classical first order theorem proving techniques. Alternatively our results can be viewed as new decidable classes of first-order formula.

5.3. Automata theory

Keywords: *XML foundations*, *characterizations*, *logic*, *tree automata*.

Participants: Florent Jacquemard, Thomas Place, Luc Segoufin.

The links between models for XML and regular tree languages has been advocated in many places. Tree automata seem to be playing for semi-structured data and XML the role of the relational algebra for relational databases. As XML is central in our research we also study tree automata and regular tree languages. Unlike the previous section that consider extensions of regular tree automata for manipulating data, we study here restrictions of regular tree automata.

A first line of research concerns the expressive power of various subclasses of regular tree languages. It is usually admitted that a fragment is completely understood, in term of expressive power, when one has a *decidable characterization* of it. That is an algorithm that given a regular tree language, presented say as a tree automata, tests whether it belongs to the class being investigated or not. This question is an active research topic that turns out to be quite challenging. We have exhibited decidable characterizations for various classes of regular tree languages: the class of languages definable with Boolean combination of $\Sigma_1(<)$ formulas for unranked trees in [17], the class of languages definable with $\Delta_2(<)$ formulas for unranked trees in [16]. Decidable characterizations for classes of languages definable with the previous two logics, but also for the class of languages definable with $EF + F^{-1}$ formulas, has been obtained for ranked trees in [23].

We have also studied the expressive power of tree walking automata. We show in [24] that nesting such automata yields a very nice class of tree languages that is closed under all usual operations, has a logical characterization in terms of transitive closure logic, and is strictly weaker than regular tree languages.

A second line of research studies the transformation of tree automata languages under various kind of rewriting systems. We have considered for instance a new kind of rewrite systems for semi-structured data, which generalizes both notions of standard term rewriting and word rewriting, and investigated properties of preservation, under this notion of rewriting, of the languages of two classes of languages of unranked ordered terms. As a consequence, the problems of reachability and regular hedge model checking are shown decidable for restricted classes of rewrite systems [22]. We are pursuing now this line of research with the study of other classes of rewrite systems defining bottom-up tree transducers and application to Associative and Associative-Commutative term rewriting.

We have also considered the transformation of tree automata languages under term rewriting with the innermost strategy – which corresponds to the *call by value* computation in programming languages. We prove in particular [21] that the set of innermost-reachable terms from a tree automata language by a shallow rewrite system is not necessarily regular, but it can be recognized by a tree automaton testing isomorphisms between subtrees at brother positions. As a consequence we show the decidability of the reachability and of the problem of regularity of the set of terms reachable by innermost rewriting with the above class of rewrite systems. This result is in contrast with plain (not necessarily innermost) rewriting for which undecidability is proved.

Finally, we propose [18] a new method for automated implicit inductive theorem proving for equational specifications made of rewrite rules with conditions and constraints. Our procedure is based on tree automata with constraints, which are used in the induction proofs on one hand as an induction scheme for the generation of subgoals at induction steps, and on the other hand for the decision of redundancy criteria by reduction to an emptiness problem.

6. Other Grants and Activities

6.1. National collaborations

Dahu is currently participating in two ANR projects:

ENUM is a research project supported by the ANR blanche on algorithmic and complexity problems raised by enumerating solutions of a query. The goal is to provide formal methods to understand and compare the complexity of enumerations problems. The partners are University of Paris-7 (with Arnaud Durand), the project-team Mostrare at INRIA-Lille (with Joachim Niehren), the university of Caen (with Etienne Grandjean) and the university of Marseille (with Nadia Creignou). Dahu is involved in the ANR as part of the Paris-7 node.

Averiss is a research project supported by the ANR SETIN (ANR-06-SETI-001-02, 2007-2009) on the development of new techniques for automatic software verification taking into account complex features of modern programming languages, including infinite data domains and procedure calls. The partners are LIAFA, University of Paris-7 (with Ahmed Bouajjani), LABRI, Bordeaux (with Igor Walukiewicz) and LSV, ENS Cachan (with Philippe Schnoebelen). Dahu is involved in this project as part of the LSV node.

