COLÓQUIO EM INFORMÁTICA: BRASIL / INRIA, COOPERAÇÕES, AVANÇOS E DESAFIOS

COLIBRI 2009

COLLOQUE D'INFORMATIQUE: BRÉSIL / INRIA, COOPÉRATIONS, AVANCÉES ET DÉFIS

COLIBRI - Colóquio em Informática: Brasil / INRIA, Cooperações, Avanços e Desafios

COLloque d'Informatique: BRésil / Inria, coopérations, avancées et défis

(COLIBRI 2009)

http://gppd.inf.ufrgs.br/colibri
http://www-direction.inria.fr/international

22-23 de julho de 2009
Bento Gonçalves
RS – Brazil

Editores
Sociedade Brasileira de Computação — SBC

Organizadores
Nelson Maculan (UFRJ)
Malik Ghallab (INRIA)

Universidade Federal do Rio Grande do Sul — Instituto de Informática
http://www.inf.ufrgs.br

Capa: Rodrigo Righi
Supervisão Gráfica: Philippe O. A. Navaux, Nicolas Maillard
Impressão: Evangraf, Porto Alegre - RS

CIP – CATALOGAÇÃO NA PUBLICAÇÃO

COLIBRI - Colóquio em Informática: Brasil / INRIA, Cooperações, Avanços e Desafios (1.: 22-23 jul 2009: Porto Alegre)

Anais / Editores: Sociedade Brasileira de Computação — SBC. — Porto Alegre: Instituto de Informática da UFRGS, 2009.

224 f.: il.

ISBN 978-85-7669-245-4

Conhecido também como COLIBRI 2009.

1. 2. I. COLIBRI (1.: 22-23 jul 2009: Porto Alegre). II. Universidade Federal do Rio Grande do Sul. Instituto de Informática. III. SBC, Sociedade Brasileira de Computação — . IV. Título.

Instituto de Informática
Endereço:
Av. Bento Gonçalves, 9500 Bloco IV
Caixa Postal 15064 – CEP 91501-970 – Porto Alegre, RS – Brasil
Telefone: (51) 3316 6165 Fax: (51) 3316 7308
E-mail: informat@inf.ufrgs.br
http://www.inf.ufrgs.br
INRIA
Endereço:
Domaine de Voluceau-Rocquencourt
BP 105 - 78153 Le Chesnay cedex, France
Telefone: +33 (0)139635511
http://www.inria.br

Sociedade Brasileira de Computação

Presidente: José Carlos Maldonado (ICMC - USP)
Vice-Presidente: Virgílio Augusto Fernandes Almeida (UFMG)

Diretorias

Administrativa: Carla Maria Dal Sasso Freitas (UFRGS)
Finanças: Paulo Cesar Masiero (ICMC - USP)
Eventos e Comissões Especiais: Marcelo Walter (UFPE)
Educação: Edson Norberto Cáceres (UFMS)
Publicações: Karin Breitman (PUC-Rio)
Planejamento e Programas Especiais: Augusto Sampaio (UFPE)
Secretarias Regionais: Aline Maria Santos Andrade (UFBA)
Divulgação e Marketing: Altigran Soares da Silva (UFAM)

Diretorias Extraordinárias

Regulamentação da Profissão: Ricardo de Oliveira Anido (UNICAMP)
Eventos Especiais: Carlos Eduardo Ferreira (USP)
Cooperação com Sociedades Científicas: Taisy Silva Weber (UFRGS)

Conselho

Mandato 2007-2011
Cláudia Maria Bauzer Medeiros (UNICAMP)
Roberto da Silva Bigonha (UFMG)
Cláudio Leonardo Lucchesi (UNICAMP)
Daltro José Nunes (UFRGS)
André Ponce de Leon F. de Carvalho (ICMC - USP)

Mandato 2005 - 2009
Ana Carolina Salgado (UFPE)
Jaime Simão Sichman (USP)
Daniel Schwabe (PUC-Rio)
Vera Lúcia Strube de Lima (PUCRS) — Raul Sidnei Wazlawick (UFSC)

Suplentes - Mandato 2007-2009
Ricardo Augusto da Luz Reis (UFRGS) — Jacques Wainer (UNICAMP)
Marta Lima de Queiroz Mattoso (UFRJ)

COLIBRI 2009 • 22-23 de julho de 2009 • ISBN 978-85-7669-245-4

COLIBRI 2009 • 22-23 de julho de 2009 • ISBN 978-85-7669-245-4

Message from the Committees

The Colloquium COLIBRI was born from the will of French and Brazilian researchers to meet at the occasion of the "Ano da França no Brasil 2009", in order to exchange experiences on the most current topics in Technologies of Information. It was also an opportunity to commemorate 35 years of continuous scientific relations.

This colloquium would not have been possible without a strong commitment from the INRIA, the SBC, and the INF/UFRGS. From the INRIA side, the Direction de Partenariats Européens (DPE - Direction of European Partnerships) wished to associate the scientific partners of both sides in European projects. The Direction de Relations Internationales (DRI - Direction of International Relationships) aimed at the opportunity to have a balance of the current and future collaborations, as well as to meet its institutional partners — especially those with whom calls for projects are issued.

The perspective of meeting students and scientists, actively engaged in their domains, at the yearly SBC conference, which is the very place where they regularly gather, has suggested to associate COLIBRI to the CSBC'09. The SBC has immediately agreed on this project and has supported it all along its preparation. Finally, nothing could have been possible without the enthusiasm of the organizers from the INF/UFRGS, who have succeeded in overcoming one by one, and with strong determination, all the obstacles.

Thus, COLIBRI has been set up as a two days colloquium, with an accumulation of events devised to foster the meetings among the participants.

The organization and the preparation have been led by a bilateral Steering Committee, gathering the partners from the INRIA, the SBC and the INF/UFRGS. The scientific program has been elaborated by a Program Committee constituted of renown researchers from both countries, under the co-presidency of Nelson Maculan and Malik Ghallab.

The call for participation has been especially well received, since 44 articles have been submitted to COLIBRI. There were 94 Brazilian authors, and 35 French, belonging respectively to 30 institutions in Brazil (from 13 States), and 23 in France. The INRIA is but one of these since many French authors belong to mixed project-teams. This slight unbalance may result from the special attention that France has had this year in Brazil. It is also due to 17 articles which have been submitted unilaterally from Brazilian teams, *vs.* 3 unilateral French papers. On the whole, most of the articles have been co-signed by Brazilian and French researchers, and highlight works of common interest.

The topics that have been treated are, in decreasing order: Multimedia Data Treatment and Virtual Reality (11), Software Engineering and Algorithms (7), Biocomputing and Applications to Health (7), Network and Nomadism (6), Automatics and Numerical Analysis (5), Impacts and Perspectives (4).

Two speakers of each country have accepted to present their most recent works, as well as grand challenges in many fields of IT, in particular in "Networks and Nomadism", and "High Performance Computing". Two panels are also scheduled, one about the cooperation between Brazil and the European Union, and a second to discuss the international cooperation between the Brazilian financial entities and the French ones.

We would like to thank here all the members of the different committees (Program, Steering and Organization) who have contributed to the success of this manifestation, sometimes giving much of

their valuable time. We also thank all the speakers and participants for their involvement. We particularly appreciate the financial sponsoring of the French Embassy, the CAPES, Bull and Microsoft. Finally, we would like to acknowledge the representation of many State and Federal Agencies, which have agreed to come and share their experience on the topics dealing with international cooperation.

We hope that these proceedings will contribute to shed a light on the most active research areas between our two countries, and to give us all the motivation to keep on progressing together.

Sumário

Comitês - Comités

Comitê de acompanhamento bilateral - Comité de suivi bilatéral

Claudia Bauzer Medeiros (UNICAMP)
Jean-Pierre Courtiat (Embaixada da França, Brasilia)
Jean-Loïc Delhaye (INRIA, DPE)
Pierre Deransart (INRIA)
Bruno Le Dantec (INRIA, DPE)
José Carlos Maldonado (SBC)
Philippe Olivier Alexandre Navaux (INF, UFRGS)
Dominique Sotteau (INRIA, DRI)

Comitê de programa científico - Comité de programme scientifique

Co-presidentes:
Malik Ghallab (INRIA) — Nelson Maculan (UFRJ)

Membros:
Roberto da Silva Bigonha (UFMG)
Mariza da Silva Bigonha (UFMG)
Pierre-Alexandre Bliman (INRIA Paris-Rocquencourt)
Jean-Daniel Boissonnat (INRIA Sophia Antipolis - Méditerranée)
Frederic Bonnans (INRIA Saclay - Île-de-France)
Francisco de Assis Tenório de Carvalho (UFPe)
Luiz Mariano Carvalho (IME-UERJ)
Paulo Roberto Freire Cunha (UFPe)
David Deharbe (UFRN)
Marie-Dominique Devignes (INRIA Nancy-Grand Est)
Afonso Ferreira (COST-EU)
Alfredo Goldman (USP)
Pascal Guitton (INRIA Bordeaux - Sud-Ouest)
Frédéric Havet (INRIA Sophia Antipolis - Méditerranée)
Isabelle Herlin (INRIA Paris-Rocquencourt)
Michel Hurfin (INRIA Rennes - Bretagne Atlantique)
Jean-Marc Jézéquel (INRIA Rennes - Bretagne Atlantique)
Yves Lechevallier (INRIA Paris-Rocquencourt)
Sergio Lifschitz (PUC-Rio)
Raimundo José de Araújo Macêdo (UFBa)
Eliane Martins (UNICAMP)

Marta Lima Queirós Mattoso (UFRJ)
Alba Cristina Magalhães de Melo (UnB)
José Palazzo Moreira de Oliveira (UFRGS)
Bruno Raffin (INRIA Grenoble - Rhône-Alpes)
Michel Raynal (INRIA Rennes - Bretagne Atlantique)
Celso Carneiro Ribeiro (UFF)
Michaël Rusinowitch (INRIA Nancy-Grand Est)
Marie-France Sagot (INRIA Grenoble - Rhône-Alpes)
Ana Carolina Salgado (UFPe)
Pierre Sens (INRIA Paris-Rocquencourt)
Patrick Valduriez (INRIA Rennes - Bretagne Atlantique)
Rosa Maria Viccari (UFRGS)
Artur Ziviani (LNCC)

Comitê de organização - Comité d'Organisation

Philippe Olivier Alexandre Navaux (Presidente)
Alexandre Carissimi
Carla Maria dal Sasso Feitas
Cláudio Fernando Resin Geyer
José Valdeni de Lima
Leandro Wives
Lourdes Tassinari
Luciana Nedel
Nicolas Maillard

Brazil - INRIA: A short history of a rewarding cooperation

The cooperation between France's Institut National de Recherche en Informatique et Automatique (INRIA) and Brazil, which has been developing steadily since its inception, has been ongoing for more than thirty years. Founded officially on 3 January 1967, IRIA, which became INRIA in 1979, is based in Rocquencourt, near Paris. It happens that the foundation of INRIA virtually coincided with the signature of the technical and scientific cooperation agreement between France and Brazil on 16 January 1967. Doubtless this was a propitious coincidence!

During the debates at France's lower house of Parliament (or "Assemblée Nationale"), concerning the creation of INRIA, the French Minister for Scientific Research stated that one of the goals pursued was "the expansion of our research effort through the formation of international links"[1]. IRIA was therefore tasked with developing high-level research in an emergent field considered as strategic in the face of ever-growing world competition. Meanwhile, the Institute would maintain strong links with Industry and back the creation of start-up companies, focusing primarily on the development of innovative software applications. As soon as it had been founded, the Institute began to strongly develop its international relations, which at first was oriented towards the universities of North America, but subsequently opened up towards the countries of the East and the South, especially South America, with which INRIA was soon taking part in a number of cooperation agreements, mainly involving Brazil. The Institute's relationship with that country grew very fast, curiously marked each decade by a large event.

In 1977 (that is to say ten years after the creation of IRIA) a series of very regular exchanges began, with courses dispensed in Rio de Janeiro, at UFRJ and LNCC, by Jacques-Louis Lions (who at the time was Head of the "research"arm of IRIA: LABORIA) on the topic of mathematical physics problems. That year, one of the first of these entirely developed at the Institute by a Brazilian student was presented.

In 1987, the Institute, which had been renamed INRIA, went one step further, with the creation of a position entitled "Head of Relations with Brazil". This move provided a boost to the already well-established relationship between INRIA and Brazil, with the launch of several doctoral theses, as well as bilateral research projects partly backed by the French Ministry of Foreign Affairs[2].

The first cooperation agreement between INRIA and CNPq was signed in Paris during a meeting of the Joint Commission on 27 November 1997. This agreement grew out of the participation of INRIA in the assessment seminars for ProTem-CC, a major CNPq program that was designed to foster the development and deployment of international-level new technology teams in Brazil. With no less than seven cooperation agreements[3] signed with Brazil, INRIA is clearly determined to give every priority to its cooperation with that country. Each year, INRIA, together with many FAPs and

[1] Except from the book "Histoire d'un pionnier de l'informatique, 40 ans de recherche à l'INRIA"[History of a pioneer of IT, 40 years of research at INRIA], by Alain Beltran and Pascal Grisset, EDP Science, 2007, 290p.

[2] See Chapter of "France Brésil, 20 ans de coopération"(science and technology) coordinated by Luiz Claudio Cardoso et Guy Martinière, Collection "Travaux et Mémoires"n° 44, IHEAL, PUG, 1989.

[3] FAPESP and FAPERJ (2003), FACEPE (2004), FAPERGS (2005), FAPEMIG (2008) and FAPESB (2009).

CNPq, offer their researchers multiple cooperation opportunities for which the financing is contributed equally by both parties. Requests for intensifying the exchanges between INRIA and Brazil are growing all the time, both on the part of INRIA and of Brazil, as witnessed by the marked increase in the budgets dedicated to the resulting bilateral and multilateral relationships.

This determination to weave links with Brazil is not merely motivated by political considerations. It takes the form, first and foremost, of multiple exchanges involving researchers and doctoral students. In 2008 INRIA hosted 59 Brazilians and dispatched over sixty researchers to Brazil. The exchanges include doctoral students and have resulted in some fifteen Brazilians being permanently based at INRIA. These impressive figures result from the fact that INRIA and its Brazilian partners have always given priority to cooperation in the field of further education, by promoting mobility among the younger strata (masters and doctoral students). INRIA takes on students from fourteen Brazilian partners as part of its Internships program[4] and hosted eight interns in 2008. This motivational program is designed to lay the ground for future collaboration between INRIA and Brazil. Also, by financing four Associate Teams[5], and by backing the many bilateral collaboration projects of its researchers, to the tune of some ten per year on average, INRIA has chosen to promote and develop cooperation ventures between INRIA and Brazil at the highest level. These bilateral exchanges are in the process of gaining new impetus through relational partnerships. Given the development of the new IT and Communication Technologies sectors as well as Mathematics in South America, regional research programs have been implemented in those very fields. INRIA has decided to commit itself to the STIC AmSud program, which encompasses France, Argentina, Brazil, Chile, Peru and Uruguay. As a result INRIA is taking part in twelve projects, six of which involve Brazil. Thanks to the MATH AmSud program, INRIA has witnessed the inception of two new projects, both of them involving Brazil. The challenge of networking the research staff of INRIA with their Brazilian counterparts has thus to a large extent been met. INRIA's project teams also make use of the resources provided by other cooperation programs backed by CAPES-COFECUB. They also take part in the ANR/FINEP invitations to tender and have the possibility of involving Brazilian partners in European projects.

The relationship, which began with INRIA supporting Brazilian research, has now become a balanced one, with Brazil acting as a strategic partner in many fields. After more than thirty years of cooperation, INRIA and Brazil can consider one another as close partners indeed[6].

Pierre Deransart
Philippe O. A. Navaux
Dominique Sotteau

[4]The Internships program is designed to develop the Institute's ability to host foreign students.
[5]INRIA financing program for cooperation between INRIA and foreign research teams.
[6]For more details and contacts see: http://www-direction.inria.fr/international/PagePays.php?pays=Bresil .

Invited Talks and Panels

COLIBRI 2009 • 22-23 de julho de 2009 • ISBN 978-85-7669-245-4

Grand Research Challenges in Computer Science in Brazil – initiatives from the Brazilian Computer Society

Claudia Bauzer Medeiros[1]

[1]UNICAMP

In 2006, the Brazilian Computer Society organized a workshop to explore major research issues in Computer Science, within the Brazilian and international scenario. The result was the identification of five grand research challenges, that cover five main research axes - data, models, new architectures, people and omnivalence (a word coined during the workshop). These challenges motivated many kinds of activities in Brazil - several public RFPs, new research initiatives. They were also at the heart of a similar initiative, in 2008, for Latin America (Charla), and had a follow-up in 2009, in Manaus. The talk will present the Challenges, and discuss some of the associated outcomes for the Brazilian and Latin American scenario.

COLIBRI 2009 • 22-23 de julho de 2009 • ISBN 978-85-7669-245-4

Les Réseaux Sans Fil: Les Défis et Les Perspectives

José Ferreira de Rezende[1]

[1]COPPE-UFRJ

COLIBRI 2009 • 22-23 de julho de 2009 • ISBN 978-85-7669-245-4

Software Transactional Memory: What? Why? How? A New Challenge?

Michel RAYNAL
IRISA, Campus de Beaulieu, 35042 Rennes Cedex, France
raynal@irisa.fr

Abstract

The recent advance of multicore architectures and the deployment of multiprocessors as the mainstream computing platforms have given rise to a new concurrent programming impetus. Software transactional memories (STM) are one of the most promising approach to take up this challenge. The aim of a STM system is to discharge the application programmer from the management of synchronization when he/she has to write multiprocess programs. His/her task is to decompose his/her program in a set of sequential tasks that access shared objects, and to decompose each task in atomic units of computation. The management of the required synchronization is ensured by the associated STM system. This talk presents the motivation and the basics for STM systems. It visits first their aim and the consistency conditions on which they rely. It then presents a few existing STM protocols, and proposes open problems that should challenge researchers.

Keywords: Concurrent programming, Consistent global state, Consistency condition, Linearizability, Lock, Logical clock, Opacity, Serializability, Shared object, Software transactional memory, Transaction.

COLIBRI 2009 • 22-23 de julho de 2009 • ISBN 978-85-7669-245-4

The challenging advent of multicore architectures The speed of light has a limit. When combined with other physical and architectural demands, this physical constraint places limits on processor clocks: their speed is no longer rising. Hence, software performance can no longer be obtained by increasing CPU clock frequencies. To face this new challenge, (since a few years ago) manufacturers have investigated and are producing what they call *multicore architectures*, i.e., architectures in which each chip is made up of several processors that share a common memory. This constitutes what is called "the multicore revolution" [3].

The main challenge associated with multicore architectures is "how to exploit their power?" Of course, the old (classical) "multi-process programming" (multi-threading) methods are an answer to this question. Basically, these methods provide the programmers with the concept of a *lock*. According to the abstraction level considered, this lock can be a semaphore object, a monitor object, or the programmer's favorite synchronization object.

Unfortunately, traditional lock-based solutions have inherent drawbacks. On one side, if the set of data whose accesses are controlled by a single lock is too large (large grain), the parallelism can be drastically reduced. On another side, the solutions where a lock is associated with each datum (fine grain), are error-prone (possible presence of subtle deadlocks), difficult to design, master and prove correct. Finally, but not the least, lock-based solutions do not compose (this is certainly one of their main drawbacks). In other words, providing the application programmers with locks is far from being the panacea when one has to produce correct and efficient multi-process (multi-thread) programs. Interestingly enough, multicore architectures have (in some sense) rang the revival of concurrent programming.

The Software Transactional Memory approach The concept of *Software Transactional Memory* (STM) is an answer to the previous challenge. The notion of transactional memory has first been proposed (fifteen years ago) by Herlihy and Moss to implement concurrent data structures [4]. It has then been implemented in software by Shavit and Touitou [7], and has recently gained a great momentum as a promising alternative to locks in concurrent programming [1, 2, 5, 6].

Transactional memory abstracts the complexity associated with concurrent accesses to shared data by replacing locking with atomic execution units. In that way, the programmer has to focus where atomicity is required and not on the way it has to be realized. The aim of a STM system is consequently to discharge the programmer from the direct management of synchronization entailed by accesses to concurrent objects.

More generally, STM is a middleware approach that provides the programmers with the *transaction* concept (this concept is close but different from the notion of transactions encountered in databases [1]). More precisely, a process is designed as (or decomposed into) a sequence of transactions, each transaction being a piece of code that, while accessing any number of shared objects, always appears as being executed atomically. The job of the programmer is only to define the units of computation that are the transactions. He does not have to worry about the fact that the base objects can be concurrently accessed by transactions. Except when he defines the beginning and the end of a transaction, the programmer is not concerned by synchronization. It is the job of the STM system to ensure that transactions execute as if they were atomic.

Of course, a solution in which a single transaction executes at a time trivially implements transaction atomicity but is irrelevant from an efficiency point of view. So, a STM system has to do "its best" to execute as many transactions per time unit as possible. Similarly to a scheduler, a STM system is an on-line algorithm that does not know the future. If the STM is not trivial (i.e., it allows several transactions that access the same objects in a conflicting manner to run concurrently), this intrinsic limitation can direct it to abort some transactions in order to ensure both transaction atomicity and object consistency. From a programming point of view, an aborted transaction has no effect (it is up to the process that issued an aborted transaction to re-issue it or not; usually, a transaction that is restarted is considered as a new transaction).

COLIBRI 2009 • 22-23 de julho de 2009 • ISBN 978-85-7669-245-4

Content of the talk The talk is an introduction to STM systems, with an algorithmic flavor. It will first presents the basics of STM systems and their motivation. It will then presents a few existing STM systems, and will finally list open problems that should challenge the Franco-Brazilian community interested in (both failure-free and failure-prone) parallel and distributed computing.

References

[1] Felber P., Fetzer Ch., Guerraoui R. and Harris T., Transactions are coming Back, but Are They The Same? *ACM Sigact News, Distributed Computing Column*, 39(1):48-58, 2008.

[2] Harris T., Cristal A., Unsal O.S., Ayguade E., Gagliardi F., Smith B. and Valero M., Transactional Memory: an Overview. *IEEE Micro*, 27(3):8-29, 2007.

[3] Herlihy M.P. and Luchangco V., Distributed Computing and the Multicore Revolution. *ACM SIGACT News*, 39(1): 62-72, 2008.

[4] Herlihy M.P. and Moss J.E.B., Transactional Memory: Architectural Support for Lock-free Data Structures. *Proc. 20th ACM Int'l Symposium on Computer Architecture (ISCA'93)*, pp. 289-300, 1993.

[5] Larus J. and Kozyrakis Ch., Transactional Memory: Is TM the Answer for Improving Parallel Programming? *Communications of the ACM*, 51(7):80-89, 2008.

[6] Raynal M., Synchronization is Coming Back, but Is it the Same? Keynote Speech. *IEEE 22nd Int'l Conf. on Advanced Inf. Networking and Applications (AINA'08)*, IEEE Computer Press, pp. 1-10, Okinawa (Japan), 2008.

[7] Shavit N. and Touitou D., Software Transactional Memory. *Distributed Computing*, 10(2):99-116, 1997.

Tensor algebras for parallelism and synchronization

Brigitte Plateau[1]
Joint work with L. Brenner and A. Sales, Phd students, funded by CAPES

[1]Professor at the University of Grenoble – Director of the Lab. of Informatic of Grenoble
Project Team INRIA Mescal

Structured modeling formalisms are based on the description of complex systems by their parts. Such an approach usually necessitates to handle a large number of states and transitions. For systems described by structured modeling formalisms, we are concerned by predicting their performance and, for that purpose, promoting techniques that can manage efficiently the combinatoric explosion of the size of the states and transitions spaces.

The systems are studied under probabilistic assumptions concerning their temporal behavior, as well as some choices that are random. The mathematical framework is that of Markov Chains. The structured formalism we use here is Stochastic Automata Networks (SAN) [34, 19] but results could be transfered to other types of approaches, such as Stochastic Petri Nets [33, 30, 8, 11, 12], or Stochastic Process algebra [27, 23, 25, 42]. Another important feature is the representation of time : continuous or dicrete and we address both. Thus we have 3 important characteristics for these models: markovian framework, structured modelling, time scale.

The key issue is to capture the mathematical structure of those so-called *structured modeling formalisms*, where structure lies a priori in the way they are described. The mathematical structure is based on a matrix representation of the parts (or components) of the sytem and appropriate operators to combine them. These operators are inspired by the classical tensor algebra for matrices [28, 7, 14, 19], and extensions are proposed, which capture the relation between components. From this mathematical structure, efficient algorithms can be derived. These algorithms implement a decomposed computation process : they handle the global model as a sequence of phases which manage the characteristics of only one component and the memory requirements are reduced.

Seminal results were obtained in [19] for continuous time scale, with a *generalized* tensor algebra. Many other results [18, 15, 9, 36, 3, 10, 2, 22, 20, 39, 21, 37] have followed. Discrete time was also treated, but with behavior restrictions [31, 24, 35, 13, 44, 4, 26, 20]. The new result we present here is a tensor algebra, called *complex tensor algebra*, which is able to capture the mathematical structure of a quite large variety of discrete time models. Discrete time models are much more complex than continuous time models. For continuous time models, the state space is subject to combinatoric explosion but the transitions are triggered by single events, thanks to the markovian assumption. Discrete time models are concerned with the combinatoric explosion not only of the state space, but also, of the events cascade that can trigger a single transition.

The presentation will tackle the following topics :

- what are these models, why are they usefull and what is hard about them ?
- the main behavioral differences of structured models according to time scales
- the main basics and assets of tensor algebra [41, 17]
- extensions of tensor algebra [40]
- related issues of structured models [18, 15, 9, 3, 10, 2, 21, 22, 20, 39, 37]
- some examples [35, 16, 29, 1, 32, 43, 5]

The talk will not be deeply technical and interested readers can refer to L. Brenner and A. Sales, Phd reports [6, 38] for new results and publications.