6.2. International collaborations

6.2.1. Cooperation within Europe

Dahu has strong connections with several universities within Europe. A joint proposal has been submitted to the FET-open call of FP7. This involves Thomas Schwentick (university of Dortmund, Germany), Mikolaj Bojańczyk (university of Warsaw, Poland) and Leonid Libkin (university of Edinburgh, Scotland) who have been collaborating with Dahu since the beginning.

6.2.2. Cooperation with Tunisia

Dahu is coordinator (on the French side) of a project INRIA-DGRSRT (Tunisian universities) on "automated verification of the conformance of firewall configurations to access-control policies", since January 2008. The other partners of the project are the CASSIS team at INRIA Nancy-Grand-Est and the security team at Sup'Com Tunis.

6.2.3. Cooperation with North America

Close links also exist with UC San Diego and the database group of Victor Vianu.

6.3. Visiting Professors and Students

This year the following researchers made a long visit to Dahu:

- Ranko Lazić, assistant professor, University of Warwick (1 month in 2008).
- Sasha Rubin, post-doc, University of Auckland, New-zeland (April and Mai 2008).
- Victor Vianu, professor, UC San Diego, USA (September to December 2008)
- Hitoshi Ohsaki, senior research scientist, AIST, Japan (2 weeks in November 2008)
- Mohammed Anis Benelbahri, Ph.D., Sup'Com Tunis, Tunisia (November 2008)

7. Dissemination

7.1. Participation in conferences organisation

Stéphane Demri has been program co-chair of the "15th International Symposium on Temporal Representation and Reasoning" (TIME 2008), Montreal, June 2008 (with Chr. Jensen).

Several members of the project have participated in program committees:

- S. Demri: International conference "Advances of Modal Logic" (AIML'08), September 2008.
- F. Jacquemard: International Conference on RIsks and Security of Internet and Systems (CRISIS), October 2008.
- C. Sirangelo: 12th International Conference on Extending Database Technology (EDBT), March 2009.

7.2. Participation to symposia, seminars, invitations

Besides the presentations of our papers accepted to international conferences, the members of Dahu made the following keynote talks to international conferences or workshops.

Stéphane Demri gave a talk "Model checking memoryful logics over one-counter automata" during the Dagstuhl seminar "Beyond the Finite: New Challenges in Verification and Semistructured Data" (April 2008).

Florent Jacquemard gave a keynote talk on "Tree Automata Techniques for Infinite State Verification of Security Protocols" during the Dagstuhl seminar "Beyond the Finite: New Challenges in Verification and Semistructured Data" (April 2008).

7.3. Scientific Animations

- Stéphane Demri is a member of the steering committee of the Tableaux conference (Intl. Conf. Automated Theorem Proving with Analytic Tableaux and Related Methods).
- Stéphane Demri is member of the publication board of the review "Technique et Science Informatiques" (among 5 members).
- Stéphane Demri and Florent Jacquemard were members of the Commission de Spécialistes of ENS de Cachan, Number 6 (Computer Science) until august 2008.
- Stéphane Demri is co-editor of the proceedings of "15th International Symposium on Temporal Representation and Reasoning", IEEE, 2008.
- Florent Jacquemard is member of the board (general secretary) of the French Association for Information and Communication Systems (ASTI).
- Luc Segoufin was a member of the Commission d'Évaluation (CE) of INRIA until September 2008.

7.4. Teaching

As a Maître de conférence Cristina Sirangelo is teaching in the department of computer science of ENS de Cachan. In this department Stéphane Demri is in charge of the second-year students. Stéphane Demri also participated to the preparatory courses to "leçons d'aggrégation".

In the Master Parisien de Recherche en Informatique (MPRI), Florent Jacquemard is teaching in the first year (M1), a course on Tree Automata Techniques and Applications and Luc Segoufin is teaching in the second year (M2) a course on descriptive complexity, finite model theory and database theory.

8. Bibliography

Major publications by the team in recent years

[1] S. ABITEBOUL, L. SEGOUFIN, V. VIANU. *Static analysis of active XML systems*, in "ACM Symposium on Principles of Database Systems (PODS)", 2008, p. 221-230.

[2] M. BOJAŃCZYK, L. SEGOUFIN. *Tree languages defined in first-order logic with one quantifier alternation*, in "International Colloquium on Automata, Languages and Programming (ICALP)", 2008.

[3] M. BOJAŃCZYK, L. SEGOUFIN, H. STRAUBING. *Piecewise testable tree languages*, in "IEEE Symposium on Logic in Computer Science (LICS)", 2008.

[4] A. BOUHOULA, F. JACQUEMARD. *Automated Induction with Constrained Tree Automata*, in "Proceedings of the 4th International Joint Conference on Automated Reasoning (IJCAR'08), Sydney, Australia", A. ARMANDO, P. BAUMGARTNER, G. DOWEK (editors), Lecture Notes in Artificial Intelligence, vol. 5195, Springer-Verlag, August 2008, p. 539-553, http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/BJ-ijcar08.pdf.

[5] H. COMON-LUNDH, F. JACQUEMARD, N. PERRIN. *Visibly Tree Automata with Memory and Constraints*, in "Logical Methods in Computer Science", vol. 4, n° 2:8, June 2008, http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/CJP-lmcs08.pdf.

[6] S. DEMRI, R. LAZIĆ. *LTL with the freeze quantifier and register automata*, in "ACM Transactions on Computational Logic", To appear, 2008, http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/DL-tocl08.pdf.

[7] S. DEMRI, R. LAZIĆ, A. SANGNIER. *Model checking freeze LTL over one-counter automata*, in "Proceedings of the 11th International Conference on Foundations of Software Science and Computation Structures (FoSSaCS'08), Budapest, Hungary", R. AMADIO (editor), Lecture Notes in Computer Science, vol. 4962, Springer, March-April 2008, p. 490-504, http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/DLS-fossacs08.pdf.

[8] F. JACQUEMARD, M. RUSINOWITCH. *Closure of Hedge-Automata Languages by Hedge Rewriting*, in "Proceedings of the 19th International Conference on Rewriting Techniques and Applications (RTA'08), Hagenberg, Austria", A. VORONKOV (editor), Lecture Notes in Computer Science, vol. 5117, Springer, July 2008, p. 157-171, http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/JR-rta08.pdf.

[9] F. JACQUEMARD, M. RUSINOWITCH, L. VIGNERON. *Tree automata with equality constraints modulo equational theories*, in "Journal of Logic and Algebraic Programming", vol. 75, n° 2, April 2008, p. 182-208, http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/JRV-jlap08.pdf.

[10] B. TEN CATE, L. SEGOUFIN. *XPath, transitive closure logic, and nested tree walking automata*, in "ACM Symposium on Principles of Database Systems (PODS)", 2008, p. 251-260.

Year Publications

Articles in International Peer-Reviewed Journal

[11] H. COMON-LUNDH, F. JACQUEMARD, N. PERRIN. *Visibly Tree Automata with Memory and Constraints*, in "Logical Methods in Computer Science", vol. 4, n° 2:8, June 2008, http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/CJP-lmcs08.pdf.

[12] S. DEMRI, R. LAZIĆ. *LTL with the freeze quantifier and register automata*, in "ACM Transactions on Computational Logic", To appear, 2008, http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/DL-tocl08.pdf.

[13] F. JACQUEMARD, M. RUSINOWITCH, L. VIGNERON. *Tree automata with equality constraints modulo equational theories*, in "Journal of Logic and Algebraic Programming", vol. 75, n° 2, April 2008, p. 182-208, http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/JRV-jlap08.pdf.