1. References

[1] A. Benoit, L. Brenner, P. Fernandes, and B. Plateau. Aggregation of Stochastic Automata Networks with replicas. In 4^{th} *International Conference on the Numerical Solution of Markov Chains*, pages 145–166, Urbana, IL, USA, September 2003.

[2] A. Benoit, P. Fernandes, B. Plateau, and W. J. Stewart. On the benefits of using functional transitions and Kronecker algebra. *Performance Evaluation*, 58(4):367–390, 2004.

[3] A. Benoit, B. Plateau, and W. J. Stewart. Memory-efficient Kronecker algorithms with applications to the modelling of parallel systems. In *Proceedings of International Parallel and Distributed Processing Symposium*, pages 275–282, Nice, France, April 2003.

[4] Anne Benoit. *Méthodes et algorithmes pour l'évaluation des performances des systèmes informatiques à grand espace d'états*. PhD thesis, Institut National Polytechnique de Grenoble, Grenoble, France, 2003.

[5] L. Brenner, P. Fernandes, J. M. Fourneau, and B. Plateau. Modelling Grid5000 point availability with SAN. *Electronic Notes in Theoretical Computer Science*, 232:165–178, 2009.

[6] Leonardo Brenner. *Réseaux d'Automates Stochastiques : Analyse transitoire en temps continu et algebre tensorielle pour une sémantique en temps discret*. PhD thesis, Institut Polytechnique de Grenoble, Grenoble, France, 2009.

[7] J. W. Brewer. Kronecker Products and Matrix Calculus in System Theory. *IEEE Transactions on Circuits and Systems*, CAS-25(9):772–780, 1978.

[8] P. Buchholz. A class of hierarchical queueing networks and their analysis. *Queueing Systems*, 15(1-4):59–80, 1994.

[9] P. Buchholz, G. Ciardo, S. Donatelli, and P. Kemper. Complexity of memory-efficient Kronecker operations with applications to the solution of Markov models. *INFORMS Journal on Computing*, 13(3):203–222, 2000.

[10] P. Buchholz and T. Dayar. Block SOR for Kronecker structured representations. In *Proceedings of the 2003 International Conference on the Numerical Solution of Markov Chains*, pages 121–143, Urbana, IL, USA, 2003.

[11] P. Buchholz and P. Kemper. On generating a hierarchy for GSPN analysis. *ACM Performance Evaluation Review*, 26(2):5–14, 1998.

[12] M.-Y. Chung, G. Ciardo, S. Donatelli, N. He, B. Plateau, W. J. Stewart, E. Sulaiman, and J. Yu. A Comparison of Structural Formalisms for Modeling Large Markov Models. In *IPDPS Next Generation Software Program - NSFNGS - PI Workshop*, Santa Fe, New Mexico, USA, April 2004.

[13] G. Ciardo. Discrete-time Markovian stochastic Petri nets. In *Proceedings of the* 2^{th} *International Conference on the Numerical Solution of Markov Chains*, pages 339–358, Raleigh, NC, USA, 1995.

[14] M. Davio. Kronecker Products and Shuffle Algebra. *IEEE Transactions on Computers*, C-30(2):116–125, 1981.

[15] C. Durbach, F. Quessette, and A. Troubnikoff. Solving Large Markov Model based on Stochastic Automata Network. In *Advances in Computer and Informmation Sciences 1998, IOS press*, Belek-Antalya, Turkey, 1998.

[16] P. Fernandes and B. Plateau. Modeling Finite Capacity Queueing Networks with Stochastic Automata Networks. In *Fourth International Workshop on Queueing Networks with Finite Capacity (QNETs 2000)*, pages 1–12, Ilkley, West Yorkshire, UK, July 2000.

[17] P. Fernandes and B. Plateau. Triangular solution of linear systems in tensor product format. *ACM SIGMETRICS Performance Evaluation Review*, 28(4):30–32, 2001.

[18] P. Fernandes, B. Plateau, and W. J. Stewart. Numerical Issues for Stochastic Automata Networks. In *4th Workshop on Process Algebras and Performance Modelling*, pages 215–234, Torino, Italy, 1996. Ed. M. Ribaudo.

[19] P. Fernandes, Brigitte Plateau, and W.J. Stewart. Efficient descriptor-vector multiplications in stochastic automata networks. *JACM*, (3):381–414, 1998.

[20] J. M. Fourneau. Discrete time stochastic automata networks: using structural properties and stochastic bounds to simplify the SAN. In *Proceedings of the* 2^{nd} *International Conference on Performance Evaluation Methodolgies and Tools*, ACM International Conference Proceeding Series, pages 1–10, Nantes, France, October 2007. ACM.

[21] J. M. Fourneau. Product Form Steady-State Distribution for Stochastic Automata Networks with Domino Synchronizations. In *Proceedings of the* 5^{th} *European Performance Engineering Workshop*, volume 5261 of *Lecture Notes in Computer Science*, pages 110–124, Palma de Mallorca, Spain, September 2008. Springer.

[22] J. M. Fourneau, B. Plateau, and W. J. Stewart. Product form for stochastic automata networks. In *Proceedings of the* 2^{nd} *International Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS'07)*, Nantes, France, 2007. ACM.

[23] H. Garavel and H. Hermanns. On Combining Functional Verification and Performance Evaluation Using CADP. In *Proceedings of the International Symposium of Formal Methods Europe*, volume 2391 of *Lecture Notes in Computer Science*, pages 410–429, Copenhagen, Denmark, July 2002. Springer-Verlag.

[24] E. Gelenbe and G. Pujolle. *Introduction to queueing networks*. John Wiley & Sons, Inc., New York, NY, USA, 1987.

[25] N. Gootz, U. Herzog, and M. Rettelbach. Tipp - a stochastic process algebra. In *Proceedings of Workshop on Process Algebra and Performance Modelling (PAPM'93)*, pages 427–430, Edinburgh, UK, 1993.

[26] O. Gusak, T. Dayar, and J. M. Fourneau. Iterative disaggregation for a class of lumpable discrete-time stochastic automata networks. *Performance Evaluation*, 53(1):43–69, 2003.

[27] J. Hillston. Compositional Markovian Modelling Using a Process Algebra. In W. J. Stewart, editor, *Proceedings of the* 2^{nd} *International Workshop on the Numerical Solution of Markov Chains*. Kluwer, 1995.

[28] L. Kronecker. Der einige Interpolationsformeln fur ganze Funktionen mehrerer Variabeln. In *L. Kroneckers Werke*, volume I, pages 133–141, Lectures at the academy of sciences, December 21 1865. Chelsea Publishing Company.

[29] L. Mokdad, J. Ben-Othman, and A. Gueroui. Quality of Service of a Rerouting Algorithm Using Stochastic Automata Networks. In *Proceedings of the* 6^{th} *IEEE Symposium on Computers and Communications*, pages 338–343, Hammamet, Tunisia, July 2001. IEEE Computer Society.

[30] M. K. Molloy. Performance analysis using stochastic Petri nets. *IEEE Transactions on Computers*, C-31(9):913–917, 1982.

[31] M. K. Molloy. Discrete Time Stochastic Petri Nets. *IEEE Transactions on Software Engineering*, 11(4):417–423, 1985.

[32] V. Nguyen, R. Mathias, and G. D. Smith. A stochastic automata network descriptor for Markov chain models of instantaneously coupled intracellular Ca^{2+} channels. *Bulletin of Mathematical Biology*, 67(3):393–432, 2005.

[33] C. A. Petri. General net theory. In *Computing System Design: Proc. of the Joint IBM University of Newcastle upon Tyne Seminar*, pages 131–169. University of Newcastle upon Tyne, 1977.

[34] B. Plateau. On the stochastic structure of parallelism and synchronization models for distributed algorithms. In *Proceedings of the 1985 ACM SIGMETRICS conference on Measurements and Modeling of Computer Systems*, pages 147–154, Austin, Texas, USA, 1985. ACM Press.

[35] B. Plateau and K. Atif. Stochastic Automata Networks for modelling parallel systems. *IEEE Transactions on Software Engineering*, 17(10):1093–1108, 1991.

[36] O.H. Roux, D. Delfieu, and P. Molinaro. Discrete time approach of time petri nets for real-time systems analysis. *8th IEEE International Conference on Emerging Technologies and Factory Automation*, 2:197–204, 2001.

[37] A. Sales and B. Plateau. Reachable state space generation for structured models which use functional transitions. In *Proceedings of the* 6^{th} *International Conference on the Quantitative Evaluation of Systems (QEST'09)*, Budapest, Hungary, September 2009. IEEE Computer Society.

[38] Afonso Sales. *Réseaux d'Automates Stochatiques : Génération de l'espace d'états atteignables et Multiplication vecteur-descripteur pour une sémantique en temps discret*. PhD thesis, Institut Polytechnique de Grenoble, Grenoble, France, 2009.

[39] I. Sbeity, L. Brenner, B. Plateau, and W.J. Stewart. Phase-type distributions in Stochastic Automata Networks. *European Journal of Operational Research*, 186(3):1008–1028, 2008.

[40] W. J. Stewart. *Introduction to the numerical solution of Markov chains*. Princeton University Press, 1994.

[41] C. Tadonki and B. Philippe. Parallel Multiplication of a Vector by a Kronecker Tensor Product of matrices. *Parallel and Distributed Computing Practices*, 2(4):53–67, 1999.

[42] A. J. C. van Gemund. Symbolic Performance Modeling of Parallel Systems. *IEEE Transactions on Parallel and Distributed Systems*, 14(2):154–165, February 2003.

[43] Verena Wolf. Modelling of biochemical reactions by stochastic automata networks. *Electron. Notes Theor. Comput. Sci.*, 171(2):197–208, 2007.

[44] A. Zimmermann, J. Freiheit, and G. Hommel. Discrete Time Stochastic Petri Nets for the Modeling and Evaluation of Real-Time Systems. In *Proceedings of the* 15^{th} *International Parallel and Distributed Processing Symposium (IPDPS'01)*, pages 1069–1074, Washington, DC, USA, 2001. IEEE Computer Society Press.

Paineis — Tables Rondes

Bruno Le Dantec — Dominique Sotteau[1]

[1]INRIA

1. How can European Commission programs foster scientific collaborations with Brazil?

Objectives. For many years, Brazilian and French teams have been working together in the field of Information and Communication Technologies, essentially in a bilateral context. This session dedicated to European programs, will offer a great opportunity to discuss the possibilities given by the European Commission to foster interactions between both countries in a European Framework. The objectives of this session organiser's are to help developing closer research and business collaborations between Brazilian, French and European teams.

Three panellists have been invited to provide the audience with accurate updated information on (i) the European Commission Seventh Framework Programme (FP7) offering an opportunity for collaborations between Brasil and the European Union, (ii) the Brazilian priorities for cooperation with the EU in the ICT domain, (iii) a concrete example of EC projects involving Brazilian partners and potential support to be provided by the B-Bice Project. A dedicated Q&A "question and answer" session will allow participants to interact with the panellists to better understand the Brazilian priorities for cooperation and the matching possibilities offered by the European Commission programmes.

Participants

- Bruno Le Dantec (session coordinator) was born in 1963 in France. He has a master degree in commercial law and a postgraduate certificate in finance. He is in charge of strategic collaborations and interregional activities at the European Partnership Department of INRIA. Until 2008, he was the ERCIM deputy Manager leading the European department.
- Mr Paulo Lopes, Lopes is Counselor for Latin America in the fields of Information Society and Media at the European Commission's Delegation in Brazil since April 2007. Before that Mr Lopes was responsible since 2003 for relations with Latin America in the Directorate General for Information Society and Media of the European Commission in Brussels. Before integrating the EC in 1993, Mr Lopes has been working with the Portuguese regulatory authority for telecommunications, ANACOM. He was born in Lisbon and has a degree in telecoms engineering and electronics by the Instituto Superior Técnico de Lisboa.
- Felipe Costi Santarosa, is the head of the Divisão da Sociedade da Informação, Brazilian Ministério das Relações Exteriores,

- Paulo Cesar Gonçalves Egler, hold a PhD in Environmental Sciences from the University of East Anglia, UK. Since 1980 he has developed activities of scientific and technological planning and management in the Brazilian Federal Government. Nowadays works as advisor for international cooperation of the Brazilian Institute of Information in Science and Technology a research institute linked to the Brazilian Ministry of Science and Technology (Ibict/MCT), and as a researcher at the Centre of Advanced Studies of Government and Public Administration at the University of Brasilia (CEAG/UnB). Also acts as the Coordinator of the New Brazilian Bureau for Enhancing the International Cooperation with Europe (Project BB.Bice).

2. Coopération internationale Brésil-France, avancées et défis

Cette table ronde doit réunir, autour de représentants de l'INRIA, des représentants d'institutions françaises et brésiliennes. Elle devrait permettre à chaque institution représentée de faire état de sa vision de la coopération aujourd'hui, de présenter les outils qu'elle met à disposition pour aider à renforcer cette coopération, de souligner les avancées obtenues grâce à ces outils et les retombées positives qu'ils ont engendrées ; il ne sera pas interdit de faire aussi état des difficultés rencontrées, ce dans un esprit positif, afin de pouvoir chercher ensemble à y remédier. Les échanges qui s'en suivront devraient ouvrir la discussion autour de l'identification des nouveaux défis posés dans le domaine des sciences de l'information et de la communication, en relation avec le monde qui nous entoure et les problèmes de la société actuelle. Ils devraient pouvoir déboucher sur de nouvelles idées adaptées pour la résolution de ces défis ensemble, faire naitre de nouvelles propositions d'actions, ce afin d'ouvrir des perspectives qui permettraient un élargissement des coopérations (synergie entre les projets existants par exemple, regroupement de compétences, etc.).

Participants

- FAPESP, FAPEMIG, FAPERGS, FACEPE, CONFAP.
- CNPq, INRIA, Ambassade de France.

I

Software Engineering and Algorithms

COLIBRI 2009 • 22-23 de julho de 2009 • ISBN 978-85-7669-245-4

A Coalescing Algorithm for Aliased Registers

Mariza A. S. Bigonha[1], **Fabrice Rastello**[2],
Fernando Magno Quintão Pereira[1], **Roberto S. Bigonha**[1]

[1] Departamento de Ciência da Computação – Universidade Federal de Minas Gerais (UFMG)
Av. Antônio Carlos, 6627 – 31.270-010 – Belo Horizonte – MG – Brazil

{mariza,fpereira,bigonha}@dcc.ufmg.br

[2]Laboratoire de l'Informatique du Paralllisme – École normale supérieure de Lyon
46 alle d'Italie, 69364 – Lyon cedex 07 – France

Fabrice.Rastello@ens-lyon.fr

Abstract. *Register coalescing is a compiler optimization that removes copy instructions such as $a = b$ from a source program by assigning variables a and b to the same register. The vast majority of coalescing algorithms described in the literature assume homogeneous register banks; however, many important computer architectures, such as x86, ARM, SPARC and ST240 contain an irregularity called register aliasing. Two registers alias if assigning a value to one of them changes the contents of the other. Most of the time registers can be divided into subclasses that hierarchically fit into each other. The objective of this research is to design, implement and test new coalescing algorithms that handle hierarchical register aliasing. We expect that an aliasing aware coalescer will be able to remove more copy instructions than an otherwise oblivious algorithm; thus, decreasing the size and increasing the performance of compiled programs.*

1. Problem Description

Register allocation is the task of mapping the variables of a source program into a finite number of registers. If the number of registers is not sufficient, then some of the variables are mapped into memory. A compiler optimization that is performed on top of register allocation is *coalescing*. This optimization consists in mapping variables related by copy instructions to the same register. For instance, we can remove the instruction $a = b$ from the source program provided that a and b are mapped to the same register `r`.

A good coalescing algorithm can improve the execution speed of a program by as much as 12% [12]. Although register coalescing is important, and many coalescing algorithms have been proposed in the literature, these algorithms do not take *register aliasing* into consideration. This project aims at filling this gap. Its objective is to design, implement and test new coalescing algorithms that handle hierarchical register aliasing.

Two registers alias if an assignment to one of them affects the contents of the other. The best known example of aliasing is found in the x86 architecture, which has four general purpose 16-bit registers - `AX`, `BX`, `CX` and `DX` - that can also be used as eight 8-bit registers. That is, the x86 architecture combines two 8-bit registers into one 16-bit register. Figure 1 shows the bank of general purpose registers used in the x86. Notice that all the registers contain some sort of aliasing, but only the upper registers are divided into two parts. Another example of aliased registers is the combination of two aligned single

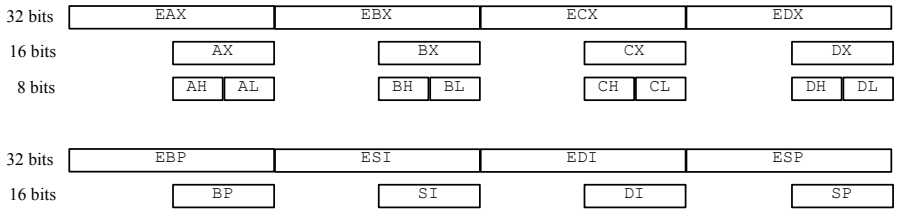


Figure 1. General purpose registers of the x86 architecture

precision floating-point registers to form one double-precision register. As for the x86, register aliasing of most architectures is restricted to hierarchical aliasing: at the lowest level the register bank is composed of atomic registers; the upper level partitions this set of atomic registers into non-intersecting register subsets; to each subset corresponds a register that aliases with its composing registers. Examples of such architectures we are concerned with include the Sun SPARC, the ARM processors, and the ST240. ARM Neon, and ST240 have two levels of aliasing while SPARC V8 has three: single precision floating point registers can be combined into double precision and even in quad precision registers. X86 has a subset of registers with just one level of aliasing - `SP`, `SI`, `DI` and `SP` - plus another subset, formed by registers `AX`, `BX`, `CX` and `DX`, with two levels of aliasing. X86 is thus an example of hybrid hierarchical aliasing.

2. Related Works

The coalescing problem is intrinsically related to the register allocation problem. For instance, Chaitin *et al.* [8] already describe a coalescing strategy in their pioneering work on register allocation via graph coloring. Since its first appearance, Chaitin's work has been the target of a slow, yet never-ending stream of improvements. One quarter century after Chaitin's seminal paper, coalescing has been one of the main forces pushing new variations in graph coloring register allocation. Bouchez *et al.* [3] summarizes some of the best known approaches for performing register coalescing:

- *Aggressive Coalescing* [8, 7]: merges move-related vertices, regardless of the colorability of the interference graph after the merging.
- *Conservative Coalescing* [5]: merges moves if, and only if, the merging does not compromise the colorability of the interference graph.
- *Optimistic Coalescing* [17, 18]: coalesces moves aggressively, and if it compromises the colorability of the graph, then gives up as few moves as possible.
- *Incremental Conservative Coalescing* [11]: removes one particular move instruction, while keeping the colorability of the graph.

Bouchez *et al.* [2, 3] have shown, by means of an ingenious sequence of reductions, that all these different materializations of the register coalescing problem are NP-complete for general interference graphs. Hack *et al.* [12] proposed a new scheme called *Recoloring Coalescing*: color the graph arbitrarily, recolor move-related nodes (to satisfy as much moves as possible) and their interfering neighborhood (to keep the coloring valid).

The finding that programs in Single Static Assignment (SSA) form have chordal interference graphs [2, 6, 13, 20] has given a new boost to research on register coalescing. *Static Single Assignment* (SSA) form is an intermediate representation in which each variable is defined at most once in the program code [9, 22]. Nowadays, there exist several

industrial and academic compilers using SSA in their back-end, such as LLVM [15], Sun's HotSpot JVM [25], IBM's Java Jikes RVM [26], LAO [1], and Firm [12]. It is possible to retain the SSA property until the end of the code generation process. Indeed, there exists polynomial time algorithms to discover the chromatic number of chordal graphs [10]; thus, register assignment has polynomial time solution for programs in the SSA representation. However, as a form of live range splitting, the SSA transformation inserts many copies into the program code, what makes register coalescing even more important. Unfortunately, as shown by Bouchez *et al.* [3], most coalescing instances remain NP-complete for chordal graphs. Two new coalescing algorithms in the context of SSA-form based register allocation have been presented in 2008: Hack *et al.* have proposed a recoloring coalescing algorithm [12], and Bouchez *et al.* have proposed a suite of conservative and optimistic algorithms [4].

Aliasing complicates register allocation substantially. For instance, finding the minimal register assignment in face of aliasing is NP-complete, even for programs in SSA-form [16]. Nevertheless, register allocators that handle aliasing have already been described [14, 23, 24]. The algorithm presented by Smith *et al.*, for instance, modifies a graph coloring approach to deal with this phenomenon. However, none of the coalescing algorithms discussed so far takes aliasing into consideration. Recently, Pereira *et al.* [21] have proposed a *Puzzle-Based* register allocator to handle hierarchical aliasing. In this paradigm, the register allocation problem is seen as a collection of puzzles that can be solved in polynomial time. This approach tends to reduce the number of variables sent to memory, at the expenses of increasing the number of copy instructions in the target code; thus, making register coalescing essential.

We can differentiate several coalescing problems. *Global* coalescing consists in minimizing the total amount of move instructions for the entire procedure. This problem is NP-complete [3], regardless of aliasing. *Local* coalescing is the restriction to a basic block. This problem has polynomial time solution in the absence of register aliasing, but Lee et al. have shown that it is NP-complete for architectures that present hierarchically aliased registers. The *biased* coloring problem is the simplest realization of the coalescing problem, consisting in minimizing the number of move instructions inserted between two instructions of a program. Pereira *et al.* have described an optimal algorithm for a particular case of biased coloring - level-1 alias hierarchy with no pre-assignment [19]. We are interested in the global version of the register coalescing problem.

3. The Proposed Approach to the Register Coalescing Problem

We will adapt the graph coloring based coalescing algorithm proposed by Bouchez *et al* [4] to handle computer architectures with hierarchical register aliasing. Our objective is to reduce the number of move instructions in the function being optimized, using, for instance, the expedient of manipulating two small values stored in different halves of the same register with one single instruction. For instance, if a, b, c and d are four single precision floating point values, and a and b are stored in the same double precision register, then we can implement the two copies $c := a$ and $d := b$ with one single register copy. In order to show the advantages of the new approach, we will compare our modified algorithm with Bouchez's original method, which does not take register aliasing into consideration. We will also compare it with the puzzle based register allocator [21], which handles register aliasing, but performs very simple coalescing, based on a biased coloring

strategy. The new register coalescing algorithm will be implemented and tested in the back-end of the Low Level Virtual Machine (LLVM) framework [15]. This framework is used, for instance, to JIT compile open-GL applications in Mac OSX 10.5.

4. Expected Results

We expect that our new algorithm will produce target code that is shorter and more efficient than the code produced by traditional coalescers that do not take register aliasing into consideration. Ideally we should be able to reduce the size of the programs produced by the puzzle based allocator [21] by 6-7%. Speed-up improvements will depend on the target architecture. We expect small improvements in the x86 processor; however, we should be able to obtain a 2-4% decrease in execution time on the PowerPC chip. Gains in efficiency should be more noticeable in embedded devises, such as the ARM Neon processor. In addition to the concrete contribution to the research community, this project will have the positive side-effect of creating an environment more favorable to further co-operations between Brazilian and French research institutes.

References

[1] Benoit Boissinot, Sebastian Hack, Daniel Grund, Benoit Dupont de Dinechin, and Fabrice Rastello. Fast liveness checking for SSA-form programs. In *CGO*, pages 35-44 IEEE, 2008.

[2] F. Bouchez. Allocation de registres et vidage en mémoire. Master's thesis, ENS Lyon, 2005.

[3] F. Bouchez, A. Darte, and Fabrice Rastello. On the complexity of register coalescing. In *CGO*, pages 102-104 IEEE, 2007.

[4] Florent Bouchez, Alain Darte, and Fabrice Rastello. Advanced conservative and optimistic register coalescing. In *CASES*, pages 147 – 156. ACM, 2008.

[5] P. Briggs, K. D. Cooper, and L. Torczon. Improvements to graph coloring register allocation. TOPLAS, 16(3), pages 428-455, 1994.

[6] P. Brisk, F. Dabiri, J. Macbeth, and M. Sarrafzadeh. Polynomial-time graph coloring register allocation. In *IWLS*, pages 150-155 ACM, 2005.

[7] G. J. Chaitin. Register allocation and spilling via graph coloring. *CC*, 17(6), pages 98–105, 1982.

[8] G. J. Chaitin, M. A. Auslander, A. K. Chandra, J. Cocke, M. E. Hopkins, and P. W. Markstein. Register allocation via coloring. *Computer Languages*, 6, pages 47–57, 1981.

[9] Ron Cytron, Jeanne Ferrante, Barry K. Rosen, Mark N. Wegman, and F. Kenneth Zadeck. Efficiently computing static single assignment form and the control dependence graph. *TOPLAS*, 13(4), pages 451–490, 1991.

[10] F. Gavril. Algorithms for minimum coloring, maximum clique, minimum covering by cliques, and maximum independent set of a chordal graph. *SICOMP*, 1(2), pages 180-187, 1972.

[11] Lal George and Andrew W. Appel. Iterated register coalescing. *TOPLAS*, 18(3), pages 300-324, 1996.

[12] S. Hack and G. Goos. Copy coalescing by graph recoloring. In *PLDI*, pages 227-237 ACM, 2008.

[13] Sebastian Hack, Daniel Grund, and Gerhard Goos. Register allocation for programs in SSA-form. In *CC*, pages 247–262, 2006.

[14] Timothy Kong and Kent D Wilken. Precise register allocation for irregular architectures. In *MICRO*, pages 297–307 ACM, 1998.

[15] Chris Lattner and Vikram S. Adve. Llvm: A compilation framework for lifelong program analysis & transformation. In *CGO*, pages 75–88 IEEE, 2004.

[16] J. K. Lee, J. Palsberg, and Fernando M. Q. Pereira. Aliased register allocation. In *ICALP*, pages 680-691, 2007.

[17] J. Park and S. Moon. Optimistic register coalescing. In *PACT*, p.196–204 IEEE, 1998.

[18] J. Park and S. Moon. Optimistic register coalescing. *TOPLAS*, 26(4):735-765, 2004.

[19] Fernando M. Q. Pereira. *Register Allocation by Puzzle Solving*. PhD thesis, University of California, Los Angeles, 2008.

[20] Fernando Magno Quintao Pereira and Jens Palsberg. Register allocation via coloring of chordal graphs. In *APLAS*, pages 315–329. Springer, 2005.

[21] Fernando M. Q. Pereira and J. Palsberg. Register allocation by puzzle solving. In *PLDI*, p. 216-226 ACM, 2008.

[22] B. K. Rosen, F. K. Zadeck, and M. N. Wegman. Global value numbers and redundant computations. In *POPL*, pages 12–27 ACM, 1988.

[23] Bernhard Scholz and Erik Eckstein. Register allocation for irregular architectures. In *LCTES/SCOPES*, pages 139–148 ACM, 2002.

[24] Michael D. Smith, Norman Ramsey, and Glenn Holloway. A generalized algorithm for graph-coloring register allocation. In *PLDI*, pages 277–288 ACM, 2004.

[25] JVM Team. The java HotSpot virtual machine. Technical Report Technical White Paper, Sun Microsystems, 2006.

[26] The Jikes Team. Jikes RVM home page, 2007. http://jikesrvm.sourceforge.net/.

Towards automating proofs for model-based software engineering

D. Déharbe[1] , P. Fontaine[2] , A. Martins Moreira[1] , S. Merz[2] , A. Santana de Oliveira[3]

[1]Universidade Federal do Rio Grande do Norte
Natal, RN, Brazil

[2]LORIA–INRIA
Nancy, France

[3]Universidade Federal Regional do Semi-Árido
Mossoró, RN, Brazil

{anamaria,anderson,david}@dimap.ufrn.br,

{Pascal.Fontaine,Stephan.Merz}@loria.fr

Abstract. *Model-based approaches are commonly used to engineering software for safety-critical applications. Several artifacts, such as abstract specifications, formal refinements, and assertions at the implementation level require computational support for formal reasoning activities. Delivering formal verification tools that address such reasoning activities in an automated, trustable and flexible fashion remains a scientific challenge. Researchers from LORIA and UFRN have been collaborating to address this challenge and are now developing an automated theorem prover targeted at trustable software verification efforts.*

1. Introduction

The software industry needs effective technologies to assist the design and development of software for safety-critical applications. This domain of application has its own safety-related standards, where strict requirements impose mathematical rigor and traceability to the development activities.

Model-driven design approaches such as the B method [1] have been used by the industry for more than a decade to address these needs. Such methods require that the designers produce a formal model of the requirements. The satisfiability and coherence of this model is then verified using theorem provers to establish the validity of verification conditions. This first model may then be used as a reference to establish the correctness of more concrete design artifacts, known as refinements. Theorem provers again play an essential role in this step, they ensure the overall soundness of the development with respect to the initial requirements. Support for these methods comes in the form of fully or partially automated provers, with varying degrees of efficiency, depending on the expressiveness of the specification language, as well as the maturity of the employed theorem proving technologies.

For the last few years, the collaboration between UFRN and LORIA has been focusing on the design of a state-of-the-art theorem prover that caters to the needs of such model-based software engineering methods (Section 2 of the paper gives a brief account

of the history the collaborations between the institutions). Details on the theorem prover currently under development are then given in Section 3. Finally, Section 4 provides a perspective on the scientific challenges that the authors aim to tackle in the future.

2. A short account of previous collaborations

Collaboration between UFRN and LORIA (encompassing the Cassis, Mosel, and Protheo/Pareo research teams) in the area of formal approaches to software development started formally in 2001 with the FERUS project, coordinated by Anamaria Martins Moreira on the Brazilian side and by Hélène Kirchner on the French side. This project was proposed in response to a joint call of CNPq and INRIA and addressed the reuse of software specification components in the algebraic realm. In addition to short scientific missions of the different partners, during the FERUS project UFRN lecturer Anamaria Martins Moreira realized a one-year long sabbatical at LORIA and her then master student Anderson Santana de Oliveira also realized a four-month stay at LORIA. Anderson Santana was later the recipient of a grant from CAPES to realize a full doctoral study at LORIA, starting in 2004. In 2008, Anderson Santana defended his thesis and after a post-doc stay at UFRN was hired as a lecturer by UFERSA.

In 2002, UFRN lecturer David Déharbe visited LORIA for a sabbatical stay with Michael Rusinowitch. This stay initiated a fruitful collaboration in the area of automatic theorem proving, involving Silvio Ranise, Christophe Ringeissen, and recently in particular Pascal Fontaine. This collaboration was formalized in the second joint UFRN/LORIA project under the sponsorship of CNPq and INRIA: Da Capo (2005-2008). Since then several former UFRN students initiated their doctoral studies at LORIA: Judson Santos Santiago (Cassis group, concluded[1]), Cláudia Fernanda Oliveira Kiermes Tavares (Pareo group, ongoing), Diego Caminho Barbosa de Oliveira (Mosel group, ongoing).

3. The veriT solver

The collaboration between Loria and UFRN in the area of automatic theorem proving has led to the active development of a common tool through which our research results are validated and made available to the larger community. Initially known as haRVey [6] whose first version was made available in 2002, it was essentially a research platform for combining reasoning tools. Under the DaCapo project, we began in 2007 to re-factor the code of haRVey and greatly enhance its capabilities. The new tool is called veriT solver; the release of a usable and stable version is imminent.

The veriT solver is best classified as a Satisfiability Modulo Theories (SMT) solver. It takes a predicate logic formula as input and checks if the formula is satisfiable or not, that is, if there is an interpretation that makes the formula true. The validity problem is trivially translated to the satisfiability problem by taking the negation of the formula; a formula is valid (always true) if and only if its negation is unsatisfiable. Most formal methods and verification techniques heavily rely on checking the satisfiability or the validity of formulas. In addition to giving a verdict on the satisfiability of its input, veriT is able to produce a proof of its result; such proof may be then checked or reused by external components. This feature is important as it makes it possible to certify the results produced by veriT.

[1] Judson Santos Santiago has also been hired as a lecturer by UFERSA.

The input language of veriT is a first-order language with uninterpreted and interpreted symbols based on the SMT-LIB format [13]. Uninterpreted symbols are particularly useful to model arrays, or functions that are partially or totally unknown, whereas interpreted symbols are essential to model the usual data structures used in computer systems such as integers, rationals, real numbers or lists. As a distinguishing feature, veriT includes a complete first-order generic prover (at the moment, the E-prover [14]; inclusion of SPASS [16] is planned). This feature makes it possible for the user to define his own data structures by writing axioms in logic that characterize the data structures and the operators working on them

Figure 1 depicts the architecture of veriT. First veriT reads the input formula, written in the SMT-LIB format [13]. Currently veriT implements two syntactical extensions to this format: lambda expressions and macros. The parser rewrites the instances of these constructs applying beta reduction and macro expansion. The result of this preprocessing may be output to a file in the SMT-LIB format without these extensions. Second, veriT computes the conjunctive normal form of the formula and stores the corresponding clauses in SAT-solver MINISAT [7], a propositional satisfiability solver, that implements a modern and efficient version of the DPLL algorithm [3]. The SAT-solver attempts to find a propositional model of the formula, i.e. a satisfiable set of literals. If no such model is found, then the formula is unsatisfiable. Otherwise, it is propositionally satisfiable. The set of literals is then incrementally input to the theory reasoning engine. The theory reasoning engine may conclude that the set of literals is satisfiable in the combination of first-order theories; in that case the original formula is satisfiable and the execution stops with this result. Another possible outcome is that the set of literals is unsatisfiable in the theory; in that case, a clause is added to the propositional SAT-solver to discard this set of literals from the set of possible models. Finally, the third possible outcome is that the theory reasoning module cannot reach a conclusion; in that case, it may provide lemmas that are translated to additional clauses that further constrains the set of possible models. In the last two cases, the SAT-solver is applied to the updated set of clauses and the process repeats until no more propositional models can be built (the formula is unsatisfiable), or no more lemmas can be generated (the solver is not able to decide the satisfiability of the input formula). For some logics, veriT is also able to produce a proof of the result, which consists of a sequence of instances of basic deduction rules that can be then checked by a third-party.

The theory reasoning module is itself composed of a decision procedure for the theory of equality with uninterpreted functions, based on congruence closure [12], a decision procedure for the so-called “difference logic”, a fragment of real and integer arithmetics [4], quantifier instantiation heuristics, and an automated deduction engine based on resolution and superposition [15]. Such module [5] implements a variant of the Nelson and Oppen combination technique [11], based on the propagation of variable equalities. Moreover, to avoid overflow and underflow errors in the manipulation of arithmetic values, veriT uses the GNU MP [9] libraries.

The language of veriT totally covers several sections of the competition of SMT solvers (SMT-COMP [2]). Although the efficiency of veriT is not yet at the level of tools winning those sections of the competition, its performances are on a par with serious general-purpose SMT solvers. We evaluated veriT, CVC3 and Z3 (both using the lat-

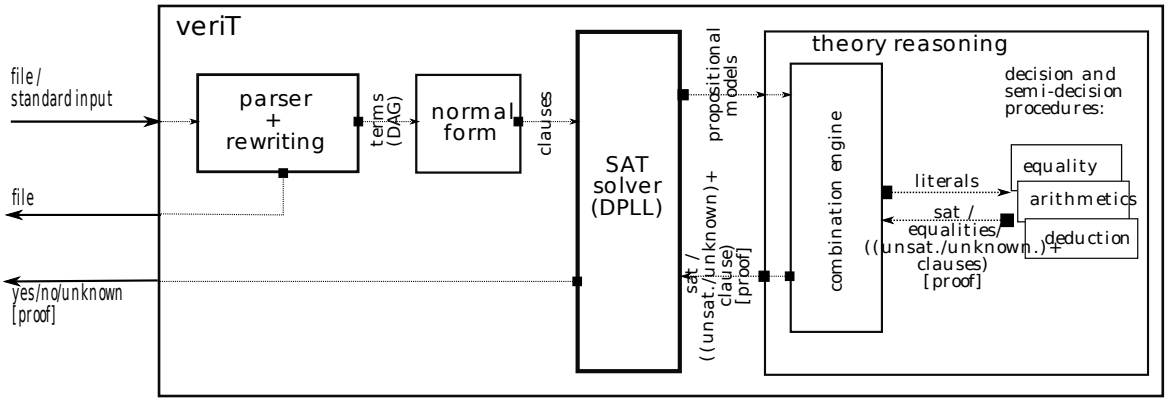


Figure 1. Software architecture of veriT

Solver	QF_UF (6656)	QF_UFIDL (432)	QF_IDL (1673)	QF_RDL (204)	all (8965)
veriT	6323	332	918	100	7673
CVC3	6378	278	802	45	7503
Z3	6608	419	1511	158	8696

Figure 2. Experimental comparison of veriT **with CVC3 and Z3**

est available version in February 2009) against the SMT-LIB benchmarks for QF_IDL, QF_RDL, QF_UF and QF_UFIDL (June 2008 version) using an Intel® Pentium® 4 CPU at 3.00 GHz with 1 GB of RAM and a timeout of 120 seconds. The Figure 2 gives, for each solver, the number of completed benchmarks.

The veriT solver has been successfully used for several verification tasks on lock-free algorithms, and to check refinements in model-based software engineering approach such as B and Circus. The tool is publicly available and distributed under the BSD open-source license. It may be downloaded at `http://www.verit-solver.org`.

Planned enhancements include better quantifier handling, hierarchic theories, new decision procedures for arithmetic fragments, and better integration of the tool within proof assistants.

4. Perspectives and conclusion

Our motivation for the development of the veriT solver is that the formal development of software should be supported by automated reasoning modules. As the solver gains maturity, another important task remains: the automation provided by veriT should be made available inside popular tool platforms supporting formal methods, such as the B method, TLA+, as well as the Isabelle and Coq proof assistants. We have several preliminary works along that line. The veriT solver has already been integrated as a plug-in within the Isabelle proof-assistant [8], some investigation to integrate veriT inside the Coq proof-assistant has been done, and veriT has been used successfully to discharge simple but very tedious proof obligations from B models [10].

In this context it is crucial to provide the users with syntactical criteria to identify the fragment of the language of their favorite formal method for which automatic proof

support is possible. This is not trivial, since syntactical transformations may often be used to convert proof obligations that occur naturally in popular formal methods towards the input language of automated solvers. For example, proof obligations for the B method are typically formulas using set theoretic concepts. A straight translation—using axioms for set theory—into the language of first-order or SMT provers leads to poor results. However veriT is able to handle some set formulas very efficiently, by understanding sets as their characteristic predicates and rewriting the usual operators on sets as boolean operators accordingly. Not every formula can be handled in this way, and notably, quantifiers over sets are not allowed in the original formula. Even in those cases, it is sometimes possible to use clever instantiations to obtain formulas that fall into the fragment handled by veriT, and it would be interesting to automatically recognize such cases. The same recognizers could also be applied to sub-goals in interactive proof assistants, thus simplifying the required expertise. As a last resort, a user aware of this automation capability could guide the manual proof to minimize the number of interactions necessary to prove a given verification condition.

The development of the veriT solver and its use as a component for tool sets supporting formal methods are at the core of the long term fruitful cooperation between the UFRN and LORIA teams.

References

[1] J.-R. Abrial. *The B-Book. Assigning Programs to Meanings*. Cambridge University Press, 1996.

[2] C. Barrett, L. de Moura, and A. Stump. SMT-COMP: Satisfiability Modulo Theories Competition. In *Computer Aided Verification (CAV)*, pages 20–23, 2005.

[3] M. Davis, G. Logemann, and D. Loveland. A machine program for theorem-proving. *Communications of the ACM*, 5:394–397, 1962.

[4] D. C. B. de Oliveira. Deciding difference logic in a Nelson-Oppen combination framework. Master's thesis, PPgSC/UFRN, 2007.

[5] D. C. B. de Oliveira, D. Dharbe, and P. Fontaine. Combining decision procedures by (model-)equality propagation. In *Brazilian Symposium on Formal Methods (SBMF2008)*, pages 51–66. Editora Grfica da UFBA, EDUFBA, 2008.

[6] D. Déharbe and S. Ranise. Bdd-driven first-order satisfiability procedures (extended version). Research report 4630, LORIA, 2002.

[7] N. Eén and N. Sörensson. An extensible SAT-solver. In *Theory and Applications of Satisfiability Testing*, volume 2919 of *LNCS*, pages 333–336. Springer, 2003.

[8] P. Fontaine, J.-Y. Marion, S. Merz, L. P. Nieto, and A. Tiu. Expressiveness + automation + soundness: Towards combining SMT solvers and interactive proof assistants. volume 3920 of *Lecture Notes in Computer Science*, pages 167–181. Springer-Verlag, 2006.

[9] T. Granlund. GNU MP: The GNU multiple precision arithmetic library, 4.1.4 ed, 2004. http://gmplib.org/.

[10] E. S. Marinho, V. de Medeiros Jr, D. Déharbe, B. Gomes, and C. Tavares. A ferramenta batcave para a verificação de especificações formais na notação b. In *XXII Simpósio*

Brasileiro de Engenharia de Software. XV Sessão de Ferramentas do SBES., pages 7–12.

[11] G. Nelson and D. Oppen. Simplification by cooperating decision procedures. *ACM Transactions on Programming Languages and Systems*, 1(2):245–257, 1979.

[12] G. Nelson and D. Oppen. Fast decision procedures based on congruence closure. *Journal of the ACM*, 27(2):356–364, 1980.

[13] S. Ranise and C. Tinelli. The SMT-LIB standard : Version 1.2, Aug. 2006.

[14] S. Schulz. System Description: E 0.81. In *Proc. of the 2nd IJCAR, Cork, Ireland*, volume 3097 of *LNAI*, pages 223–228. Springer, 2004.

[15] S. Schulz. System Description: E 0.81. In *Proc. of the 2nd IJCAR, Cork, Ireland*, volume 3097 of *LNAI*, pages 223–228. Springer, 2004.

[16] C. Weidenbach, R. Schmidt, T. Hillenbrand, R. Rusev, and D. Topic. System description: Spass version 3.0. In *Conference on Automated Deduction (CADE)*, volume 4603 of *LNCS*, pages 514–520. Springer, 2007.

Vérification automatique de systèmes à base de règles avec le narrowing stratégique

Anderson Santana de Oliveira[1], David Déharbe[2], Pierre-Etienne Moreau[3]

[1]Departamento de Ciências Exatas e Naturais – Universidade Federal Rural do Semi-Árido(UFERSA) – Mossoró – RN – Brésil

[2]Departmento de Informática e Matemática Aplicada– Universidade Federal do Rio Grande do Norte (UFRN) – Natal– RN – Brésil

[3]INRIA-Nancy-Grand Est— France

Résumé *Nous proposons la construction d'un outil pour la vérification automatique des buts d'atteignabilité sur des systèmes à base de règles contrôlés par une stratégie d'exécution explicite. La procédure de vérification est basée sur le narrowing, qui permet de simuler les exécutions possibles d'un ensemble de règles avec des variables. Le langage TOM sera utilisé pour le développent de cet outil, étant donné ces capacités de réécriture avec des stratégies.*

Abstract *We propose to build a tool aimed to the automated verification of reachability goals over rule-based systems controlled by strategies. The verification procedure is bases on narrowing, which allows one to simulate the possible executions of the rewrite rules with variables. The Tom language will be employed in the development of the tool, given its strategic rewriting capabilities.*

1. Introduction

La réécriture est un formalisme général et expressif pour la spécification et la vérification de divers types de systèmes axiomatisés par des règles de réécriture. La réécriture fournit un style de programmation déclaratif qui peut être utilisé dans la modélisation de la quasi-totalité des applications dans ce domaine grâce aux théories équationnelles et aux stratégies de réécriture ajoutées à la réécriture simple. Ces aspects apparaissent dans la vérification des programmes, mais aussi dans des domaines plus particuliers comme la modélisation de protocoles cryptographiques (Escobar, Meadows and Meseguer 2007) ou des politiques de sécurité (Dougherty, et al. 2007).

Étant donné qu'un ensemble de règles de réécriture définit un système de transitions (dont les états sont des ensembles de termes dans la même classe équationnelle) et que les règles établissent les transitions atomiques entre ces états, il est naturel de poser des questions d'atteignabilité sur ces systèmes. Par exemple, à partir d'un état initial t d'un système, est-il possible d'atteindre une configuration t' qui n'est pas considéré sûre selon une politique de sécurité donnée, c'est-à-dire, si $t \rightarrow^* t'$, quelles que soient les valeurs des variables de t?

À l'origine proposé comme une technique de résolution des buts dans des théories équationnelles ($\exists x.\ (t = t')$), le narrowing a été reconnu comme un mécanisme clé pour répondre à des questions d'atteignabilité, car il fournit les abstractions nécessaires pour simuler les exécutions d'un

ensemble de règles de réécriture. Son avantage par rapport aux techniques de model-cheking traditionnelles est de permettre la vérification des systèmes avec un nombre potentiellement infini d'états.

Notre but est donc de généraliser la technique de résolution basée sur le narrowing, suggéré initialement par (Escobar, Meseguer and Thati 2007) de façon à considérer des systèmes imposant une certaine stratégie d'application des règles. Ceci est important dans la spécification de modèles de sécurité assez courants, comme les systèmes de pare-feu (Kirchner, Kirchner and de Oliveira 2009), qui attribuent des priorités aux règles par exemple. Les stratégies ont été considérées seulement comme une façon d´accélérer la résolution des buts. Dans notre proposition, nous avons une contrainte additionnelle, qui est la stratégie adoptée par le système en train d'être analysé. De cette manière le narrowing stratégique est perçu comme un complément aux approches déjà existantes pour la vérification de modèles avec une infinité d'états (Clarke, Talupur and Veith 2008), telles que les tests d´atteignabilité basés sur les automates d'arbre (Boichut, et al. 2007), et l'approche originale basée sur le narrowing de (Escobar, Meseguer and Thati 2007).

Le produit principal de cette démarche sera un système pour la vérification automatique de programmes à base de règles avec des stratégies, plus particulièrement, nous sommes intéressés par les politiques de sécurité (Cirstea, Moreau and de Oliveira 2009). Ce système sera développé dans le langage Tom (Balland, et al. 2007), qui intègre la réécriture avec des stratégies dans des langages de programmation générales.

2. Le système Tom

Le système Tom fournit un moyen générique d'intégrer les signatures algébriques et le filtrage dans les langages de programmation existants, comme Java, C ou ML. Le langage Tom n'est pas conçu comme un langage à part entière. Rappelant le concept de Domain-Specific Language, sa conception repose sur l'idée d'« îlot formel » : un programme Tom est composé de constructions algébriques nouvelles, correspondant aux notions de signature, de filtrage, de règle, et de stratégie. Ces constructions, imagées par la notion d'îlot, sont de petits espaces isolés dans un ensemble d'une autre nature : le programme écrit dans le langage hôte. Il y a peu de restriction sur la nature du langage hôte et celui-ci peut sans difficulté correspondre aux langages C, C++, C#, Java, Eiffel, Python, Caml, etc.

Tom ajoute principalement deux nouvelles constructions : « `%match` » et « ` ` » (appelé backquote). La première est une extension de switch/case permettant de discriminer sur un terme plutôt que sur des valeurs atomiques comme des entiers ou des caractères. Les motifs sont utilisés pour discriminer et récupérer de l'information dans la structure de données algébrique filtrée. La seconde, inspirée du langage Lisp, permet de construire un terme.

Exemple 1 Un exemple élémentaire pour présenter Tom est la définition de l'addition sur les entiers de Peano, représentés par la constante `Zero()` et le successeur `Suc(Nat)`. En considérant Java comme langage hôte, l'addition peut être définie en Tom de la façon suivante :

```
public class Peano {
   public final static void main(String[] args) {
     Nat two = `Suc(Suc(Zero()));
```

```
      System.out.println("2 + 2 = " + plus(two,two));
}

  Nat plus(Nat t1, Nat t2) {
     %match(t1, t2) {
       x, Zero() -> { return `x; }
       x, Suc(y) -> { return `Suc(plus(x,y)); }
     }
  }
}
```

Dans cet exemple, nous distinguons bien les parties îlots introduites par `%match` et « ` », dont la portée correspond à un terme bien formé. Il est à noter qu'un îlot peut contenir des morceaux écrits en langage hôte, appelés « lacs », comme le sont les actions introduites par `-> { ... }`. Ces lacs, pouvant bien sûr contenir des îlots.

La construction « ` » est utilisée pour exprimer la construction de termes de manière algébrique. Elle permet d'utiliser les variables qui sont instanciées par filtrage (comme x et y), des constructeurs algébriques (`Suc`), et des fonctions du langage hôte, telles que plus dans cet exemple. La définition de plus est donnée par filtrage et correspond à une fonction Java, qui peut être utilisée de manière classique, partout ailleurs.

Afin d'être le plus fidèle possible au concept d'îlot formel, le compilateur Tom ne parse pas les parties écrites en langage hôte. Comme l'illustre la Figure 1, seules les constructions ajoutées sont examinées pour être traduites en des instructions du langage hôte. Cette étape de « dissolution » est faite in situ : la traduction se fait en lieu et place des constructions reconnues, sans modifier ni déplacer le code hôte. Cette caractéristique est essentielle pour permettre le déboguage des applications ainsi produites.

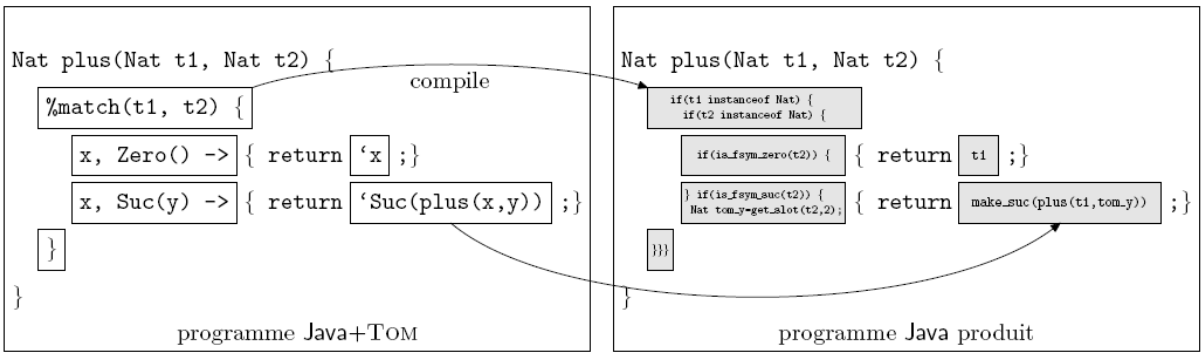


Figure 1 - Compilation de Tom

La troisième contribution, essentielle dans l'environnement Tom, consiste à définir un ensemble minimal d'opérateurs de stratégie, dont la combinaison permet de décrire des parcours et des transformations arbitraires. Le langage proposé autorise des définitions récursives et rend explicite la notion de position dans un terme, permettant ainsi de décrire des stratégies d'exploration non-déterministe. Les stratégies sont réifiées au niveau des termes, permettant de filtrer, de construire, et même de transformer dynamiquement une stratégie. Enfin, une stratégie étant un terme, il est possible d'appliquer une stratégie sur une stratégie. L'ensemble est implanté et intégré à Tom, ce qui rend le langage de stratégie facilement utilisable en présence des toutes les constructions de Tom.