International Peer-Reviewed Conference/Proceedings

[14] S. ABITEBOUL, L. SEGOUFIN, V. VIANU. *Static analysis of active XML systems*, in "ACM Symposium on Principles of Database Systems (PODS)", 2008, p. 221-230.

[15] C. ARECES, D. FIGUEIRA, S. FIGUEIRA, S. MERA. *Expressive Power and Decidability for Memory Logics*, in "Proceedings of the 15th Workshop on Logic, Language, Information and Computation (WoLLIC'08), Edinburgh, Scotland, UK", W. HODGES, R. DE QUEIROZ (editors), Lecture Notes in Computer Science, vol. 5110, Springer, July 2008, p. 56-68, http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/AFFM-wollic08.pdf.

[16] M. BOJAŃCZYK, L. SEGOUFIN. *Tree languages defined in first-order logic with one quantifier alternation*, in "International Colloquium on Automata, Languages and Programming (ICALP)", 2008.

[17] M. BOJAŃCZYK, L. SEGOUFIN, H. STRAUBING. *Piecewise testable tree languages*, in "IEEE Symposium on Logic in Computer Science (LICS)", 2008.

[18] A. BOUHOULA, F. JACQUEMARD. *Automated Induction with Constrained Tree Automata*, in "Proceedings of the 4th International Joint Conference on Automated Reasoning (IJCAR'08), Sydney, Australia", A. ARMANDO, P. BAUMGARTNER, G. DOWEK (editors), Lecture Notes in Artificial Intelligence, vol. 5195, Springer-Verlag, August 2008, p. 539-553, http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/BJ-ijcar08.pdf.

[19] C. DAVID. *Complexity of Data Tree Patterns over XML Documents*, in "Intl. Symp. on Mathematical Foundations of Computer Science (MFCS)", 2008, p. 278-289.

[20] S. DEMRI, R. LAZIĆ, A. SANGNIER. *Model checking freeze LTL over one-counter automata*, in "Proceedings of the 11th International Conference on Foundations of Software Science and Computation Structures (FoSSaCS'08), Budapest, Hungary", R. AMADIO (editor), Lecture Notes in Computer Science, vol. 4962, Springer, March-April 2008, p. 490-504, http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/DLS-fossacs08.pdf.

[21] A. GASCÓN, G. GODOY, F. JACQUEMARD. *Closure of Tree Automata Languages under Innermost Rewriting*, in "Proceedings of the 8th International Workshop on Reduction Strategies in Rewriting and Programming (WRS'08), Castle of Hagenberg, Austria", A. MIDDELDORP (editor), Electronic Notes in Theoretical Computer Science, To appear, Elsevier Science Publishers, July 2008, http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/GGJ-wrs08.pdf.

[22] F. JACQUEMARD, M. RUSINOWITCH. *Closure of Hedge-Automata Languages by Hedge Rewriting*, in "Proceedings of the 19th International Conference on Rewriting Techniques and Applications (RTA'08), Hagenberg, Austria", A. VORONKOV (editor), Lecture Notes in Computer Science, vol. 5117, Springer, July 2008, p. 157-171, http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/JR-rta08.pdf.

[23] TH. PLACE. *Characterization of Logics Over Ranked Tree Languages*, in "Proceedings of the 16th Annual EACSL Conference on Computer Science Logic (CSL'08), Bertinoro, Italy", M. KAMINSKI, S. MARTINI (editors), Lecture Notes in Computer Science, vol. 5213, Springer, September 2008, p. 401-415, http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/place-csl08.pdf.

[24] B. TEN CATE, L. SEGOUFIN. *XPath, transitive closure logic, and nested tree walking automata*, in "ACM Symposium on Principles of Database Systems (PODS)", 2008, p. 251-260.