3. Un moteur de narrowing stratégique

Dans toute sa généralité, le narrowing est une méthode de résolution de buts d´atteignabilité correcte mais pas forcément complète. Par contre, dans (Escobar, Meadows and Meseguer 2007) ont montré que le narrowing est une procédure complète dans des cas avec des intérêts assez pratiques, par exemple quand nous cherchons des solutions normalisées.

L'efficacité du narrowing a été améliorée avec l´utilisation de certaines stratégies, surtout dans le sens de trouver des séquences de dérivations optimales où seulement les positions strictement nécessaires sont rétrécies. En général, les résultats de complétude demandent des restrictions assez fortes sur la forme des règles. Par contre, dans (Escobar, Meadows and Meseguer 2007) les auteurs ont montré que ces restrictions peuvent être relaxées dans le cadre de l'analyse d'atteignabilité, car en général nous sommes intéressés par des formes normales seulement.

Nous proposons le développent d'un outil de vérification en utilisant le système TOM. Le système consistera dans un moteur capable d'exécuter des pas de narrowing selon une stratégie donné. Le langage de stratégie de Tom est suffisamment extensible pour comporter la définition de nouvelles situations typiquement issues du narrowing, comme des stratégies pour rendre le processus plus efficace. Néanmoins, ce processus sera contraint par la stratégie sous-jacente au système sous analyse. Nous devrons donc étudier des méthodes pour déterminer quand une stratégie est une sous-stratégie contenant des pas de réécriture strictement compris parmi les dérivations d'une stratégie dominante.

Le développement d'un tel outil devra comprendre la construction d'une interface avec l´utilisateur pour rendre facile l'introduction du modèle à être vérifié aussi bien que les propriétés qui doivent être validés. Tom est bien adapté pour le traitement d'un langage d'entrée que pour la conception de ce moteur de narrowing, étant donné qu'il intègre la réécriture et un langage de stratégie.

La présente proposition fera partie d'une demande de financement en réponse à l'appel Capes-Coffecub qui comprendra la collaboration entre les équipes Pareo et Mosel, de l'INRIA Nancy-Grand-Est et de l'UFRN et de l'UFERSA.

5. Références

Balland, Emilie, Paul Brauner, Radu Kopetz, Pierre-Etienne Moreau, et Antoine Reilles. «Tom: Piggybacking Rewriting on Java.» 2007. 36-47. «Term Rewriting and Applications, 18th International Conference, RTA 2007, Paris, France, June 26-28, 2007, Proceedings.» Springer, 2007.

Boichut, Yohan, Thomas Genet, Thomas P. Jensen, et Luka Le Roux. «Rewriting Approximations for Fast Prototyping of Static Analyzers.» 2007. 48-62. «Term Rewriting and Applications, 18th International Conference, RTA 2007, Paris, France, June 26-28, 2007, Proceedings.» Springer, 2007.

Cirstea, Horatiu, Pierre-Etienne Moreau, et Anderson Santana de Oliveira. «Rewrite Based Specification of Access Control Policies.» *Electronic Notes in Theoretical Computer Science* 234 (2009): 37-54.

Clarke, Edmund M., Muralidhar Talupur, et Helmut Veith. «Proving Ptolemy Right: The Environment Abstraction Framework for Model Checking Concurrent Systems.» 2008. 33-47. «Tools and Algorithms for the Construction and Analysis of Systems, 14th International Conference, TACAS 2008, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2008, Budapest, Hungary, March 29-April 6, 2008. Proceedings.» Springer, 2008.

Dougherty, Daniel J., Claude Kirchner, Hélène Kirchner, et Anderson Santana de Oliveira. «Modular Access Control Via Strategic Rewriting.» 2007. 578-593. «Computer Security - ESORICS 2007, 12th European Symposium On Research In Computer Security, Dresden, Germany, September 24-26, 2007, Proceedings.» Springer, 2007.

Escobar, Santiago, Catherine Meadows, et José Meseguer. «Equational Cryptographic Reasoning in the Maude-NRL Protocol Analyzer.» *Electr. Notes Theor. Comput. Sci.* 171 (2007): 23-36.

Escobar, Santiago, José Meseguer, et Prasanna Thati. «Narrowing and Rewriting Logic: from Foundations to Applications.» *Electr. Notes Theor. Comput. Sci.* 177 (2007): 5-33.

Kirchner, Claude, Hélène Kirchner, et Anderson Santana de Oliveira. «Analysis of Rewrite-Based Access Control Policies.» *Electronic Notes in Theoretical Computer Science* 234 (2009): 55-75.

Components for Rule-Based Constraint Programming in the Large

Pierre Deransart, François Fages[1]**,**
Jacques Robin [2]

[1]Institut de Recherche en Informatique et Automatique (INRIA)
CR de Paris Rocquencourt

[2]Universidade Federal de Pernambuco (UFPe)
Campus da UFPe

{pierre.deransart,francois.fages}@inria.fr, robin.jacques@gmail.com

Abstract. *In this paper we present how constraint based applications may benefit from software engineering modeling approaches. Extension of constraint handling rules are currently developed as a general rule based constraint programming environment including constraint solving, deduction, default reasoning, abduction, belief revision, and planning tasks and methods. But it still lacks general application modeling methods in order to be used in large scale software products. We show how the current collaborative project C4RBCP between UFPe and INRIA brings a contribution to this objective.*

1. Introduction

In this paper we show how Component-Based Software Engineering (CBSE) may give a boost to the use of Rule Based Constraints Programming (RBCP) for large software engineering applications. These two areas bring complementary principles to the general issue of software reuse for Automated Reasoning (AR).

Constraint Logic Programming supports a great ambition for programming: the one of making of programming essentially a modeling task, with equations, constraints and logical formulas. Rule-Based Constraint Programming (RBCP) is the combination of rule systems and constraint resolution. There are huge repositories of global constraints [Beldiceanu et al. 2008], including constraints with rules. [Fages and Martin 2008] introduce a general purpose rule-based modeling language for constraint programming, named Rules2CP which allows in particular to express search strategies and heuristics as preference orderings on variables, values, and *and/or* formulae.

Another paradigm is Constraint Handling Rules (CHR) originally developed by [Fruehwirth and Abdennadher 2003]. Several extensions of constraint handling rules are currently developed as a general rule based constraint programming environment including constraint solving, deduction, default reasoning, abduction, belief revision, and planning tasks and methods. Despite of these many approaches and improvements of modeling with constraints, constraints programming is not used as much as it could be. It still lacks of general modeling methods in order be effectively used for modeling very large software applications.

Component-Based Software Engineering (CBSE) [Atkinson et al. 2002] brings the idea of developing any software in two largely orthogonal steps. The first, design

for reuse step identifies services recurrently needed in a variety of contexts and encapsulate them into components that clearly separate the hidden services realizations from their public specifications as a provided interface. Other components can use the services provided by this interfaces by connecting to their associated ports. This offers them the possibility to outsource by design part of their own provided services realizations. This part must be publicly specified as required interfaces with associated ports. The second, design with reuse step of CBSE can then assemble at low-cost a variety of complex services by just connecting component ports with matching required and provided interfaces.

These CBSE concepts of components, interfaces, ports and assembly are currently only supported by imperative programming platforms (whether object-oriented or not) such as .Net or EJB. They are not supported by RBPC platforms, making constraint solving rule bases monolithic. Our project is to investigate how these CBSE concepts can be adapted to the RBPC language CHR^V (Constraint Handling Rules with Disjunctive bodies), a simple yet powerful extension of CHR, and implemented in the adaptive CHR^V engine CHROME (CHR Online Model-driven Engine) currently under development at CIn-UFPE.

After presenting some aspects of constraint programming and of CBSE, we present some results and perspectives resulting from a collaboration between the INRIA research team “Constraints” and the UFPe-CIN “SE” team.

2. Constraints Programming

Constraint Programming is a field born during the mid 70s from Logic Programming [Kowalski 1974, Colmerauer and Roussel 1996], Linear Programming coming from Operations Research [Schrijver 1998], and Constraint Propagation techniques coming from Artificial Intelligence [Marriot and Stuckey 1998, Rossi et al. 2006]. Its foundation is the use of relations on mathematical variables to compute with partial information. The successes of Constraint Programming for solving combinatorial optimization problems, from pure problems to real problems in industry or commerce, owe much to the bringing of, on the one hand, new local consistency techniques, and, on the other hand, declarative languages which allow control on the mixing of heterogeneous resolution techniques: numerical, symbolic, deductive and heuristic.

Several constraint solvers paradigms have been developed, in particular the Constraint Handling Rule (CHR) approach [Fruehwirth and Abdennadher 2003] which combines in an elegant and effective way the declarativeness of logic programming and the flexibility and empowerment of combining hybrid solvers. In particular CHR offers a good base to specify and implement Rule-Based Constraint Programming (RBCP). CHR has been proven successful in over 100 applications (see `WebCHR: http://www.cs.kuleuven.be/ dtai/projects/CHR/`).

The INRIA “Constraints” team is developing SiLCC (Linear Concurrent Constraint programming) an imperative and concurrent constraint programming language based on a single paradigm: the one of Vijay Saraswat’s concurrent constraint programming extended with constraint systems based on Jean-Yves Girard’s Linear Logic. In the late 90’s he developed the theory of this extension and is now working on its implementation. The LCC paradigm offers at the same time a theoretical framework for analysis, and a valuable guide for practical language design and implementation.

The UFPe group worked on an extension of CHR which shares many similarities with SiLCC, called CHR^V. It is a versatile RBCP language that can serve as a unifying basis to represent constraint solving, deduction, default reasoning, abduction, belief revision, and planning tasks and methods.

In parallel, INRIA contributed to develop tools for constraint solvers behavior analysis, in particular by the OADymPPaC project [Deransart & al 2004] (2001-2004). The recent progresses made on the theory and implementation of traces in CP (definition of the generic trace format GenTra4CP) put it into a firm ground for developing tracers for rule based systems. The figure 1 illustrates variable domain reduction during a 40 queens puzzle resolution.

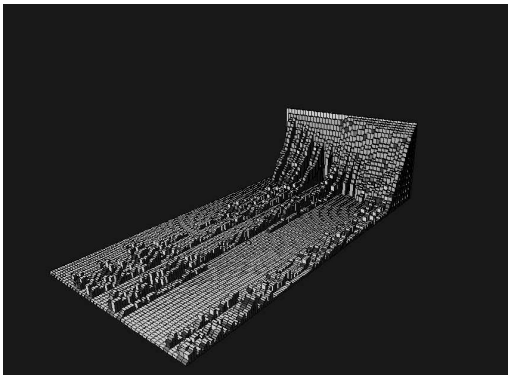


Figure 1. An example of visualisation of constraint propagation (40 queens problem, OADymPPaC project)

3. Component-Based Software Engineering

Components are the leading cost-cutting software reuse technology. A component encapsulates the realization of a service set and makes them accessible from outside through a provided interface with its access port. It may delegate part of this realization to external components by specifying them as required interfaces, each one with its access port. Applications and larger component realizations can be quickly assembled from smaller components by connecting the required ports of client components to the provided ports of server components with matching associated interfaces. In this regard, type systems [Russell and Norvig 2003] provide a first way to ensure some level of consistency in the assembling. Furthermore, each operation of a component's provided and required interfaces can be specified in detail using pre and post conditions. By providing detail service specifications while completely hiding their realizations, components with matching specifications but distinct realizations can be plugged out and in without affecting the rest of the assembly.

A component can possess a publicly visible state lifecycle: only a subset of its provided interface operations can be called in a given state and these calls can in turn alter this state. Each component in an assembly can follow its own independent thread, using its ports as asynchronous communication channels with the other components. The KobrA method [Gross 2004] prescribes how to leverage UML to develop component-based artifacts throughout the entire software lifecycle from requirements, to design and testing.

The .Net and J2EE platforms provide sophisticated component assembly implementation, deployment facilities.

The long term topic of our project is thus the introduction of components to the most practical and challenging variety of RBCP: adaptive solving that integrates constraint simplification, propagation and search, and provides an insightful visual solving trace to ease debugging and evolution of constraint programs.

4. Recent results

CHR^V has emerged as a versatile knowledge representation language, usable for an unparalleled variety of automated reasoning tasks: constraint solving, optimization, classification, subsumption, classical deduction, abduction, truth-maintenance, belief revision, belief update and planning. [da Silva et al. 2008] add default reasoning to this list, by showing how to represent default logic theories in CHR^V. It is also discuss how to leverage this representation together with the well-know correspondence between default logic and Negation As Failure (NAF) in logic programming, to propose an extension CHR$^{V;naf}$ of CHR^V allowing NAF in the rule heads.

[Fages et al. 2008] introduce a modular version of the Constraint Handling Rules language CHR, called CHRat for "modular CHR with ask and tell". Any constraint defined in a CHRat component can be reused both in rules and guards in another CHRat component to define new constraint solvers. Unlike previous work on modular CHR, this approach is completely general as it does not rely on an automatic derivation of conditions for checking entailment in guards, but on a programming discipline for defining both satisfiability (tell) and entailment (ask) checks by CHRat rules for each constraint. The operational and declarative semantics of CHRat are defined, a transformation of CHRat components to flat CHR programs is provided, and the preservation of the semantics is proven. They give examples of the modularization of classical CHR constraint solvers and of the definition of complex constraint solvers in a modular fashion.

5. Conclusion

INRIA and UFPe teams have started a regular cooperation since 3 years in particular through the C4RBCP project, founded by FACEPE, INRIA and MAEE French Ministry. Our cooperation proved to be extremely fruitful supported by complementary objectives.

Both teams share a common interest and previous researches in RBCP and CHR. However, they bring complementary expertise to different aspects of the issues to investigate in the proposed project. The INRIA Rocquencourt team specific expertise fields in constraint programming are theoretical foundations, trace generation and visualization, as well as modules, and type systems, two precursors of software components. The CIn-UFPE specific expertise fields are model-driven engineering , software architecture, component-based modeling, and object-oriented implementation of RBCP engines. This motivates common works.

Furthermore cross cooperation with other European country [Robin et al. 2007] suggests to look for European cooperation. This motivated the authors to elaborate a STREP proposal submitted in the FP7 ICT 2007-3.2.2 framework.

References

Atkinson, C., Bayer, J., Bunse, C., Kamsties, E., Laitenberger, O., Laqua, R., Muthig, D., Paech, B., Wust, J., and Zettel, J. (2002). *Component-based Product Line Engineering with UML*. Addison-Wesley.

Beldiceanu, N., Carlsson, M., and Rampon, J.-X. (2008). Global constraint catalog.

Colmerauer, A. and Roussel, P. (1996). The birth of Prolog. In Bergin, T. J. and Gibson, R. G., editors, *History of Programming Languages*. ACM Press/Addison-Wesley. 1992 draft: http://alain.colmerauer.free.fr/ArchivesPublications/HistoireProlog/19november92.pdf.

da Silva, M. A. A., Fages, F., and Robin, J. (2008). Default reasoning in chr^{v}. In *Proceedings of the 5th Workshop on Constraint Handling Rules (CHR'08)*, Hagenberg, Austria.

Deransart & al, P. (2004). Outils d'Analyse Dynamique Pour la Programmation Par Contraintes (OADymPPaC). Technical report, Inria Rocquencourt and École des Mines de Nantes and INSA de Rennes and Université d'Orléans and Cosytec and ILOG. Projet RNTL. `http://contraintes.inria.fr/OADymPPaC`.

Fages, F., de Oliveira Rodrigues, C. M., and Martinez, T. (2008). Modular chr with ask and tell. In *Proceedings of the 5th Workshop on Constraint Handling Rules (CHR'08)*, pages 95–110, Hagenberg, Austria.

Fages, F. and Martin, J. (2008). From rules to constraint programs with the Rules2CP modelling language. INRIA Research Report RR-6495, Institut National de Recherche en Informatique.

Fruehwirth, T. and Abdennadher, S. (2003). *Essentials of constraint programming*. Springer Verlag.

Gross, H. (2004). *Component-based Product Line Engineering with UML*. Springer Verlag.

Kowalski, R. A. (1974). Predicate logic as programming language. In *IFIP Congress*, pages 569–574.

Marriot, K. and Stuckey, P. J. (1998). *Programming with constraints: An introduction*. The MIT Press.

Robin, J., Vitorino, J., and Wolf, A. (2007). Constraint programming architectures: Review and a new proposal. In *11th Brazilian Symposium on Programming Languages (SBLP'07)*, Natal, RN, Brazil.

Rossi, F., Beek, P. V., and Hardbound, T. W. (2006). *Handbook of constraint programming*. Elsevier.

Russell, S. and Norvig (2003). *Artificial Intelligence: A Modern Approach (2nd Ed.)*. Prentice-Hall.

Schrijver, A. (1998). *Theory of Linear and Integer Programming*. John Wiley & sons.

Visualization of Parallel Applications: Results of an International Collaboration

Lucas Mello Schnorr[1,2], **Guillaume Huard**[2], **Philippe O. A. Navaux**[1]

[1] Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

[2]INRIA Moais research team, CNRS LIG Laboratory, Grenoble University
ZIRST 51, avenue Jean Kuntzmann - 38330 - Montbonnot Saint Martin - France

{lmschnorr,navaux}@inf.ufrgs.br, Guillaume.Huard@imag.fr

Abstract. *The performance visualization of parallel applications is a necessary step for a better comprehension of application behavior and its correspondence with the execution environment. In the scope of grid environments, there are new challenges for the visualization of parallel applications, mainly related to the visual observation of the network influence and the scalability of the visualization techniques. We present in this paper two proposals that address these problems. We also present the details of the international collaboration that allowed the effective development of this work.*

1. Introduction

The visualization analysis of parallel applications appeared almost at the same time the first graphical screens were available. As parallel and distributed systems evolved, so did the visualization techniques. Today, there are considerable challenges to the visual analysis of parallel applications, especially when we consider grid environments characteristics and their influences over applications. Grid platforms, such as the Grid'5000 [Capello et al. 2007], have highly hierarchical network interconnections that might influence parallel applications behavior. Current visualization techniques are not able to properly identify problems of applications that are caused by the network. Another characteristic of Grids is their size, making the development of larger applications possible. In this context, traditional visualization techniques are not able to visualize applications with thousands or even millions of processes.

We show here two proposals to address these issues. The first one is focused on the network influence over parallel applications, where our proposal resides in a new three dimensional visualization technique that is able to show the network topology and its relation with the application's components behavior. The second one uses an information visualization technique called Treemap [Johnson and Shneiderman 1991] to the analysis of parallel applications with hierarchical organization of monitoring data. This second approach attacks the problem of visualization scalability present in most of available visualization tools.

The thesis of the first author is developed within the scope of a collaboration between the Federal University of Rio Grande do Sul (UFRGS), located at Porto Alegre, Brazil, and the Grenoble Institute of Technology (INPG), in France. The collaboration had the financial support by the CAPES/Cofecub Project 4602/06-4, with

an established *co-tutelle* aggreement between the two universities. During 18 months, the first author was located in France at the INRIA MOAIS Project Team, part of the CNRS Lig Laboratory. The main results of the joint work were published in the IEEE/ACM Grid'2008 [Schnorr et al. 2008] and the IEEE/ACM CCGrid'2009 conferences [Schnorr et al. 2009]. The collaborations between Brazil and Grenoble in the performance visualization area started in the middle of the 90's, with the thesis of Benhur de Oliveira Stein, advised by Jacques Chassin de Kergommeaux. The current collaboration is an effort in keeping joint works in this area.

The rest of the paper is divided in 4 sections. Section 2 briefly lists related work. Section 3 presents the proposal of an abstract model for the two new visualization techniques. Section 4 presents the implementation of the model in the prototype *3va*. We end the paper with a conclusion.

2. Contextualization

Several visualization tools were developed in the last 25 years to the analysis of parallel applications. A relevant set of tools is composed of ParaGraph (1990), Paradyn (1995), Vampir (1996), Pajé [Stein et al. 2000] and ParaProf (2003). Most of these tools are focused in the analysis of parallel applications executed in homogeneous and controlled environments, such as clusters. If we consider their visualization techniques in grid platforms, we observe the lack of support for important characteristics of grids that might influence the analysis. Next section details these problems and our approach to them.

3. Visualization Techniques Developed

Parallel application analysis, when applied to distributed environments such as grids, poses a new set of problems to be solved. These problems are related to how the grid is organize and its characteristics, such as the heterogeneity of hardware and software, dynamism of resources and larger size. The techniques developed to visualize parallel applications behavior were mostly focused in cluster-like environments, viewed as a more controlled and homogeneous environment.

Considering this context of the need of visualization techniques for the analysis of grid applications, we isolated two key problems. The first one is related to the importance of analysing applications taking into account network characteristics. The topology, bandwidth and latency of the network might have an important influence over the behavior of the programs. This happens especially in network-bound applications, where the amount of data being transferred or the need for low-latency communications is determinant. By analysing related work, we can observe that none of existing visualization techniques for application analysis show in a clear form the correlation between application and network use. The only tool that had in some way this analysis was ParaGraph, but it was built only for small-scale parallel applications.

The second problem is related to the size of grid parallel applications and the visualization techniques used for their analysis. Grids might scale up to thousands and possibly millions of resources. The Grid'5000 platform [Capello et al. 2007], for instance, was conceived to be composed of 5000 processing cores. Even if this project limitation exists, the addition of new resources can continue indefinitely. Therefore, the potential utilization of these resources increases resulting in larger applications that might attain

millions of processes. Most of the visualization techniques that attack the problem of analysing applications of this size work by grouping, reducing and filtering out processes behavior to later plot them using classical visualization techniques.

With the main goal of trying to solve these problems, we have proposed two visualization techniques. The first one, focused in the problem of the network influence over the application, resulted in the proposal of a three dimensional visualization technique that uses two dimensions to render a network topology and the third dimension as timeline [Schnorr et al. 2008]. This allows developers to be able to track up when the limitations of the network influence the application. The second visualization technique proposed is focused in the problem of analysing large-scale applications. Our proposal in this case uses an information visualization technique called Treemap [Johnson and Shneiderman 1991], combined with a hierarchical organization of monitoring data, to depict the behavior of processes in different time intervals [Schnorr et al. 2009].

Figure 1 depicts the proposed component model used to generate the new visualizations techniques for parallel applications. The input of the model is depicted at right, composed by the traces from the parallel applications and the resources description. The two outputs are at right, as a form of a 3D Visualization (bottom right) or a Treemap View (top right). The flow of data occurs from left to right, mainly influenced by new monitoring data arriving or changes in the resources description. Configurations and interaction mechanisms act from the visualizations to the middle components.

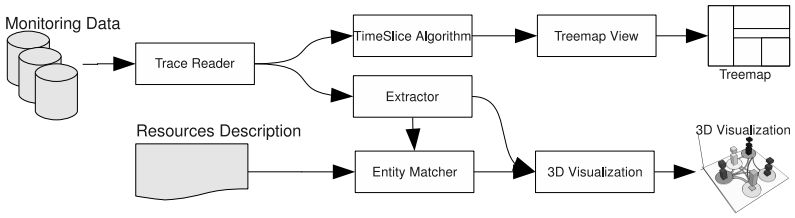


Figure 1. Model of the proposed solutions.

4. The *3va* Prototype

The model created for the proposed visualization techniques is implemented in a prototype called *3va*, that stands for Three Dimensional Visualization Analysis. The prototype is developed in the Objective-C and C++ languages, with each component implemented separately. The trace reader was implemented using the API defined by the DIMVisual library [Schnorr et al. 2006], and the Pajé Simulator [Stein et al. 2000]. The two visualizations components (Treemap and 3D), were implemented from scratch using the Wxwidgets and Ogre3D libraries. The components that implemented most of the algorithms proposed in the model (TimeSlice Algorith, Extractor and Entity Matcher) were fully implemented using the Objective-C language. Two resources description were used in this work: network topology and hierarchical structure. The monitoring data input was mostly composed of traces from parallel applications developed with the KAAPI library [Gautier et al. 2007], although other types of traces, such as from MPI applications, can be easily adapted to the be used within *3va*. Figure 2 shows the 3D visualization generated by the prototype.

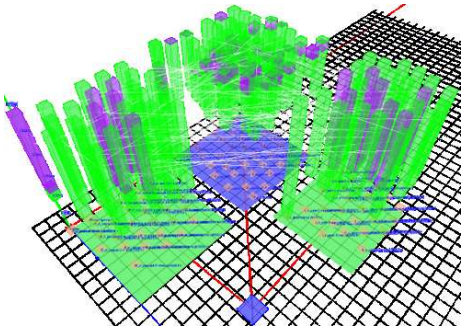


Figure 2. 3D visualization generated by the prototype *3va*.

5. Conclusion

Grid applications have particular characteristics that current visualization tools are not able to show. This paper focused on two of them: the visual analysis of the network topology influence over the parallel applications and the visualization scalability. Our approach to these issues is to use a three dimensional visualization to be able to show network topology and application behavior together; and to use hierarchical organization of monitoring data with the treemap technique to be able to analyze larger applications. We have developed a prototype that implements these approaches and generated results with the analysis of KAAPI applications executed in the Grid'5000 platform. As future work, we intend to refine the mechanisms used to create the visualizations to turn more easy the analysis. This work is developed within the scope of a collaboration Brazil/France, funded by a CAPES/Cofecub project. We support the continuity of this type of collaborations because of the positive aspects of this international environment to research and the know-how acquired over the years by both sides in the performance visualization area.

References

Capello, F., Plateau, B., Priol, T., Vandromme, D., Cappello, F., Dayde, M., Desprez, F., Jeannot, E., Jegou, Y., Lanteri, S., Melab, N., Namyst, R., Quinson, M., Richard, O., and Primet, P. V.-B. (2007). Grid5000. http://www.grid5000.fr. Last accessed April, 24, 2009.

Gautier, T., Besseron, X., and Pigeon, L. (2007). Kaapi: A thread scheduling runtime system for data flow computations on cluster of multi-processors. In *Proceedings of the international workshop on Parallel symbolic computation*, pages 15–23, New York, NY, USA. ACM.

Johnson, B. and Shneiderman, B. (1991). *Tree-Maps: a space-filling approach to the visualization of hierarchical information structures*. IEEE Computer Society Press Los Alamitos, CA, USA.

Schnorr, L. M., Huard, G., and Navaux, P. O. A. (2008). 3d approach to the visualization of parallel applications and grid monitoring information. In *Proceedings of the 9th IEEE/ACM International Conference on Grid Computing*, Washington, DC, USA. IEEE Computer Society.

Schnorr, L. M., Huard, G., and Navaux, P. O. A. (2009). Towards visualization scalability through time intervals and hierarchical organization of monitoring data. In *The 9th IEEE International Symposium on Cluster Computing and the Grid, CCGRID*. IEEE Computer Society.

Schnorr, L. M., Navaux, P. O. A., and de Oliveira Stein, B. (2006). Dimvisual: Data integration model for visualization of parallel programs behavior. In *Proceedings of the Sixth IEEE International Symposium on Cluster Computing and the Grid*, pages 473–480, Washington, DC, USA. IEEE Computer Society.

Stein, B., de Kergommeaux, J. C., and Bernard, P.-E. (2000). Pajé, an interactive visualization tool for tuning multi-threaded parallel applications. *Parallel Computing*, 26:1253–1274.

RobustWeb: Development and Validation of SOA Applications based on *Web Services*

Eliane Martins[1], Ana Cavalli[2], Jean Arlat[3], Regina Moraes[1], Taisy Weber[4]

[1]Universidade Estadual de Campinas - UNICAMP
[2]TELECOM & Management SudParis
[3] LAAS-CNRS, Université de Toulouse
[4]Universidade Federal do Rio Grande do Sul - UFRGS

eliane@ic.unicamp.br, ana.cavalli@it-sudparis.eu, jean.arlat@laas.fr, regina@ceset.unicamp.br,taisy@inf.ufrgs.br

Abstract. *RobustWeb is a cooperation project joining Brazilian and French researchers to develop methods and tools that aim reaching Web Services robustness. The project encompasses seven institutions in both countries. All partners have long experience in the complete life cycle of software engineering, which constitute a solid basis to reach relevant results to face important challenges in the area of service-oriented architecture.*

1. Introduction

Service-orientation is an emerging paradigm for the development of distributed and e-business applications, in which the services (autonomous platform-independent computational elements that can be described, published, discovered and accessed over the Internet using standard protocols) are the new elements for development and reuse. This way of structuring an application in a set of interacting services is referred as Service-oriented Architectures (SOA).

In recent years, various approaches for implementing service-oriented architectures have appeared, and the majority of them uses Web Services (WS) as underlying technology for integration between different software applications. Web Service is a client-server system specially structured to make the best use as possible of Web standards, like XML for example. Today, different software vendors are launching their own solutions, as for instance Microsoft with .Net platform, Sun with Sun One initiative, and IBM with its WSDL tool kit development and execution environment. Web Services standards and their implementations are important to enable SOA wide acceptance. Various initiatives exist in that sense, such as the Web Services Interoperability Organization (WS-I), which is an open, industry organization committed to promoting interoperability among Web Services based on common, industry-accepted definitions and standards. Another one is the Organization for the Advancement of Structured Information Standards (OASIS), which defined a reference model for SOA.

These initiatives represent an essential step in the widespread adoption of WS and SOA in practice. However, composing a high quality SOA for the development of business or mission critical applications still stands at an early stage. The sources of failures for highly dynamic, distributed applications are innumerous. Besides, security aspects in this distributed, collaborative context are also an issue. In this respect, it is

important to provide guidelines on how to use the available techniques for the design of reliable architectures to compose their services. For some examples of relevant research efforts aimed at coping with these issues, see [Salatgé & Fabre 2007, Abou El Kalam & Deswarte 2008, Mallouli 2008]. Validation is another important aspect. The investigation of how to use testing techniques to the context of WS and SOA, as well as the modeling and evaluation of performance in this continually changing context will also be useful to guarantee that the desired level of quality is assured and maintained in spite of changes. Some of these issues are studied in the following works: [Cavalli 2007], [Mallouli 2008].

SOA-based architectures with Web Services are thus promising for the development of *omnivalent systems*, which are systems which present characteristics such as ubiquity, dependability and security. According to the document entitled *Grand Research Challenges in Computer Science in Brazil*, prepared in 2006 by the Brazilian Computer Society (www.sbc.org.br), these systems represent one of the research challenges nowadays. Developing and maintaining this kind of system represents a challenge in that combining the aforementioned properties is far from being easy. It has also been mentioned in [Laprie 2005], that dependability and security drop significantly for these emerging systems, compared with the traditional ones. On the other hand, the cost of failures in services is growing, given the increase in the presence of computer systems in everyone's lives. System development, and in particular SOA-based applications, has to deal with two main, sometimes contradictory, driving forces: increasing complexity and reduced time-to-market. Services reuse addresses the latter, but does not avoid the former. On the contrary, the complexity tends to increase, due to heterogeneity and the geographically distributed development and operation of Web Services based applications. These factors lead to more residual development defects, vulnerabilities and interaction mistakes, which represent threats to dependability and security of the applications [Laprie 2005].

The project *RobustWeb*, granted by CAPES/Cofecub, addresses this challenge. With this objective in mind, the project is aimed at covering the following issues presented by SBC:

- To integrate researchers into a multidisciplinary group that encompasses diverse research domains such as Dependability and Security Design and Validation, in this way covering not only the aspects relative to the construction of reliable and secure systems but also, those relative to the determination of whether the desired level of quality was achieved, which encompasses functional testing and robustness assessment activities, as well as performance evaluation and risk assessment.
- To promote meetings among the researchers, in one hand, to identify key research directions in the development and validation of reliable and secure SOA-based applications with Web Services. On the other hand, to organize workshops to help in propagate the best practices among practitioners and scientists.
- To increase the cooperation between Brazilian and French researchers, by bringing to Brazil visiting scholars that are active in the different domains covered by this project.

2. Partner institutions

The expertise needed to carry out the project is well covered by the gathered teams. All partners have long experience in the complete life cycle of software engineering: from fault-tolerant architectural modeling and formal specifications, to test specification and execution for applications, services and protocols. This includes timed aspects, automatic test generation, fault injection and building software engineering tools. More precisely, the main experience of each partner is:

- TELECOM & Management SudParis: methods and tools for test generation and monitoring, tool prototyping, Web Services technologies, web protocols and protocol architectures, engineering and standardization;
- LRI: random testing, statistical testing, modeling of Web Services; Web Services technologies including their standardization;
- LAAS-CNRS: analytical modeling and evaluation of fault tolerance, software testing, probabilistic generation of test data, fault injection, fault tolerance of web services, security.
- Unicamp: exception handling mechanisms; fault tolerance of concurrent and distributed object-oriented systems; dependable software architecture and component-based development; model-based testing; fault injection; methods and tools for building and testing dependable systems, risk assessment.
- II-UFRGS: developing of fault injection tools, communication fault injection, fault modeling of distributed systems, Internet protocols and their validation, fault tolerance in distributed systems, generation of fault loads for test campaigns.
- UFES: analytical modeling and evaluation of fault tolerance; performance evaluation of large scale distributed systems, Internet and web services measurements;
- INPE: model-based testing, methods and tools for test execution of embedded systems, tools for test execution and organizational aspects of test cases designed and executed along with the software development.

3. Project goals

The general objective of this project is to put together researchers in different areas, mainly: application architecture, security, testing and performance evaluation, to create guidelines and frameworks to assist service users and providers in the design and validation of services as well as the applications built by composing these services using the standards and technologies available for Web Services development and use.

The guidelines and frameworks been developing are based on best practices for software development, adapted to the context of WS and SOA. The intent is to help practitioners in the development of robust, fault-tolerant architectures, and that the services and applications implemented satisfy their functional requirements, but also present the required level of robustness, security and performance. The good practices help in reducing the costs as well as in increasing the productivity in the development or acquisition of services, as well as in their composition. More specifically, the intent is to provide a road map for the adoption of WS-based SOA to mission or business critical

applications. To illustrate the methods and techniques proposed, we are using as case study a space application from INPE, the Spacecraft Monitoring and Control (SM&C) system, which refers to end-to-end services between functions, resident on-board a spacecraft or based on the ground, that are responsible for mission operations.

The outcomes from the project will be knowledge, publications, training and technology transfer. In addition, the creation of new joint projects is also intended.

- Knowledge created by the project will be disseminated through conferences and journals.
- Education, training and technology transfer will also be sustained ways for disseminating and exploiting the results.
- Case studies are useful by themselves, as they are a springboard for realistic application of better network engineering and validation methods. By creating 'how to' guides based on case studies, the project will smooth the adoption of best network engineering and validation practices in industry.

4. Technical challenges

The project addresses important technical challenges [Martins 2008]. The first one is the design of an architecture that can maintain the desired level of reliability, availability and performance in the service composition, even when the services change at runtime and the only information about them is the specification of its public interface. Two approaches are possible. The first one defines how to use techniques to construct customizable fault-tolerant connectors to add fault-tolerance to unreliable WS [Salatgé & Fabre 2007]. These connectors insert detection actions (e.g. runtime assertions) and recovery mechanisms (based on various replications strategies). The connectors can use identical or equivalent available service replicas. The other approach adapts techniques for fault-tolerant composition of components, which are static, to the context of WS.

The next challenge is to adapt currently available model-based testing techniques for the testing of continually changing services and dynamic, loosely coupled applications to ensure that the composite services work properly. Two testing techniques will be investigated. One is active testing, in which test cases are generated from a model representing system behavior [Cavalli 2004]. We require new testing architectures that can take into account the tests executed remotely, with a minimum interference with the operation of the service being tested. Also, the dynamic changes of services must be considered. Another technique is based on passive testing, or monitoring, in which the traces collected during execution are used to determine whether a service behaves as expected. One important aspect from which guidelines are to be given concerns how to express these properties in a dynamically changing context.

Concerning test so far, the problem of testing the robustness of services without interfering with the normal operation of these services must also be solved. This challenge raises some issues related to managing the fault handler events, such as raising and handling of exceptions. Two approaches will be considered. One is property-oriented, which uses the specification of a property to drive the testing process. The aim is to exercise a system to observe whether the property is violated or not [Abdellatif-Kaddour 2003]. The properties of interest are any high level requirement

related to the most critical failure modes of the system. Another approach is fault injection that consists in the deliberate introduction of faults into a system to observe its behavior. Here also a model-based testing approach is to be investigated, the model representing either the architecture or the behavior of the service or the application. In case of behavior-based testing, the derivation of suitable models will be useful to fulfill this objective [Ambrosio 2005]. In case of an architectural model, the objective is to perform a dependency analysis to guide fault injection. In this way, we can assess the failure tolerance of the interfaces regarding component failures and corruptions that may enter into system from external sources [Moraes 2005].

How to evaluate the risk of using a service in a given application in a dynamic, constantly evolving context is another important challenge. The composition of the WS may come up with the choice among components that provide the same functionality. In order to guide the selection, a risk assessment can be performed. Previously, assessment of component risk was proposed, so that the component that represents lower risk to the overall system should be selected [Moraes 2007]. The approach was also based on fault injection, in which internal faults (representing faults occurring during development) were introduced into the component. It is necessary to adapt such approach to the WS-based SOA world, in which the services can change dynamically and should be tested sometimes remotely, with no interference with its operation.

The last challenge the project addresses is how to evaluate whether the desired level of performance is guaranteed in a dynamic, continually evolving context. The performance evaluation of web services involves traditional techniques which can be combined in a unified approach. In particular, our approach relies on the combination of analytical-based models and large scale experimental-based evaluation. From the application designers' point of view, it is critical to understand how different components of the distributed infrastructure supporting the provided service might affect the service. Based on the experimental results, analytical models will be built in order to support such analysis.

5. Conclusion

The project promotes opportunities of knowledge exchange among the partners. Most of these exchanges are been realized remotely, with the support of the Internet. In addition some project themes correspond to PhD thesis advised by the researchers in France and in Brazil. Doctoral as well as post-doctoral missions are envisaged between France and Brazil, encouraging the participation of the students in the project activities and stimulating their mobility.

Other planned project activities are the technical visits among the different institutions. These visits are aimed to consolidate the partnership and also to give rise to new collaborations. These researchers will present tutorials and talks to graduate students of the visited institutions, promoting therefore the transfer of knowledge.

References

Abdellatif-Kaddour, O.; Thévenod-Fosse, P.; Waeselynck, H. (2003). "A stepwise strategy for property-oriented testing", LAAS report no. 02351, September 2002. In Proc. 18th ACM Symposium on Applied Computing (SAC'2003), Melbourne, Florida, USA.

Abou El Kalam, A. and Deswarte, Y. (2008). "Critical Infrastructures Security Modeling,

Enforcement and Runtime Checking", in Proc. 3rd Int. Workshop on Critical Information Infrastructures Security (CRITIS 2008), Rome, Italy, pp.115-128.

Ambrosio, A.M.; Martins, E.; Vijaykumar N. L.; Carvalho, S.V. (2005). "A Methodology for Designing Fault Injection Experiments as an Addition to Communication Systems Conformance Testing," In Proc. of the First Workshop on Dependable Software – Tools and Methods in the IEEE Conference on Dependable System and Network, 28 June – 1 July 2005, Yokohama, Japan.

Cavalli ,A.; Maag, S.; Papagiannaki, S.; Vergiakis, G.; Zaïdi, F. (2004). "A Testing Methodology for an Open Software E-Learning Platform". In 18th IFIP World Computer Congress - TC10/WG10.5 Edutech Worskshop, pages 165-174, Toulouse, France. Kluwer Academic Publishers, ISBN 1-4020-8161.

Cavalli, Ana; Maag, Stephane; Morales, Gerardo (2007). "Regression and Performance Testing of an e-learning Web application: dotLRN". In 3rd ACM / IEEE International Conference on Signal-Image Technology and Internet-Based Systems (SITIS'07), Shangai, China.

Laprie, J.-C. (2005). "Resilience for the Scalability of Dependability". In Proceedings of the Fourth IEEE International Symposium on Network Computing and Applications (NCA'05). Cambridge, MA, USA.

Mallouli, Wissam; Lallali, Mounir, Morales, Gerardo; Cavalli, Ana (2008). "Modeling and Testing Secure Web-Based Systems: Application to an Industrial Case Study". In The Fourth ACM/IEEE International Conference on Signal Image Technology and Internet Based Systems SITIS '08, pages 128-136, Bali, Indonesia.

Martins, E. ; Moraes, R. ; Weber, T. S. ; Rubira, C. M. F. ; Ambrosio, A. M. ; Martinello, M. (2008). Desafios na construção e validação da robustez de aplicações orientadas a serviços. In Seminário Integrado de Software e Hardware, XXXV, Belém, pp. 62-76.

Moraes, R.; Martins, E.; (2005). "Fault Injection Approach based on Architectural Dependencies". Architecting Dependable Systems III, Book Chapter, Lecture Notes in Computer Science, Springer-Verlag Berlin Heidelberg New York, pp. 300-321.

Moraes, R.; Durães, J.; Barbosa, R.; Martins, E.; Madeira, H. (2007). "Experimental Risk Assessment and Comparison using Software Fault Injection". In The International Conference on Dependable Systems and Networks - DSN 07, Edimburgo.

Salatgé, N.; Fabre, J.-C. (2007) "Fault Tolerance Connectors for Unreliable Web Services". In 37th Annual IEEE/IFIP Int. Conf. on Dependable Systems and Networks (DSN-2007), Edinburgh, UK, pp.51-60.

Particle-Based Simulation and Visualization of Tubular Flows

Tiago H. C. Nobrega[1], **Diego Dias Bispo Carvalho**[1], **Aldo von Wangenheim**[1]

[1]LAPIX – Laboratory for Image Processing and Computer Graphics
UFSC – Federal University of Santa Catarina
Florianópolis, Santa Catarina, Brazil

{tigarmo,diegodbc,awangenh}@inf.ufsc.br

Abstract. *In this paper a method to efficiently simulate fluid flows through tubular structures, such as pipes and blood vessels, is introduced. The structure's centerline acts as a guide in pushing the fluid, which is simulated through the Smoothed Particle Hydrodynamics method. We employed a color coded classification on the particles to show variations of physical quantities along the tubular length.*

1. Introduction

Liquid flow (hereafter referred to simply as "flow") can be applied in applications like CAD systems, video games and virtual surgery simulators. The method in this paper is concerned with interactivity and flexibility allowing its use in real time environments, depending on the model complexity and level of detail. The method aims to perform a simplified guided flow simulation through arbitrary tubular structures, employing a Lagrangian (particle-based [Müller et al. 2003]) method to simulate the fluid, the Smoothed Particle Hydrodynamics (SPH) method [Monaghan 2005].

This paper's contributions are two-fold: First, a new method to efficiently obtain the approximate centerline of tubular structures is briefly discussed, and then the same centerline is used to simulate guided flows inside arbitrary tubular triangle meshes with no connectivity information.

2. Method

Approximate Centerline

According to Jiang, a centerline is "a curve that traverses the center of a hollow organ" [Jiang and Gu 2005]. In this work, we use an updated version of [Carvalho et al. 2006] to obtain cross-sections of tubular objects. Given two user-positioned planes, the algorithm works by firing lines between the planes and checking the lines' intersection with the object. If an intersection is found, a new plane is placed at the line segment's midpoint, and the algorithm proceeds recursively (Figure 1). The centerline is built by linking the centroids of neighboring cross-sections with line segments.

Simplified Guided Flow Simulation

The SPH method's appeal relies on the possibility to work with discrete values of physical quantities that control the fluid's state for each particle and are interpolated with the aid of *Smoothing Kernels*, which are radial symmetric weighing functions. Müller [Müller et al. 2003] discussed a simplified version of the Navier-Stokes equation for the

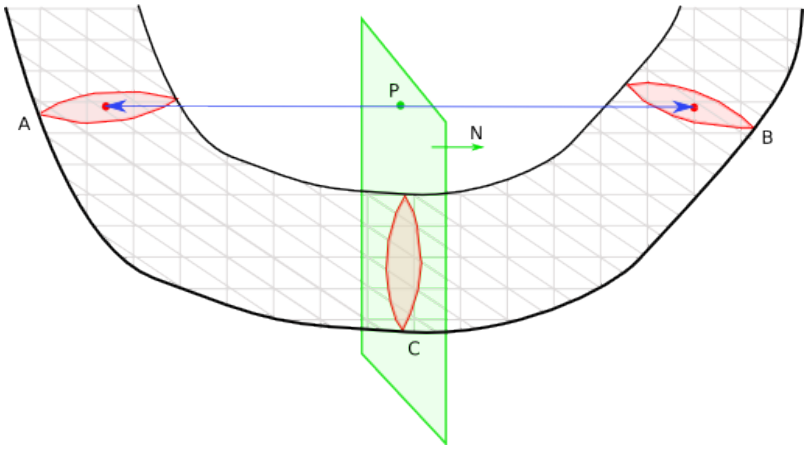


Figure 1. A line (in blue) is traced between the centroids of the two cross-sections *A* and *B*, resulting in a new plane with normal *N* which generates the new cross-section *C*.

conservation of momentum for incompressible fluids involving physical quantities such as density, pressure and viscosity which are represented by continuous fields with difficult analytical solutions. Müller applied the SPH method to solve the equation's right hand side to calculate the force acting on each particle. The simplified equation supports the application of an external force, in our case the centerline guided force.

External Centerline Force

The idea is to use the approximate centerline as a simplification of the tubular structure's shape to guide the fluid particles. The centerline line segments' direction vector is used to apply an external force that pushes the particle forward on its field of influence (Figure 2) — the particle is under the influence on the closest centerline segment. At the simulation's initial step each particle's "region" is identified and stored. At every simulation step, each particle trajectory is checked to confirm it has crossed any of the cross-sections. If a particle's trajectory crossed the cross-section S_i and if the centerline sets (the line segments L and the polygonal cross-section S) are ordered such that the first vertex of line L_i corresponds to the centroid of polygon S_i, then the new direction of the centerline force acting on this particle is the direction of line L_i.

Since the sets containing the cross-sections and the line segments are ordered, it is expected that a particle that crossed the cross-section S_i last will cross the polygon S_{i+1} next. In practice, due to the dynamics of the fluid it is possible that the particle be pushed back momentarily against the flow. Still, only the trajectory's intersections with S_{i+1} and S_i need to be checked, making the process very efficient. The result is a spatial subdivision in which a particle is always at a known region delimited by two cross-sections.

The last cross-section in the set S is a special case; as Figure 2 shows, if there are n line segments, there are $n + 1$ cross-sections. If a particle crosses the last cross-section it is not associated with any force vector, because there is no line segment to provide the direction. We decided to move the particle back to the beginning of the centerline in this case. The positive points about that are that the fluid is constrained to the region of interest on which the centerline was built and that the particles are reused avoiding the need to track, destroy and create new ones.

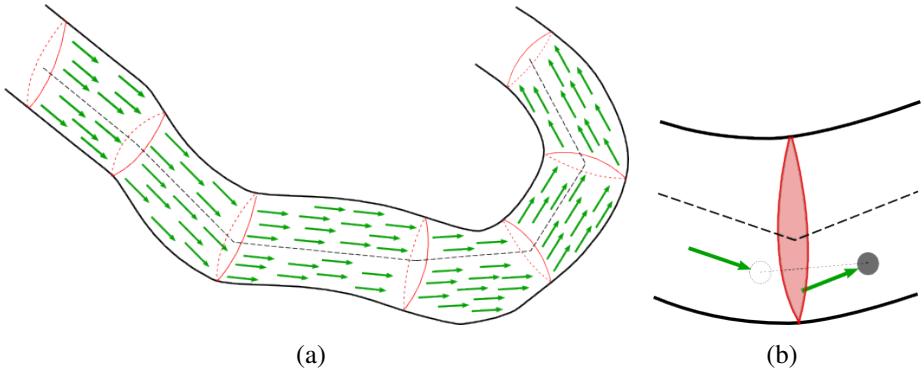


Figure 2. Particles suffer the influence of the line segment in the centerline region they're in. The force changes direction once the particles passes through a cross-section.

Although the set of line segments and cross-sections are only an approximation of the tubular structure's shape, the actual resolution of the mesh is irrelevant to the force computation once the centerline has been generated, as only the cross-sections and the line segments are used.

3. Visualization Experiments

In this section we present the results of visualization and simulation experiments performed on common desktop-grade hardware. The centerline algorithm was applied in a polygonal mesh representing an artery reconstructed by the marching cubes algorithm. This segment is interesting from a simulation point of view because it presents a few anomalies, such as a stenosis-like narrowing and an aneurysm-like bulge.

Figure 3(a) shows three different representations of the same frame of the simulation near the bulge area, with the fluid flowing from the structure's right to its left. The left-most figure displays all particles rendered with a solid blue color. On the middle image, the particles are color-coded according to their *pressure* value. Particles that fell in the bulge get "squeezed" together and are red-colored, representing an area of high fluid pressure. The traffic of particles near that region has a direct impact on their *velocities*, as shown in the final image. Particles in the bulge are almost stationary and blue-colored, as they cannot get out. The flow of the rightmost particles is red and fast because of the lack of collision obstacles and the higher degree of freedom to move, but as the particles near the trouble bulge-area their speed slows down considerably.

Figure 3(b) is a different case. Each image shows the state of the simulation of a small amount of fluid in different points in time. The centerline force is being exerted from the right to the left of the structure. On the first image, the particles have just arrived at the beginning of the uphill inclination — the particles that hit the artery's wall directly are under heavier pressure. The middle image shows the centerline force acting on pushing the particles upwards, towards the stenosis-like narrowing. This narrowing limits the intensity of the flow so that the particles begin to accumulate at the narrow opening's location. Thus, the pressure on the region increases.

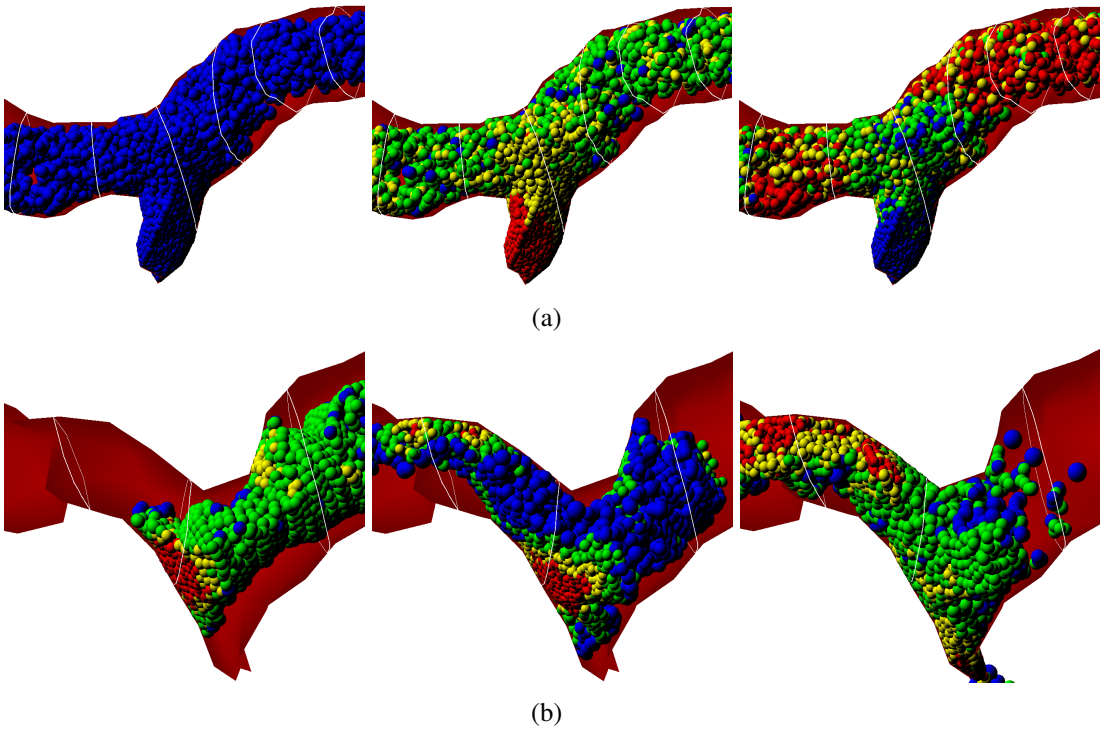


Figure 3. Example simulations with color-coded particles to visualize physical quantities.

4. Discussions and Conclusions

In our opinion the centerline approach worked as an efficient and realistic manner to simulate an external force acting on the fluid. The approximate centerline algorithm takes an average of 60 milliseconds to process a polygonal mesh with about 2000 triangles. Once the centerline has been generated, the actual computation of the centerline force acting on the particles takes an average of 2 percent of the time of each simulation step. For a scene like the one on Figure 3(a), the centerline force computation takes an average of 2 milliseconds per frame for 4000 particles, and 4 milliseconds for 8000 particles for a centerline composed of regular line segments. This value is linearly dependent on the number of particles, but independent of the number of triangles in the mesh, since the method deals only with the cross-sections and the line segments.

The inner characteristics of the SPH algorithm allowed to obtain updated values of the physical quantities of fluid's particles and to apply them on our color coded classification. In the experiments on Section 3 we could observe the variations of the pressure and velocity among the particles, in a simple and intuitive way.

Although the experiments data were taken from medical images, the fluid's parameters used were artificial. We aimed at evaluating the flow behavior regarding the reaction to an external force and on the collision with polygons of an arbitrary mesh. A future step is the acquisition of patient-custom anatomical information and the adaptation of the method to support such data, both in the fluid simulation and the external force computation. This implies the use of deformable structures to model the vessels.

Because of the lack of proper anatomical parameters and the absence of de-

formable structures to model the structure, it is hard to compare the method with previous, existing techniques. For example, the coupling of a 3D model of the cardiovascular system in a specific region of interest with a 1D representation for the rest of the system has been employed to perform a more accurate simulation with physically feasible parameters. This approach is taken by Blanco et al. [Blanco et al. 2007] using tetrahedral meshes built from real medical image data to represent the artery in 3D to evaluate variations in the inflow boundary between the 1D and the 3D models. Other methods include the adoption of pseudo-particles to implement a repulsion-attraction force [Müeller et al. 2004], and impulse-based techniques to add "stickiness" to simulate cohesion and perform the interaction with rigid bodies [Clavet et al. 2005].

The Asclepios research team at INRIA [Ayache et al. 2008] has extended experience in the field of analysis and simulation of medical images, and such expertise can be invaluable to obtain better anatomical data and in the realistic simulation of cardiovascular phenomena.

References

Ayache, N., Clatz, O., Delingette, H., Malandain, G., Pennec, X., and Sermesant, M. (2008). Asclepios: a research project-team at inria for the analysis and simulation of biomedical images. In *From semantics to computer science: essays in honor of Gilles Kahn*. Cambridge University Press.

Blanco, P. J., Pivello, M. R., Feijoó, R. A., and Urquiza, S. (2007). Sensitivity of blood flow at the carotid artery to the heart inflow boundary condition. *III International Congress on Computational Bioengineering*.

Carvalho, D. D. B., dos Santos, T. R., and von Wangenheim, A. (2006). Measuring arterial diameters for surgery assistance, patient customized endovascular prosthesis design and post-surgery evaluation. *Computed-Based Medical Systems (CBMS)*, 0:225–230.

Clavet, S., Beaudoin, P., and Poulin, P. (2005). Particle-based viscoelastic fluid simulation. In *SCA '05: Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 219–228, New York, NY, USA. ACM.

Jiang, G. and Gu, L. (2005). An automatic and fast centerline extraction algorithm for virtual colonoscopy. In *IEEE-EMBS 2005. 27th Annual International Conference of the Engineering in Medicine and Biology Society, 2005.*, pages 5149–5152.

Monaghan, J. J. (2005). Smoothed particle hydrodynamics. *Reports on Progress in Physics*, 68(8):1703–1759.

Müeller, M., Schirm, S., Teschner, M., Heidelberger, B., and Gross, M. (2004). Interaction of fluids with deformable solids. *Computer Animation and Virtual Worlds*, 15(3-4):159–171.

Müller, M., Charypar, D., and Gross, M. (2003). Particle-based fluid simulation for interactive applications. In *SCA '03: Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 154–159, Aire-la-Ville, Switzerland, Switzerland. Eurographics Association.

II

Automatics and Numerical Analysis

COLIBRI 2009 • 22-23 de julho de 2009 • ISBN 978-85-7669-245-4

High Performance Applications on Hierarchical Shared Memory Multiprocessors

Christiane Pousa Ribeiro[1], Márcio Castro[2], Fabrice Dupros[3]
Alexandre Carissimi[4], Luiz Gustavo Fernandes[2], Jean-François Méhaut[1]

[1]INRIA Mescal Research Team,
LIG Laboratory, University of Grenoble, France

{christiane.pousa,jean-francois.mehaut}@imag.fr

[2]GMAP Research Team
PPGCC, PUCRS, Brazil

{marcio.castro,luiz.fernandes}@pucrs.br

[3]BRGM - Orléans Cedex 2, France

f.dupros@brgm.fr

[4]Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre, Brazil

asc@inf.ufrgs.br

Abstract. *Collaborative research works between Brazilian and French universities are not recent. In the past years, these cooperations have allowed research groups to exchange knowledge and experiences in many different fields. High Performance Computing (HPC) is one of these research areas in which research groups from both countries are collaborating in order to achieve meaningful results. This work presents the research carried out between Brazil and France concerning the impact of hierarchical shared memory multiprocessors platforms on high performance applications. Hierarchical shared memory multiprocessors platforms are becoming very common in HPC due to the large number of cores in modern chips. In such platforms, the non uniform memory access is an important constraint and affinity between data and threads must be present. In this paper, we introduce two solutions for memory affinity on such platforms and the results we have obtained with high performance applications.*

1. Introduction

Researches on HPC carried out by universities in Brazil and France have led to cooperation which allowed researches groups to exchange experiences. As a result, several Brazilian students became professors and researches in Brazilian universities, like Gustavo Fernandes [Fernandes 2002] and Alexandre Carissimi [Carissimi 1999] who have worked on this actual paper. Nowadays, the exchange still occurs, allowing uninterrupted cooperations between both countries[1].

[1]Christiane Pousa is under grant CAPES and Marcio Castro is under grant INRIA.

One of these collaborations concerns the impact of hierarchical shared memory multiprocessors platforms on high performance applications. In such platforms, a high number of processing units (multiple cores) and an extremely large shared memory are the main characteristics. To minimize memory contention problems, the shared memory is physically distributed into several memory blocks interconnected by a network. Due to this, memory access costs are not symmetric and can be very high. The effects of applications memory access costs in such platforms can be reduced through the use of affinity between data and threads.

Specifically, in the studies carried out between Brazilian and French researchers, we have analysed the impact of different cache-coherent Non Uniform Memory Access (ccNUMA) architectures on two geoscientific applications ([Méhaut 2007, Dupros et al. 2009][2] and [Castro et al. 2009]). In both applications, the concept of memory affinity has been used in order to comprehend their behavior when different memory policies are applied. In this work, we present our solutions and the obtained results through this collaboration.

This paper is organized as follows: we have first defined NUMA architectures and memory affinity. Then, we have discussed different ways to manage memory affinity on NUMA architectures. We have explained the results that were obtained with the two geoscientific applications. Finally, we have shown our findings and present our future works.

2. NUMA Architectures and Memory Affinity

A ccNUMA platform is a multiprocessed system in which the processing elements are served by multiple memory levels, physically distributed through the platform. These multiple memory levels are seen by the developer as a single shared memory. ccNUMA platforms combine the efficiency and scalability of Massively Parallel Processing (MPP) with the programming facility of Symmetric Multiprocessor (SMP) machines. However, due to the fact that the memory is distributed through the machine nodes, the time spent to access data is conditioned by the distance between the processor and memory banks. A memory access by a given processor could be local (if data is in the processor local memory) or remote (it has to use the interconnection network to access the data) [Ribeiro and Méhaut 2008].

As these platforms are being widely used in HPC and memory access cost can be high, it is important to assure memory affinity on them. Memory affinity is the guarantee that processing units will always have their data close to them. Thus, to reach the maximum performance on ccNUMA architectures, it is necessary to schedule processors and data in such way that the distance between them become the shortest [Carissimi et al. 2007]. To assure memory affinity in these platforms, many different solutions were proposed by research groups. Such solutions are based on algorithms, mechanisms and tools responsible for memory pages allocation, migration and replication to guarantee memory affinity. However, they do not allow the use of different memory policies in the same application. Additionally, most of them must be explicitly integrated in the source code or operating system and it may be a complex task.

[2]NUMASIS http://numasis.gforge.inria.fr/

3. Analysis and Control of Memory Affinity

Our research is based on how memory affinity impacts on the performances of several applications. Our goal is to find the most efficient way to manage memory affinity. In order to do this we must find solutions which will gather three main characteristics: simplicity, portability and performance.

Before studying ways to manage memory affinity, we did several experiments considering different ccNUMA platforms and applications. Such experiments helped us understand not only the importance of memory affinity management in these platforms, but also the relation between memory affinity techniques, ccNUMA platforms and applications characteristics (as presented in [Ribeiro et al. 2008] and [Ribeiro and Méhaut 2008]). Thus, we have identified what is important to consider while developing memory affinity solutions.

Memory Affinity interface (MAi) is the first solution we have proposed to manage memory affinity on ccNUMAs [3]. MAi is a user level Application Programming Interface (API) that provides a simple way to control memory affinity on applications over ccNUMA platforms. It simplifies memory affinity management issues, since it provides simple and high level functions that can be called in the application source code. MAi main characteristics are its simplicity of use (less complex than other solutions: NUMA API, numactl, etc.) and its fine control of memory (several variable based memory policies), its portability (it works in different Linux based platforms and with different compilers) and its performance (better performance than other standard solutions). A full description of MAi and some more experiments can be found in [Ribeiro and Méhaut 2008].

We named our second solution Memory Affinity preprocessor (MApp). It is a preprocessor (support for C language) that enables an implicit control of memory affinity in Linux based ccNUMA platforms. MApp makes optimizations in the application considering memory affinity aspects at the compile time. This preprocessor applies a memory policy on the application variables considering the ccNUMA platform characteristics. MApp main characteristics are its simplicity of use (implicit memory affinity optimizations, no source code modifications) and its portability (it works in different platforms and with different compilers GNU C Compiler-GCC and Portland Group C compiler-PGI).

4. Results

In this section, we present an overview of the first results of MAi and MApp. The results were obtained from several experiments carried out on ccNUMA platforms (AMD Opteron [4], Itanium 2 [5] and SGI [6]) using Geoscience applications (ICTM [Castro et al. 2009] and Ondes 3D [Dupros et al. 2009]).

Interval Categorizer Tessellation Model (ICTM) is a multi-layered tessellation model for the categorization of geographic regions considering several characteristics (relief, vegetation, climate, etc.), using information extracted from satellite images. During the categorization, ICTM executes sequential phases in two dimensional matrices that are

[3]MAi can be download from http://mai.gforge.inria.fr/
[4]AMD - http://www.amd.com/
[5]Intel - http://www.intel.com
[6]SGI - http://www.sgi.com

accessed in an irregular way. Each phase is responsible for computing data stored in different matrices of two dimensions. ICTM was implemented in C++ using OpenMP to code parallelization.

ONDES3D is an application for the simulation of seismic wave propagation in three dimensional geological media. It is developed by the French geological survey (BRGM - French geological survey - www.brgm.fr) and it is mainly used for strong motion analysis and seismic risk assessment. This application has three main steps: data allocation, data initialization and propagation calculus. In all steps the three dimensional arrays are accessed in a regular way (same data pattern access). ONDES3D was implemented in C using OpenMP to code parallelization.

In Figure 1, we present the speedups obtained for ICTM with the eight memory policies of MAi (our solution) and the Linux default memory policy (*first-touch*). Each bar represents the speedup with one of the memory policies in the three ccNUMA platforms. For these results, we have used 16 threads and 2Gbytes as problem size. As we have observed, MAi memory policies presented the best speedups for all platforms. Linux default policy places data in the node that causes the first page fault. In ICTM, the first step is done sequentially and is responsible for all data allocation and initialization. Due to this, all data is placed in the master thread node as this thread does the ICTM first step. Thus, the other threads execute a large number of remote accesses to get their data. MAi memory policies present better results, since they spread data between the platform nodes, assuring thread and data affinity.

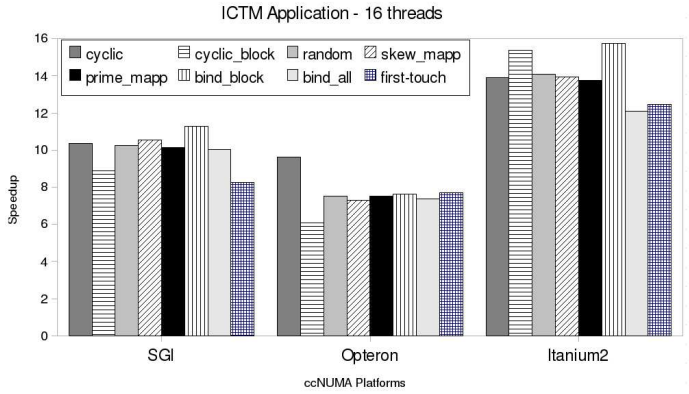


Figura 1. ICTM speedups with 16 Threads in SGI, Opteron and Itanium2 platforms.

The experiments carried out with MAi have shown that the applications implemented with it have better performance than standard solutions. The results show average gains of 30% in the considered platforms. Additionally, similar gains have also been obtained in the platforms with different compilers (GCC, PGI and Intel C Compiler-ICC). The use of MAi is less complex than other APIs and it has seven memory policies to manage memory affinity.

We present, in figure 2, the speedups obtained for ONDES3D with MAi (using the best memory policy for the application), MApp and the Linux default memory policy (*first-touch*). Each bar in the figure represent the speedup with one of the solutions in the three ccNUMA platforms. For these experiments, we have used 16 threads and a size

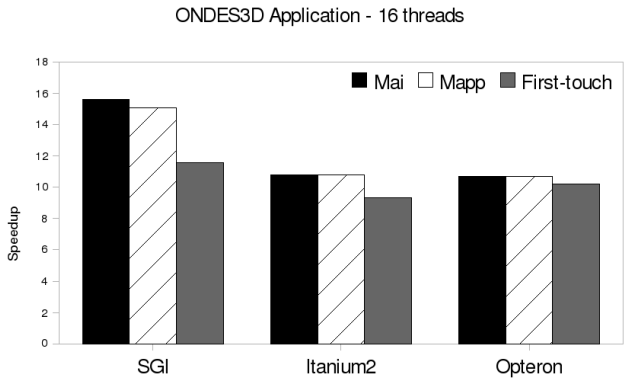


Figura 2. ONDES3D speedups with 16 Threads in SGI, Opteron and Itanium2 platforms

problem of 4Gbytes. MApp presented great speedups when compared with Linux default memory policy (about 20% better than *first-touch*) for SGI and Opteron machines. This preprocessor does some automatic code optimizations to assure memory affinity, considering the machine and application characteristics. Additionally, we have observed that MApp and MAi had very similar speedups for the machines. MApp performs its memory affinity optimizations using only two memory policies (the first one minimizes latency impacts and the second one maximizes bandwidth) whereas, MAi has eight memory policies that can be used in the applications code. Thus, specialists can manage memory affinity issues considering the application memory access patterns.

We have obtained positive results concerning the first experiments with MApp. First of all, the experiments have shown that the preprocessor solution works efficiently in different platforms. Secondly, MApp could do memory affinity optimizations considering the architecture characteristics with good results. Support for other compilers are being implemented in MApp (ICC).

5. Conclusions and Future Works

In this paper, we have presented the research that has been carried out on memory affinity management by Brazilian (PUCRS and UFRGS) and French (LIG and Grenoble University) researchers. This cooperation led to the development of MAi and Mapp softwares that manage memory affinity on cache-coherent NUMAs. Furthermore, we have presented the results obtained with such solutions using Geoscience applications over three ccNUMA platforms.

We have observed some improvement in the applications when the solutions we have proposed were applied. Indeed, performance gains were about 30% in the considered platforms. Additionally, our solutions are simple to use, less intrusive and portable (performance portability).

Thanks to the collaboration between both countries, five of our works were submitted/published and we had the opportunity to present these works in conferences worldwide [Carissimi et al. 2007], [Ribeiro et al. 2008], [Ribeiro and Méhaut 2008], [Castro et al. 2009], [Pousa et al. 2009] and [Dupros et al. 2009]. Since ccNUMA platforms are becoming widely used in HPC, this actual work is also aimed at promoting and developing the successful exchanges between both countries.

In our current works, still in collaboration with researchers in Brazil, we focus on the integration of MAi and MApp in Minas (Memory affInity maNAgement Software) [7] - a software to manage memory affinity in automatic way in large scale hierarchical shared memory multiprocessors platforms. This successful collaboration is the first step to future collaborations with other countries (International laboratory INRIA-Illinois).

References

Carissimi, A. (1999). *Athapascan-0 : Exploitation de la multiprogrammation légere sur grappes de multiprocesseurs*. PhD thesis, Institut National Polytechnique de Grenoble.

Carissimi, A., Dupros, F., Mehaut, J.-F., and Polanczyk, R. V. (2007). Aspectos de Programação Paralela em arquiteturas NUMA. In *VIII Workshop em Sistemas Computacionais de Alto Desempenho*.

Castro, M., Fernandes, L. G., Pousa, C., Méhaut, J.-F., and de Aguiar, M. S. (2009). NUMA-ICTM: A Parallel Version of ICTM Exploiting Memory Placement Strategies for NUMA Machines. In *PDSEC '09: Proceedings of the 23rd IEEE International Parallel and Distributed Processing Symposium - IPDPS*, Rome, Italy. IEEE Computer Society.

Dupros, F., Pousa, C., Carissimi, A., and Méhaut, J.-F. (2009). Parallel Simulations of Seismic Wave Propagation on NUMA Architectures. In *ParCo'09: International Conference on Parallel Computing (to appear)*, Lyon, France.

Fernandes, L. G. (2002). *Parallélisation d'un Algorithme d'Appariement d'Images Quasi-dense*. PhD thesis, Institut National Polytechnique de Grenoble.

Méhaut, J.-F. (2007). Architectures: Informatiques pour la Simulation en Géosciences. *Journal for a Sustaintable Earth - Géosciences*, 6:54–57.

Pousa, C., Castro, M., Fernandes, L. G., Carissimi, A., and Méhaut, J.-F. (2009). Memory Affinity for Hierarchical Shared Memory Multiprocessors. In *International Symposium on Computer Architecture and High Performance Computing - SBAC-PAD (submitted)*, São Paulo,Brazil. IEEE Computer Society.

Ribeiro, C., Dupros, F., Carissimi, A., Marangozova-Martin, V., Méhaut, J.-F., and de Aguiar, M. S. (2008). Explorando Afinidade de Memória em Arquiteturas NUMA. In *WSCAD '08: Proceedings of the 9th Workshop em Sistemas Computacionais de Alto Desempenho - SBAC-PAD*, Campo Grande, Brazil. SBC.

Ribeiro, C. P. and Méhaut, J.-F. (2008). Memory Affinity Interface. Technical Report RT-0359, INRIA.

[7] Minas Homepage: http://pousa.christiane.googlepages.com/minas

Estudo sobre Métodos de Krylov e seus Aspectos Geométricos

Rafael Ferreira Lago[1], Luiz Mariano Carvalho[2], Michael de Souza[1], Nelson Maculan[1]

[1]COPPE/PESC – Universidade Federal do Rio de Janeiro
21941-972 – Rio de Janeiro – RJ – Brasil

[2]Departamento de Matemática Aplicada – IME – Universidade Estadual do Rio de Janeiro
20559-900 – Rio de Janeiro – RJ – Brasil

rafael@cos.ufrj.br,luizmc@ime.uerj.br,{michaelfs,maculan}@cos.ufrj.br

Abstract. *Iterative methods for solving sparse linear systems have been widely applied in many kind of problems. This research field seems to be a promising development, incentiving us to study such methods.*
We base this contribution on [1], where methods are classified into two families, some relations are stabilished as well as a geometric interpretation of them.

Resumo. *Métodos iterativos para solução de sistemas lineares esparsos de grande porte tem sido largamente utilizados em diversos tipos de aplicações. Esta área ainda promete uma grande evolução, incentinvando um estudo mais atento do assunto.*
Nos basearemos em [1], onde os métodos são classificados entre duas famílias, são estabelecidas várias relações bem como uma interpretação geométrica.

1. Representação Gráfica do GMRES e do FOM

O método GMRES [2] tem sido um dos métodos mais utilizados para solução iterativa de sistemas lineares não hermitianos do tipo $Ax = b$, com $x, b \in \mathbb{C}^m, A \in \mathbb{C}^{m\times m}$.

Tomando um ponto inicial x_0 e o seu respectivo resíduo $r_0 = b - Ax_0$, o GMRES busca na k-ésima iteração um vetor $c_k \in \mathcal{K}_k$, tal que $x_k = x_0 + c_k$ minimize a norma do resíduo

$$r_k = b - Ax_k = b - Ax_0 - Ac_k = r_0 - Ac_k, \tag{1}$$

onde $\mathcal{K}_k = span\{r_0, Ar_0, \ldots, A^{k-1}r_0\}$ é o espaço de Krylov de ordem k. Definindo $\mathcal{W}_k := A\mathcal{K}_k$, este problema pode ser visto como

$$\min_{w\in\mathcal{W}_k} \|r_0 - w\|_2 . \tag{2}$$

Sabemos que a projeção ortogonal de r_0 em $\mathcal{W}_k$ resolve o problema acima, como é mostrado em [3], p.435. Portanto

$$r_k = r_0 - P_{\mathcal{W}_k} r_0 = (I - P_{\mathcal{W}_k})r_0 \tag{3}$$

que é a projeção ortogonal em $\mathcal{W}_k^{\perp}$. De posse destas informações, elaboramos uma representação gráfica do método GMRES na k-ésima iteração que é mostrada na figura 1.

Vamos introduzir agora um método semelhante em alguns aspectos ao GMRES, cujo nome é FOM [4]. Este método resolve

$$r_k = r_0 - w \perp \mathcal{K}_k, w \in \mathcal{W}_k \tag{4}$$

a cada iteração k. Um resultado conhecido cuja prova pode ser encontrada em [1] nos dá que $w = P_{\mathcal{W}_k}^{\mathcal{K}_k} r_0$, a projeção oblíqua de r_0 em $\mathcal{W}_k$ ortogonal a $\mathcal{K}_k$ resolve o problema em (4). A figura 2 mostra uma representação gráfica deste método.

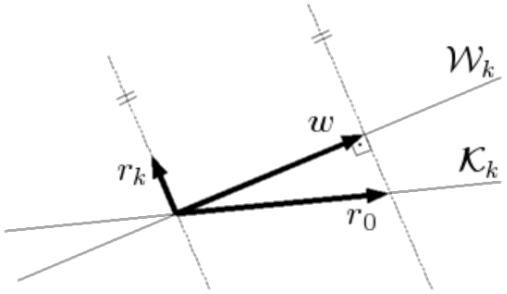


Figura 1. Representação gráfica do GMRES

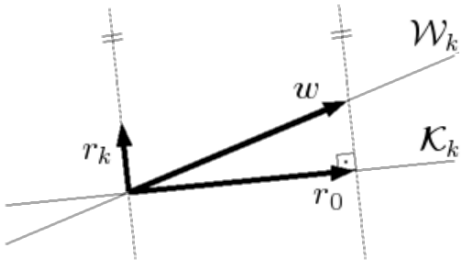


Figura 2. Representação gráfica do FOM

A seguir enunciamos um teorema que pode ser encontrado em [5]

Teorema 1. Se o polinômio mínimo da matriz nãosingular A tem grau L, então a solução de $Ax = b$ pertence ao espaço $\mathcal{K}_L(A, b)$.

Isto mostra que estes algoritmos convergem[1], ao menos considerando precisão infinita, para a solução em no máximo L passos, onde L é o grau do polinômio mínimo de A. Como há garantia de convergência em um número finito de passos, tais métodos tornam-se interessantes. Daqui em diante, consideraremos L como sendo a iteração na qual o método teria convergido considerando precisão infinita.

Tanto o FOM quanto o GMRES utilizam o *processo de ortogonalização de Arnoldi*, que cria bases $\{v_j\}_{j=1}^{k}$ ascendente aninhada ortonormal para $\mathcal{K}_k$, assim como uma base $\{w_j\}_{j=1}^{k}$ ascendente aninhada ortonormal para $\mathcal{W}_k$. Lembrando que definimos $\mathcal{W}_k = A\mathcal{K}_k$, teremos uma decomposição relacionada a essa base que é

$$W_k = AV_k, \qquad AV_k = V_{k+1}\tilde{H}_k = V_k H_k + U_k. \tag{5}$$

onde $V_{k+1} = [v_1, v_2, \ldots, v_k + 1]$, $W_k = [w_1, w_2, \ldots, w_k]$, H_k é uma matriz de Hessenberg superior e

$$\tilde{H}_k = \begin{bmatrix} H_k \\ \eta(e_k^k)^T \end{bmatrix}, \qquad U_k = [0 \ \ldots \ 0 \, \eta \, v_{k+1}] \in \mathbb{C}^{k\times k}, \qquad r_0 = V_{k+1}\beta e_1^k. \tag{6}$$

Substituindo em (2), e notando que V_{k+1} é uma matriz ortogonal, obtemos

$$\min_{y\in\mathbb{C}^k} \left\| V_{k+1}(\beta e_1^{k+1} - \tilde{H}_k y) \right\|_2 = \min_{y\in\mathbb{C}^k} \left\| \beta e_1^{k+1} - \tilde{H}_k y \right\|_2 \tag{7}$$

o que nos leva à solução do sistema $\tilde{H}_k y = \beta e_1^{k+1}$. De posse do y da iteração podemos então encontrar $w_k = V_{k+1}\tilde{H}_k y = W_k y$ e $c_k = V_k y$. Similarmente, de (4)

$$V_{k+1}(\beta e_1^{k+1} - \tilde{H}_k y) \perp \mathcal{K}_k. \tag{8}$$

Se fizermos $y = (H_k)^{-1}\beta e_1^k$, as k primeiras colunas de V_{k+1} desaparecerão em (8), restando apenas algum múltiplo de v_{k+1}, que é ortogonal a K_k. Repare que em (2) o problema envolve uma matriz retangular enquanto em (8) envolve uma matriz quadrada.

[1] métodos que utilizam a projeção oblíqua mencionada, com o FOM, podem ter rupturas

2. Diminuindo Recursos

O problema dos métodos apresentados na seção anterior é que se faz necessário o armazenamento de $\{v_j\}_{j=1}^{k}$ para encontrar v_{k+1}. Entretanto, caso H_k possua apenas uma banda de largura p não nula, precisaremos apenas dos v_j, $j = p..k$ para calcular v_{k+1}.

De (5) vemos que caso $A^H = A$ então $H_k^H = H_k$, o que significa que H_k é tridiagonal pois é uma matriz de Hessenberg superior. Caso isto ocorra, só precisaremos de três vetores a cada iteração para calcular o próximo vetor. Dois métodos que utilizam essa estratégia são o método de Lanczos e o Gradientes Conjugados, sendo muito semelhtanes ao FOM. O método dos Resíduos Conjugados, utiliza a simetria da matriz para resolver (7), assemelhando-se ao GMRES.

Caso a matriz A não seja hermitiana, uma das estratégias para diminuir a banda de H_k é utilizar uma base biortogonal, como no método de Lanczos não-hermitiano que gera vetores tais que $\langle v_i, \bar{v}_j \rangle = 0$, $i \neq j$, onde $\{v_j\}_{j=1}^{k}$ e $\{\bar{v}_j\}_{j=1}^{k}$ são bases para $\mathcal{K}_k$ e $\bar{\mathcal{K}}_k = span\{r_0, A^H r_0, \ldots, (A^H)^{k-1} r_0\}$ respectivamente. Em [6] temos uma demonstração de que utilizando tais bases obteremos uma matriz $\tilde{H}_k$ tridiagonal.

Entretanto, se V_{k+1} não é ortogonal não podemos tirá-lo da norma em (7). Considerando uma matriz V tal que $V_L^H V V_L = I$ e $v, w \in \mathcal{V}_L$, teremos

$$\begin{aligned} &\langle v, w \rangle_V = \langle V_L x, V_L y \rangle_V = \langle x, y \rangle \\ &\min_{y \in \mathbb{C}^k} \left\| V_{k+1}(\beta e_1^{k+1} - \tilde{H}_k y) \right\|_V = \min_{y \in \mathbb{C}^k} \left\| \beta e_1^{k+1} - \tilde{H}_k y \right\|_2 \end{aligned} \tag{9}$$

Notemos que com isso todos os cálculos devem ser feitos em consideração a este produto interno, incluindo ortogonalizações. Logo, neste caso o resíduo r_k será V-ortogonal à $\mathcal{W}_k$. Dentre os métodos que utilizam este artifício citamos o QMR e o BCG, associados ao GMRES e FOM respectivamente.

3. RM e RO

Nesta seção iremos estabelecer uma relação mais sólida entre os métodos apresentados. Durante toda esta seção assumiremos que $\{\mathcal{W}_j\}_{j=0}^{L}$ é uma sequência aninhada de espaços onde $dim(\mathcal{W}_j) = j$, e $\mathcal{V}_j = \{span\{r_0\} + \mathcal{W}_{j-1}\}$[2]. Começaremos citando de [1] o seguinte teorema:

Teorema 1. Seja $\{h_j\}_{j=0}^{L}$ uma sequência de aproximações $r_0 \in \mathcal{H}$, tal que $h_j \in \mathcal{W}_j$ e $h_L = r_0$. Então um produto interno $\langle ., . \rangle_V$ para $V_L = W_L$ tal que

$$\left\| r_0 - h_j \right\|_V = \min_{w \in \mathcal{W}_j} \| r_0 - w \|_V \tag{10}$$

para $j = 1, 2, \ldots, L-1$, existe se e somente se $h_j \in \mathcal{W}_{j-1}$ implica em $h_j = h_{j-1}$, para $j = 1, 2, \ldots, L-1$.

A prova deste teorema pode ser encontrada no referente artigo. Repare que caso encontremos tal $h_L = r_0$ teremos $r_k = 0$, deixando claro em $r_k = b - Ax_k$ que encontramos a solução procurada. Isto está intimamente ligado ao fato de $\mathcal{V}_L = \mathcal{W}_L$, pois com isso $w_j \in \mathcal{V}_j$; o que nos leva, por exemplo no GMRES, a $h_L = P_{\mathcal{W}_L} r_0 = P_{\mathcal{V}_L} r_0 = r_0$.

[2] embora nos métodos mencionados até agora $\mathcal{V}_j = \mathcal{K}_j$, trataremos de forma mais genérica por hora

Podemos então definir *Resíduo Mínimo* como sendo a família de métodos que, de posse de uma base aninhada de subespaços $\{\mathcal{W}_j\}_{j=0}^{L}$ tenta encontrar uma aproximação de r_0 através de (10), para alguma norma $\|.\|_V$, e que termina com $h_L = r_0$.

Isto engloba os métodos GMRES, QMR, CR e o GCR (uma versão do CR para matrizes não simétricas). Para os demais métodos, temos o seguinte teorema, bem semelhante ao anterior:

Teorema 2. Seja $\{h_j\}_{j=0}^{L}$ uma sequência de aproximações $r_0 \in \mathcal{H}$, tal que $h_j \in \mathcal{W}_j$ e $h_L = r_0$. Então um produto interno $\langle .,. \rangle_{\tilde{V}}$ para $V_L = W_L$ tal que

$$r_0 - h_j \perp_{\tilde{V}} \mathcal{V}_j, \qquad j = 1, 2, \ldots, L-1 \tag{11}$$

existe se e somente se $h_j \in \mathcal{W}_{j-1} \backslash \mathcal{W}_j$, ou seja, $h_j \in \mathcal{W}_{j-1}$ e $h_j \notin \mathcal{W}_{j-1}$, para $j = 1, 2, \ldots, L-1$.

Do mesmo modo, quando $h_L = r_0$ teremos a solução ótima. Definimos então *Resíduo Ortogonal* como sendo a família de métodos que através dos espaços mencionados, busca uma solução para o problema (11) para algum produto interno $\langle .,. \rangle_{\tilde{V}}$, e que termina com $h_L = r_0$, englobando os métodos FOM, CG, e BCG.

Os teoremas apresentados aqui garantem que, mediante ajustes no produto interno, qualquer método que possua subespaços conforme definidos aqui podem ser classificados como métodos RM ou RO, o que engloba a grande maioria dos métodos.

Mesmo com estas modificações, as representações gráficas dos métodos RM e RO são análogas às figuras 1 e 2 respectivamente, pois pelas equações (10) e (11), ainda podemos interpretar a aproximação h_j de r_0 como uma projeção V-ortogonal a $\mathcal{W}_j$ e como uma projeção oblíqua em $\mathcal{K}_j$, V-ortogonal a $\mathcal{W}_j$, respectivamente.

4. Relações Entre Subespaços e Ângulos

De posse das definições da seção anterior, também podemos estabelecer propriedades mais abrangentes como os seguintes teoremas, cujas provas podem ser encontradas em [1]:

Teorema 1. Sejam $\mathcal{W}_j$ e $\mathcal{V}_j$ dados conforme no início da seção §3. Então as aproximações RM e RO de r_0 satisfazem

$$\left\| r_j^{RM} \right\|_V = s_j \left\| r_{j-1}^{RM} \right\|_V, \quad s_j = \sin \measuredangle_V(r_{j-1}^{RM}, \mathcal{W}_j) \tag{12}$$

$$\left\| r_j^{RM} \right\|_V = c_j \left\| r_j^{RO} \right\|_V, \quad c_j = \cos \measuredangle_V(r_{j-1}^{RM}, \mathcal{W}_j) \tag{13}$$

$$\left\| r_j^{RM} \right\|_V = s_1 s_2 \ldots s_j \|r_0\|_V, \quad \left\| r_j^{RO} \right\|_V = s_1 s_2 \ldots s_j \|r_0\|_V \tag{14}$$

$$\tag{15}$$

Teorema 2. Sejam $\mathcal{W}_j$ e $\mathcal{V}_j$ dados conforme no início da seção 3. Então o maior ângulo canônico entre eles é dado por

$$\measuredangle_V(\mathcal{V}_j, \mathcal{W}_j) = \measuredangle_V(r_{j-1}^{RM}, \mathcal{W}_j) \tag{16}$$

Além disso, os outros $m-1$ ângulos canônicos entre $\mathcal{V}_j$ e $\mathcal{W}_j$ são zero.

Isto nos mostra uma relação entre a evolução do algoritmo e o ângulo entre os subespaços a cada iteração. Na equação (12) fica claro que quando $s_j = 0$, então a norma de r_j^{RM} será igual a zero, e portanto, teremos encontrado a nossa solução. Analisando esse resultado geometricamente, podemos finalmente compreender que os métodos RM simplismente aproximam o subespaço $\mathcal{W}_j$ de $\mathcal{K}_j$ gradativamente, até que coincidam, concordando com $\mathcal{W}_L = \mathcal{V}_L$.

Na equação 3.2 de [1] ainda, temos

$$\left\| r_j^{RM} \right\|^2 = \left\| r_{j-1}^{RM} \right\|^2 - |\langle r_0, w_j \rangle|^2. \tag{17}$$

Adaptando para o caso de um produto interno genérico, caso $\langle r_0, w_j \rangle_V = 0$, não teremos uma melhora no resíduo do método RM na j-ésima iteração, o que vem a ser uma idéia natural, pois se estamos procurando uma aproximar r_0 construindo uma base W_j, não faz sentido procurar tal base na direção ortogonal a r_0.

Utilizando esse resultado em (12), vemos que se $\langle r_0, w_b \rangle_V = 0$ então $s_b = 1$, o que obriga $c_b = 0$, implicando numa divisão por zero em (13).

Quando isso acontece, a aproximação RO falha. Entretanto, métodos para evitar tal falha estão além do escopo deste trabalho, e não discutiremos tais métodos, nos limitando apenas a mostrar o significado geométrico deste fato; tal falha acontecerá apenas quando $\measuredangle_V(r_{j-1}^{RM}, \mathcal{W}_j) = \pi/2$, ou equivalentemente, quando $\measuredangle_V(\mathcal{V}_j, \mathcal{W}_j) = \pi/2$. Com esta observação torna-se óbvia a inexistência de uma projeção em $\mathcal{W}_j$ que seja V-ortogonal a $\mathcal{V}_j$. Em [1] existe uma discussão mais profunda a esse respeito, e uma demonstração de que H_j^{-1} existe se e somente se $c_j \neq 0$.

Algo interessante a notar é que, caso o produto dos senos em (14) tenda a zero, então o algoritmo em questão terá convergência superlinear.

Consideremos construções especiais dos subespaços. Suponhamos que $\mathcal{V}_j = \mathcal{C}_j$ e $\mathcal{W}_j = A\mathcal{C}_j$, com $\mathcal{C}_j$ como *subespaço de correção*. A princípio poderíamos escolher qualquer subespaço de correção, mas conforme já foi mostrado pelo teorema 1, uma solução sempre poderá ser encontrada no subespaço de Krylov, desde que A seja nãosingular, e portanto temos garantia de convergência caso nenhum outro problema ocorra. Entretanto, existem exemplos onde uma matriz de dimensão n não apresenta redução do resíduo até a $(n-1)$-ésima iteração, pois $\measuredangle_V(r_{j-1}^{RM}, \mathcal{W}_j) = \pi/2$, para $j = 1, 2, \ldots, n-1$, o que faz com que a escolha do espaço de correção como sendo o espaço de Krylov nem sempre seja a melhor opção, embora ainda seja a mais utilizada.

5. Conclusão

Neste trabalho, estudamos o funcionamento básico de alguns dos principais métodos iterativos para solução de sistemas lineares de grande porte, e estabelecemos relações entre eles, classificando-os como métodos semelhantes pertencentes a uma das duas famílias apresentadas. Dentre as relações estabelecidas, as que envolvem aspectos ajudam na compreensão do funcionamento do método.

Referências

[1] EIERMANN, M.; ERNST, O. G. Geometric aspects in the theory of krylov subspace methods. *Acta Numerica*, v. 10, n. 10, p. 251–312, 2001.

[2] SAAD, Y.; SCHULTZ, M. H. Gmres: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing*, SIAM, v. 7, n. 3, p. 856–869, 1986.

[3] MEYER, C. D. *Matrix Analysis and Applied Linear Algebra*. [S.l.]: SIAM, 2001.

[4] SAAD, Y. Krylov subspace methods for solving large unsymmetric linear systems. *Mathematical Computing*, n. 37, p. 105–126, 1981.

[5] IPSEN, I. C. F.; MEYER, C. D. The idea behind krylov methods. *American Mathematical Monthly*, v. 105, p. 889–899, 1998.

[6] MEURANT, G. *Computer Solution of Large Linear Systems*. [S.l.]: Elsevier North Holland, 1999. (Studies in Mathematics and its Applications, v. 28).

Colaboração Científica na UFSC em Ciências e Tecnologias da Informação, da Comunicação e da Automática

J.M. Farines, J.E.R. Cury, G. Bittencourt[1], E.B. Castelan, E. De Pieri, U.F. Moreno, L.B. Becker, A. Trofino

Depto de Automação e Sistemas - Universidade Federal de Santa Catarina (DAS-UFSC)
Caixa Postal 476 – 88.040-900 – Florianópolis – SC – Brasil

{farines,cury,gb,eugenio,edson,moreno,lbecker,trofino}@das.ufsc.br

Résumé. Cet article présente un resumé de l'activité de coopération scientifique entre l'UFSC et des laboratoires et centres de recherche français, en sciences et technologies de l'information, communications et de l'automatique, pendant ces dernières années, ainsi qu'une perspective des domaines d'interêt pour une collaboration future.

Resumo. Este artigo apresenta um resumo das atividades de cooperação entre a UFSC e instituições de pesquisa francesas, em ciências e tecnologias da informação, da comunicação e da automática, nestes últimos anos, bem como uma perspectiva de temas para futuras colaborações.

1. Introdução

Este artigo apresenta algumas das atividades de colaboração entre a UFSC e diversos laboratórios e centros de pesquisa franceses, em ciências e tecnologias da informação, da comunicação e da automática, nestes últimos anos, bem como uma perspectiva de temas para futuras colaborações. As colaborações iniciaram na década de 70 com a permanência de pesquisadores franceses no Departamento de Engenharia Elétrica, que iniciava sua pós-graduação, e se intensificaram nas duas últimas décadas no contexto do Departamento de Automação e Sistemas (www.das.ufsc.br) e, mais recentemente, da Pós-Graduação em Engenharia de Automação e Sistemas (www.pgeas.ufsc.br).

2. Histórico de cooperações da UFSC em STICs

•"Formação e pesquisa em sistemas de controle automático" (CAPES-COFECUB N° 56/86), com vigência de janeiro de 1987 até dezembro de 1991. Este projeto visava a formação para a pesquisa e pela pesquisa, de professores e pesquisadores da UFSC, e a busca de áreas de pesquisa comuns. Os resultados mais tangíveis deste projeto foram: a formação de diversos professores da UFSC ou que vieram a ingressar nesta universidade, formação ocorrida prioritariamente no LAAS-CNRS, mas também no LAG-INPG, no IRISA, e no INRIA-Lorraine; a definição e consolidação de temas de

[1] In memoriam

pesquisa conjuntos que vieram a se viabilizar na forma de outros acordos de cooperação.

•"Análise e síntese de leis de controle quadráticas robustas" (CNPq-CNRS), com vigência de 1995 a 1996. Deste projeto participaram um professor da UFSC em colaboração com 2 professores do Laboratoire d'Automatique de Grenoble do ENSIEG/INPG. Neste projeto foram realizadas pesquisas na área de controle robusto.

• "Controle sob Restrições de Sistemas Dinâmicos" (CNPq/CNRS N° 910 183/94-5, 910 176/96-5), com vigência de janeiro de 1995 a dezembro de 1998. Este projeto, que teve participação de 3 professores da UFSC e foi realizado em cooperação com pesquisadores do LAAS-CNRS e da Unicamp, teve como objetivo o tratamento de problemas de controle de sistemas lineares considerando limites sobre variáveis de estado ou de controle, devidos a limites tecnológicos e de segurança, utilizando como ferramentas os conceitos de invariância positiva e de estabilidade de Lyapunov.

•"Controle de Sistemas Singulares - Teoria e Aplicações" (CNPq/CNRS N° 910 196/97-4 e 910066/99), com vigência de janeiro de 1999 a dezembro de 2001. A pesquisa realizada por 2 professores da UFSC em cooperação com pesquisadores do LAAS visava tratar os problemas de controle específicos de sistemas singulares, formados por equações diferenciais e algébricas, levando em consideração restrições práticas estruturais e de limites sobre variáveis de estado e de controle, com vistas as aplicações práticas, principalmente em sistemas de potência.

•"Interação entre modelos formais para sistemas de supervisão e informação" (CAPES-COFECUB), com vigência de janeiro de 2003 a dezembro de 2006, teve a participação de 4 professores da UFSC e 5 professores da Université de Toulouse I. O tema central do projeto foi o estudo de formalismos para tratar problemas de modelagem de sistemas complexos, como controle e gerência de sistemas de automação e desenvolvimento de sistemas distribuídos de informação, utilizando a tecnologia de sistemas multiagentes.

• "Concepção de Sistemas Distribuídos de Tempo Real: Aplicação aos Sistemas Críticos e aos Sistemas Multimídias" (CAPES-COFECUB N° 450/04/06), com vigência de janeiro de 2004 a dezembro de 2007. Este projeto de pesquisa no qual participaram 5 professores da UFSC em cooperação com pesquisadores do LAAS-CNRS, abordou questões de qualidade de serviço, restrições temporais, mobilidade, segurança de funcionamento, que permitem tratar a complexidade destes sistemas do ponto de vista de formalismos, arquiteturas, aplicações e metrologia. Ele permitiu a participação da UFSC em dois projetos do LAAS-CNRS: o projeto TOPCASED, coordenado pela AIRBUS para a construção de um Ambiente de Desenvolvimento de Software "Open Source", e o projeto europeu EuQoS, sobre qualidade de serviços em redes heterogêneas.

• "Controle de Sistemas Mecânicos por Realimentação Linearizante Robusta e Otimização Multi-critério" (CAPES-COFECUB N° 489/05), com vigência de janeiro de 2005 a dezembro de 2008. A pesquisa realizada por 3 professores da UFSC em cooperação com pesquisadores da ENS-Cachan e do CNAM-Paris visava o desenvolvimento de estratégias de controle avançado para sistemas eletromecânicos. Os resultados obtidos no estudo e o desenvolvimento de técnicas de controle robusto como

um problema de otimização multi-critério, tratado dentro do contexto da teoria de jogos diferenciais, apresentaram resultados de grande consistência teórica e de simulação.

3. Projetos em andamento e submetidos

•"Timing Analysis and Program Implementation On Complex Architectures - TAPIOCA" (Projeto STIC-Amsud/CAPES N° 003/07), com vigência de janeiro de 2008 a dezembro de 2009. Este projeto tem a participação de 2 professores da UFSC, professores do LAAS/CNRS, do IRIT/UPS, e da VERIMAG/CNRS, e de duas universidades argentinas (UADE e UBA). Ele tem como finalidade o desenvolvimento e integração de metodologias e ferramentas de modelagem e verificação automática que permitam produzir softwares que garantam o respeito das propriedades quantitativas temporais impostas pelas exigências específicas das aplicações do tipo sistemas "embarcados de tempo real", e das restrições da plataforma de execução.

•"Composição e Módulos para Engenharia de Ontologias" (Projeto CNPQ-INRIA), com vigência em 2008 e 2009. Participam 1 professor da UFSC e pesquisadores do INRIA. O projeto tem como objetivo projetar uma linguagem para definir módulos de ontologias que possam permitir reuso e composição. Uma questão central na definição desta linguagem será a definição da interface entre os módulos de forma a acomodar semanticamente o encapsulamento de módulos de ontologias atrás de sua interface.

• "Controle de Sistemas a Eventos Discretos. Metodologia e Aplicação aos Sistemas de Produção e Redes de Transporte" (Projeto Capes-Cofecub N° 642/09), com vigência de janeiro de 2009 a dezembro de 2012. O projeto tem 2 participantes da UFSC, professores da Unicamp e UFMG, e pesquisadores franceses do IRCCyN, Nantes e do ISTIA, Angers. O projeto visa a concepção de sistemas de controle para instalações complexas, modeladas por sistemas a eventos discretos (SED), em especial redes de transporte e sistemas de produção. O objetivo é o estudo de problemas decorrentes da natureza do sistema, composto de sub-sistemas, das especificações que consideram objetivos concorrentes, e do sistema de controle dividido em módulos. Os principais formalismos usados são a teoria de controle supervisório e álgebra dos dióides.

• "Controle e Análise com Integração de Performances para Sistemas Interconectados em Rede vistos como uma Arquitetura Híbrida (CAIPIRAH)" – em avaliação. Submetido ao CNPq em novembro de 2008 e ao CNRS em março de 2009, tem por foco o desenvolvimento de ferramentas para análise de estabilidade e síntese de controladores no contexto de sistemas chaveados submetidos a restrições de saturação e visando a garantia de critérios de desempenho. Este projeto tem 3 participantes da UFSC e pesquisadores franceses do CRAN/INPL, Nancy, e SATIE/ENS, Cachan.

4. Perspectivas

Nesta seção são apresentados alguns temas de interesse para a continuidade e consolidação das pesquisas conjuntas entre professores e pesquisadores do DAS/PPGEAS/UFSC e pesquisadores de instituições de pesquisa na França.

4.1. Robótica móvel

A interação de um conjunto de robôs móveis heterogêneos com o meio, para a realização de tarefas cooperativas, é um dos principais desafios da pesquisa e

desenvolvimento tecnológico em robótica móvel, com potencial de aplicações em tarefas de inspeção, exploração, resgate, ou automação de processos fabris. A complexidade deste estudo requer a atuação de equipes multidisciplinares, com conhecimentos complementares em diferentes campos de conhecimento, tais como: sistemas de controle (controle não-linear, inteligente, supervisório); sistemas de comunicação (redes ad-hoc); sistemas embarcados; sistemas multi-agentes; projeto de sistemas críticos (formalismos) [1]. Entre os objetivos de colaborações nesta temática, vislumbra-se disponibilizar ambientes de simulação e plataformas experimentais conjuntas, e propor novas ferramentas e metodologias de análise e projeto aplicadas à interação em robôs móveis com rodas, manipuladores, e robôs aéreos.

4.2. Desenvolvimento de sistemas embarcados

Os sistemas embarcados, encontrados hoje nas mais diversas aplicações (entre outras, aviônicas, automotivas, e de telecomunicações) tem se caracterizado pela necessidade de um rigor cada vez maior na sua concepção. Aos requisitos funcionais do sistema, vêm se somar exigências de respeito às restrições temporais e de segurança, requisitos de distribuição, limitações devidas aos diversos hardwares suportados e à energia disponível no equipamento. Estes sistemas apresentam uma grande complexidade no seu desenvolvimento, necessitando de novas metodologias, formalismos, e ferramentas, para um processo de concepção que vai da especificação à implementação de hardware e de software, passando pela arquitetura deste sistema [2]. Conceber, construir e testar um ambiente de desenvolvimento orientado a aplicações de sistemas embarcados que seja “open source” e que permita manipular esses formalismos, integrar ferramentas, mas também facilitar a comunicação de dados entre estas, num processo de engenharia dirigida pelos modelos (MDE), são desafios que a UFSC em cooperação com laboratórios franceses se sente capacitada a enfrentar.

4.3. Análise e controle de sistemas de dinâmica híbrida

Muitos sistemas dinâmicos combinam comportamentos típicos de sistemas de dinâmica contínua com comportamentos típicos de sistemas a eventos discretos. O interesse em Sistemas Híbridos cresceu enormemente a partir dos anos 90, nas áreas das Ciências da Computação e Sistemas de Controle. Nesse sentido, autômatos híbridos e sistemas chaveados constituem algumas das classes particulares de modelos para sistemas com dinâmica híbrida. Do ponto de vista da automática, este é um quadro teórico adaptado para tratar uma gama de sistemas que cada vez mais aparece na prática, como é o caso de redes elétricas, sistemas biológicos, sistemas embarcados, sistemas interconectados, e sistemas controlados via redes [3]. A grande abrangência e relevância dos problemas neste domínio motivam a cooperação entre a UFSC e laboratórios franceses na área de Sistemas Híbridos.

Referências

[1] Parker, L. E., "Multiple Mobile Robot Systems", em Springer Handbook of Robotics, B. Siciliano and O. Khatib, editores, 2008.

[2] Wolf, Wayne. Computers as Components: Principles of Embedded Computing System Design. Second edition, Elsevier, 2008.

[3] Goebel, R., Sanfelice, R.G., Teel, A.R. – Hybrid Dynamical Systems. IEEE Control Systems Magazine, pp 28-92, abril 2009.

Methodology for Automatic Observation of sky patterns

Sylvio Luiz Mantelli Neto[1,2,3], Aldo Von Wangenheim[2,3], Eros Comunello[2], Enio Bueno Pereira[1]

[1]INPE-CPTEC Instituto Nacional de Pesquisas Espaciais - Centro de Previsão do Tempo e Estudos Climáticos. Av. dos Astronautas 1758 São José dos Campos SP Brasil 12227-010

[2]UFSC-INE-LAPIX Universidade Federal de Santa Catarina - Departamento de Informática e Estatística - Laboratório de Processamento de Imagens. Florianópolis SC Brasil.

[3]UFSC-EGC Universidade Federal de Santa Catarina - Departamento de Engenharia e Gestão do Conhecimento Florianópolis SC Brasil.

sylvio@lepten.ufsc.br, awangenh@inf.ufsc.br, enio.pereira@cptec.inpe.br

Abstract. *Surface observers on monitoring meteorological stations regularly classify and register sky patterns based on human cognitive training. Substitute those observers by automatic systems is highly desirable to reduce subjective analysis. Most automatic systems use thresholding methods and ignores other physical phenomena existent in atmosphere obtaining only binary results (clouds and sky). The current work proposes a method that will establish the correspondence between observations percept from those patterns and theory using Bayesian approach embedded in Intelligent Agents in the following steps. Model the cognitive patterns on color space obtained from surface cameras and analysis. Validation and refinement of system results could be done using sun photometers and LIDAR systems.*

Keywords: Automatic pattern recognition, Cognitive Bayesian model, Multivariate statistics, atmospheric patterns.

1. Introduction

The main purpose of the current work is the characterization of sky patterns on Celestial Hemisphere (CH) observable from the surface. This characterization will improve the classification and qualification of observed optical atmospheric phenomena allowing the use of intelligent agents (IA) (NORVIG and RUSSEL, 2003). The most widely used characterization of sky patterns present on CH is established by World Meteorological Organization WMO. Synoptic observers use WMO (1975, 1987 and 1996) criteria to classify cloud type, height and amount according current atmospheric conditions. The human observer is intended to be replaced from that task by a surface camera equipped with “fish eye” lens and image interpretation algorithms, reducing efforts imposed by regular human working shifts, subjective aspects and misinterpretation. Sky monitoring by cameras and related methods are also a part of the efforts to environment monitoring activities (WCRP Report 2007). The automatic sky imagers available on the market usually presents binary evaluation of CH patterns (sky and clouds), reducing the

dimension of color space restricting the capabilities present on observed color domain (LONG et al., 2006). WMO classification was established to be used by an observer that has a very well developed and trained cognitive system to watch, classify and register the CH patterns. The synoptic observer has a good performance on qualitative analysis and poor performance on quantitative of the present CH conditions. With automatic systems is the inverse. The defined rules established by WMO are not suitable to be used by IA to replace human observation. With current paradigms IA are not capable yet to fully classify different patterns from those ones defined on the WMO domain environment[1]. If the domain is restricted, so is the program analysis or IA action. The classification criteria established by WMO could be used as a target objective for IA classification, but not as a domain. The sensed domain to be considered in that case is broader than cloud observation parameters established by WMO, leading IA´s to misinterpretation. This misinterpretation increases the uncertainty of the observation and imposes a disappointment on automatic systems outcomes. The current work will show that WMO standards are not suitable to be used as a domain environment, but as a target outcome. It proposes an alternative domain based on optical atmospheric phenomena percept by humans to be used by IA. To achieve that an exploratory data analysis (EDA) is proposed to be performed in color space to find atmospheric patterns typical locus on a multivariate context. As the pattern identification is highly oriented to human interpretation, cognitive modeling using a Bayesian approach (supervised learning and analysis) related tools is intended to be performed. To validate the aforementioned model a field experiment is also proposed to validate and refine the proposed patterns analysis using LIDARS and SUNPHOTOMETERS. The next sections presents a little further details about the proposed method.

2. Cognitive Environment Definition and Physical Context

Proper environment definition is a key issue in the current work. The environment will be defined according to the work proposed by Newell & Simon (1972). In that work it is suggested the definition of three different levels to be taken into account by the IA: research, task and problem environments. Restrictions on those environments could lead to inaccessible states on the problem solution (MANTELLI, 2001). The current studies consider only the visible part of radiation, because sensor and human cognitive systems only respond to visible part of the electromagnetic spectrum. Two principal physical phenomena dominate the interaction of sun light with the atmosphere, scattering and absorption (LILLESAND and KIEFER, 1994; LENOBLE, 1993; IQBAL, 1984). The principal observable scatter causes are: Rayleigh, Mie, aerosol atmospheric turbidity, non selective and sun light diffusion. The principal observable effect of absorption is the dark and grey levels of clouds. More detailed information about proposed environment could be observed on figure 1. In this context an EDA of those patterns on multivariate color space need to be performed to characterize their typical locus.

[1] Domain is a section of the world about which we wish to express some knowledge. P. 197 (RUSSEL AND NORVIG 2003).

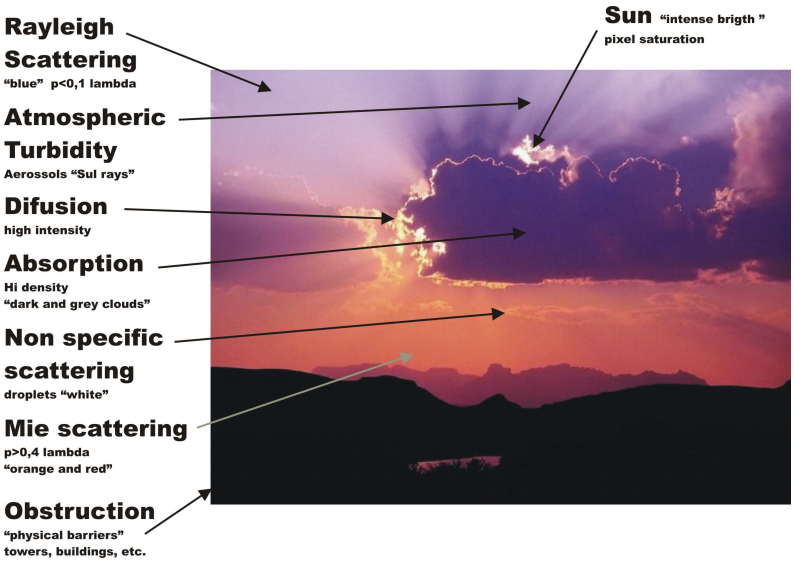


Figure 1. A typical scenario of proposed observable sky patterns from surface.

Categorization of clouds according to WMO is also an issue to be deal with. It is a point to be investigated that clouds could be discriminated by texture. It is considered that cloud texture is the same but different altitudes of same texture causes different properties that could be distinguished. Fourier analysis, wavelets and similar methods is intended to be used to investigate that. From physical context above several patterns may be proposed to take part on the environment and could be easily identified by its results. If not defined properly automatic systems are far behind to percept patterns properly. Task environment defined that way is further beyond WMO standards. The first suggestion is to reduce the identified patterns and map into a WMO goal as suggested on physical context. The difference now is that mapping is done specifically by the system and not unpredictably and improperly by the software.

3. State-of-the art and future work

Based on the aforementioned reasons it was proposed an environment defined by induction using Bayesian cognitive modeling (TENENBAUM, 2006). In that context current research leaded to multivariate statistical exploratory data analysis (EDA) of proposed image patterns. Although the current research is working in EDA definition the appropriate IA inference still have to be defined. The model validation is proposed in two steps. The first step is comparing the system performance against a synoptic observer. The second one, by using advanced field equipments and atmospheric modeling. Among then it is proposed a sun photometer and a LIDAR[2] techniques to measure simultaneous atmospheric and visual parameters such as clouds distances by texture.

4. International cooperation and French-Brazilian teams

The state-of-the-art of sky imaging is being addressed by several research institutes and is being targeted by international scientific community (WCRP Report 2007). The

[2] LIDAR – Laser Detection and Ranging.

original approach and modeling could be done in a joint cooperation between Brazil and France. Several INRIA Research groups could be involved in these tasks.

The first one in formal modeling, representation and interpretation of sky patterns into its symbolic atmospheric meaning. Checking if the proposed method is the more appropriate to be applied on the specific cognitive area. The effects of image illumination variation along the day should also be investigated. (INRIAS's PERCEPTION research group)

The second one is the model selection and statistical learning for pattern recognition in the multivariate image domain of the atmospheric patterns. Patterns should be investigated for their typical distribution (normal or non-normal) on color space. The study could be extended by adding and verifying the effects (or factors) of typical solar variables (global, direct and diffuse) on image. (INRIAS's SELECT research group)

The third potential research group could be on model validation. Validation could be done in several ways. It will be proposed two methods. The first one demands high cost equipments not available to the Brazilian counterpart (i.e. sun photometers, trackers and LIDARS) or satellite images. Some of these manufacturers are based in France (CIMEL), and are present in several French research centers. For satellite images INRIA has a research group working on environment data processing and computer vision techniques (INRIAS's CLIME research group).

5. Expected Scientific results

Several results are expected from current research as listed. Development of a suitable IA in applied area. An improvement of automatic sky imaging observation methodology to be employed on surface monitoring stations. Cost reduction and rationalization of current synoptic operations. Test of a suitable monitoring point for cross validation of satellite images on visible spectrum for clear, cloudy and mixed conditions skies. A better evaluation of atmospheric models on mixed cloudy conditions. A comparison of proposed methodologies based on tropical and tempered climates.

6. Collaboration proposals.

Actually UFSC-INE-LAPIX (Laboratory of Image Processing) do not have cooperation program with any official INRIA partners. Current potential Brazilian Partners includes INPE-CPTEC (Brazilian Institute for Space Research – Center of Weather Forecasts and Environment Studies). The proposed activities for Brazilian partners involve modeling of domain and target sky patterns. Joint affords could be done with PERCEPTION and SELECT groups on cognitive aspects of synoptic observation and statistics modeling. CLIME group could support on validation and atmospheric physics issues of atmospheric modeling. The proposed activities with INRIA partners involve modeling and operational support for field validation of proposed modeling.

7. Conclusions

Although Celestial Hemisphere pattern classification has a satisfactory result when done by a trained operator, automatic classification using surface cameras are subject to uncertainties due to inappropriate approach. Current work proposes a new paradigm on classification matching physical and optical phenomena based on current atmospheric

conditions. This approach is believed to open a new perspective based on human cognitive observations to automatic systems. Once the factors that cause uncertainties are tagged, it is easier to be addressed and overcome them. Independently of the **how** the solution will be implemented **what** to do to solve the problem is now established. Define which intelligent agent to do the job and the best technique to solve it is the next step. The new approach presented here is believed to be a suitable approach for classification of Celestial Hemisphere patterns.

References

Iqbal, M., Introduction to Solar Radiation. Academic Press 1984.

Lenoble, J., Atmospheric Radiative Transfer. 1993 A Deepak Publishing.

Lillesand, T. M., Kiefer, R. W. Remote Sensing and Image Interpretation 1994 John Wiley & Sons Inc.

Long C. N., Sabburg J. M., Calbó J., Pagès D., 2006: Retrieving Cloud Characteristics from Ground-Based Daytime Color All-Sky Images. *J. of Atmos. and Oceanic Technology* 23, issue 5, 633–652

Mantelli S. L., Desenvolvimento de Metodologia para a estimativa da cobertura de nuvens usando uma câmera de superfície e comparando com as imagens de satélite. 125f. Dissertação (Mestrado em Ciências da Computação). Universidade Federal de Santa Catarina, Departamento de Informática e Estatística, Florianópolis, 2001.

Newell, A., Simon H. A., Human Problem Solving. Englewood Cliffs, NJ: Prentice Hall 1972.

Russel S. J. and Norvig P., Artificial Intelligence a Modern Approach. 2003. Prentice Hall.

Tenenbaum J. B., Griffiths T. L. and Kemp C., 2006: Theory-based Bayesian models of inductive learning and reasoning. Special Issue: Probabilistic models of cognition. *TRENDS in Cognitive Sciences*, 10, 309-318

WCRP World Climate Research Program, 2007: *Section of the Baseline Surface Radiation Network*, Lindenberg Germany, Informal Rep. No. 9 54 pp. [Available online at http://www.gewex.org/BSRN/BSRN-9-REPORT.pdf.]

World Meteorological Organization, 1975: International Cloud Atlas: Manual on the observation of Clouds and Other Meteorol. Volume I, WMO-No. 407, Geneva.

World Meteorological Organization, 1987: International Cloud Atlas. Volume II, WMO-No. 407, Geneva.

World Meteorological Organization, 1996: Guide to Meteorological Instruments and Methods of Observation, WMO-No. 8, Geneva.

An Agent Model Using Polychronous Networks*

Júlio L. R. Monteiro[1], Philippe Caillou[2], Marcio Lobo Netto[1]

[1]University of São Paulo – EPUSP
Av. Prof. Luciano Gualberto, t3, 380
05508-970 - São Paulo – SP - Brazil

[2]INRIA Saclay- Île-de-France
Laboratoire de Recherche en Informatique (LRI)

Université Paris Sud XI - IUT de Sceaux

julio.monteiro@poli.usp.br, caillou@lri.fr, lobonett@lsi.usp.br

Abstract. *In this paper, we present an agent model based on computation with polychronous groups on spiked neural networks, that is able to learn to return to known initial situations, without any guidance.*

1. Introduction

The recent developments on biologically inspired spiked neural networks [Gerstner and Kistler 2002] along with plasticity models such as STDP [Froemke and Yang 2002] allows the exploration of a new way to store and manipulate information spread over networks, exploring a feature known as polychronous groups, suggested by [Izhikevich 2006].

In this kind of neural network, neurons fire (or spike) when their activation reaches a defined threshold. Incoming synapses carry potentials from neurons that just fired, or spiked, according to a weight, that determines how much of the firing potential will be transported to the connected neurons.

The synaptic weights vary in accordance to some plasticity model, such as STDP (spike timing dependent plasticity), where the weights increase if the pre-synaptic neuron fires just before the post-synaptic one, and decrease if it fires afterward.

A key feature that allows the emergence of neuronal groups is the existence of nonzero propagation times in the synapses, meaning that once a neuron fire, a post-synaptic neuron will only receive the potential change after some specified propagation time. Another important requisite for having polychronous groups is the existence of excitatory and inhibitory neurons, with different firing equations, in a way that once the global network activation rate is over a threshold, and a certain amount of excitatory neurons are activated, they reach most of the inhibitory neurons, causing the network to reset. This process iterated over time causes the network to have a rhythmic behavior.

The polychronous groups can be understood as assemblies of neurons, that fire in the same temporal pattern, repeatedly in accordance to network inputs. Groups are dynamic entities due to the plasticity model. As soon as the synaptic weights begin to

* This research was made possible by a grant from the Internship program at INRIA Saclay Île-de-France, from June to October in 2008.

change, some groups disappear and new ones emerge. If the same inputs are presented over and over, it is observed that the number of groups attain an equilibrium [Izhikevich 2006], as the network "aligns" to the given inputs. In this sense, then, groups can be understood as "memories" in the network, as the network learns to associate this groups to sets of inputs and coordinate them with related outputs.

So far, polychronous network has been studied for theoretical purposes [Izhikevich 2006] [Maier and Miller 2008] and reservoir computing [Paugam-Moisy et al. 2008] implementations. However, embedding this kind of network in an agent in a simulated virtual world had not yet been attempted.

Our contribution in this work is to simulate one such network embedded into an agent that roams a virtual environment in search of food, receiving stimuli about its own energy level and from the environment (the nearest food direction), and being able to choose the direction to turn. Without any kind of rule to indicate that the agent should feed, we study if the use of Izikevitch network, STDP plasticity and the initial conditions will influence the behavior of the agent and if some "food searching" strategy will emerge.

2. Agent Model

The simulated agent is a blue tetrahedron that starts in the center and moves about with constant speed in a flat 20 x 20 field containing 50 randomly placed small green spheres that represent food. The agent also has an energy level, that starts at 1.0 and gradually decreases towards 0.0. Nothing special happens if the energy reaches zero.

As soon as the agent touches some food, its energy level is increased and the food is replaced with another one placed at random on the field. The agent is blocked from leaving the field by invisible walls.

The agent has two sensors, the first, called **directional sensor**, indicates the angle towards the nearest food object, with a 30 degrees arc and a radius of 10 units of distance. The second sensor is called **energy sensor** and measures the current energy level of the agent. A challenging aspect of this kind of simulation is converting the inputs from the environment to neural activation patterns, and convert neural patterns into output values to command the effectors. Our approach, not detailed here, was based on the way the desert scorpion finds its prey, using 8 neurons located in the tips of its legs, that measures oscillations from waves in the sand [Stürzl 2000]. Using only 8 excitatory and 8 inhibitory neurons, in a way that each 3 inhibit the opposite one, it is possible to convert an incoming wave into a firing pattern. In our case, with a simple adaptation of the same algorithm, we convert numbers representing angle or energy levels to a linear pattern of activation of neurons representing inputs and convert timed activations into numbers in the reverse way.

To act in the world, the agent may choose the angle to turn using its **directional actuator**, that received a value from 0.0 to 1.0 indicating the turning direction, in a 30 degrees arc.

Each of the sensors and actuators comprise a defined set of neurons, usually 5 for each that can be activated in patterns depending on the inputs or the network activity.

3. Experimental Settings

To study the effect of Izhikevich network and STDP plasticity on the behavior of the agent, we define several scenarios. We want especially to attest if by setting the energy initially to its maximum value, it will induce the behavior of “feeding” in the agent, or moving toward the food. On the other hand, if by setting the agent's initial energy to the minimum value would induce the behavior of avoiding food.

Below is a list of the used simulation scenarios:

- **Rand**: as a benchmark, we simulate an agent which moves entirely at random (*Random scenario*).
- **IZN**: a random 180 neurons network is generated using Izhikevich's neurons equations, with 80% of excitatory neurons and 20% of inhibitory ones. Each neuron has 30% of chance of being connected to each one of the others. This network is simulated for 10 min simulated time. This scenario has STDP *disabled*.
- **IZS**: This scenario uses IZN as a base, however having the STDP feature *enabled*.
- **ENEMAX**: This is similar to IZS, but starts with maximum energy for 10 min, after that the energy is allowed to change normally for another 10 min.
- **ENEMIN**: To compare with ENEMAX, this scenario starts with energy fixed at zero for 10 min, then enable variation for another 10 min.

4. Results and Conclusions

Each of the experiments was run 20 times using 20 different random seeds. The same 20 seeds were used for all the scenarios, allowing similar initial food setup and energy progression. All comparisons below were done on the averages of the 20 simulations.

We observe that by enforcing initial values for a long period on the network we can direct the polychronous group selection, thus allowing some form of training for the agent.

Simply by having STDP active, the groups start to change and take shape in accordance to the inputs. For that reason it is not enough to have the energy start at maximum and then immediately begin to decrease, because as we can observe in the Rand scenario, the energy quickly tends back to low values, and the agent will end up in fact learning to remain in that low energy values.

This is why we developed the scenarios ENEMAX and ENEMIN, that enforce the “desired” energy input for an extended period, allowing the network to stabilize and select the best groups that are compatible with that energy input.

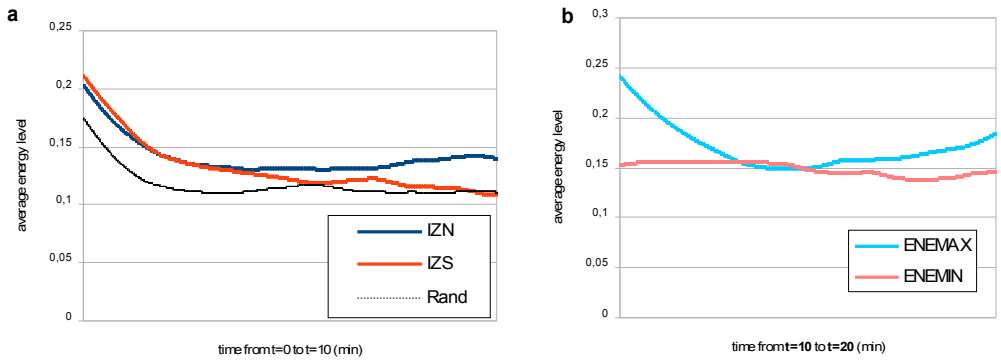


Figure 1. For each time step t, average energy level of the agent between t and the end of the simulation (end=10mn). In (a) we compare results for Rand(black, thin), IZN(blue) and with IZS(red). In (b) we compare ENEMAX (light blue) and ENEMIN(light red) scenarios, only showing the evolution of energy after it becomes variable.

With the simulation scenarios used, we can observe in Figure 1a that has STDP enabled right from the start of the simulation(IZS) will actually behave less efficiently than without STDP at all (IZN), because since the energy level will drop significantly right at start, the STDP will strengthen the polychronous groups associated to lower energy input patterns, inducing the agent to select output angles to avoid food.

On the other hand, on Figure 1b, when keeping the energy level at maximum for 10 minutes (ENEMAX), after a brief adaptation period, up to the inflection point in ENEMAX(around the middle part of the graph), the STDP reenforces groups related to full energy input patterns, in this way, most of the agent's output actions will tent to direct him towards food, so that the high energy levels can be maintained.

When the energy is kept at minimum for 10 min (ENEMIN), however, the behavior is even worse than IZS, because the network was taught to remain in lower energy states (bu reenforcing the lower energy polychronous groups), then the agent will learn to move away from food to remain in these states.

This shows that Izhikevich networks with STDP can be used as an alternative agent model, with abstract trainable inputs patterns that can direct the agent behavior.

5. Future Work

Presently, we are working on improving the performance of the simulator to be able to do even more tests and experiments to confirm our findings in larger scale networks and in other less random spatial configurations.

Another interesting possibility, aimed for future releases, is to include poisonous food (that reduces the energy level by an amount) to see how the agents react and what kind of learning behavior could be observed. If the agent learns to avoid poisonous food, we could be close to understanding how higher level concepts can be expressed within a spiked neural network.

This research continues as a collaboration between the Cognitio Research team at the Polytechnic School in University of São Paulo and researchers at LRI in INRIA Saclay Île-de-France, enabling interchange and joint work, considered relevant to both sides.

6. References

Izhikevich, E. M. 2006. Polychronization: computation with spikes. Neural Computation, 18:245-282.

Stürzl, W., Kempter, R. & van Hemmen, JL 2000. Theory of Arachnid Prey Localization. Physical Review Letters 84, pp.5668 – 5671.

Froemke, R. C. & Yang D. 2002. Spike-timing-dependent synaptic modification induced by natural spike trains. Nature Vol. 416, pp. 433-438, March 2002.

Gerstner, W. & Kistler, W. 2002. Spiking Neuron Models. Cambridge Univ. Press, Cambridge.

Paugam-Moisy, H., Martinez, R. & Bengio, S. 2008. Delay learning and polychronization for reservoir computing. Neurocomputing 71, 1143–1158

Maier, W. & Miller, B. 2008. A minimal model for the study of polychronous groups. arXiv:0806.1070v1 [cond-mat.dis-nn]

III

Impacts and Perspectives

COLIBRI 2009 • 22-23 de julho de 2009 • ISBN 978-85-7669-245-4

Gerência de Contexto e suas Aplicações

Ana Carolina Salgado[1], Flávia Maria Santoro[2], Marcos R.S. Borges[3], Patrick Brézillon[4], Renata Mendes de Araujo[2], Vaninha Vieira[5]

[1]Centro de Informática - Universidade Federal de Pernambuco
[2]Programa de Pós-Graduação em Informática – NP2Tec
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
[3]NCE/IM - Universidade Federal do Rio de Janeiro
[4]LIP6 – Université Pierre et Marie Curie (Paris 6)
[5]Centro de Processamento de Dados – Universidade Federal da Bahia

acs@cin.ufpe.br; {flavia.santoro, renata.araujo}@uniriotec.br; mborges@nce.ufrj.br; Patrick.brezillon@lip6.fr; vaninha@ufba.br

Abstract. This paper presents the research initiatives among Brazilian universities (UFRJ, UNIRIO, UFPE and UFBA) and French researchers at UPMC in the context management area.

Resumo. Este artigo apresenta as interações em pesquisa na área de gestão de contexto realizadas por grupos brasileiros (UFRJ, UNIRIO, UFPE e UFBA) e pesquisadores da UPMC, França.

1. Introdução

Contexto é o conhecimento que permite definir o que é ou não relevante em uma dada situação. Em sistemas computacionais, o contexto é uma importante ferramenta de apoio à comunicação entre os sistemas e seus usuários. As áreas da Computação Ubíqua e Inteligência Artificial foram pioneiras nos estudos e utilização do conceito de contexto [Brézillon 1999]. Pesquisas recentes vêm utilizando esse conceito para beneficiar aplicações ligadas a outras áreas como: *Sistemas Colaborativos*, visando melhorar os serviços de percepção e prover assistência aos grupos; *Hipermídia Adaptativa*, possibilitando a personalização e adaptação do conteúdo de sites web; *Integração de Dados*, facilitando a resolução de conflitos semânticos; e *Interface Humano-Computador*, para tornar mais intuitiva a interação dos usuários com os sistemas.

Brézillon e Pomerol [1999] definem contexto como "aquilo que restringe algo sem intervir nele explicitamente". Os autores consideram o que está sendo restringido como um foco para um ator. Vários elementos justificam esta definição, os três principais são (1) contexto é relativo ao foco; (2) como o foco evolui, evolui também o seu contexto; e (3) contexto é altamente dependente do domínio. Este enfoque entende que não se pode falar de contexto de uma forma abstrata e, portanto, divide contexto em conhecimento externo e conhecimento contextual [Brézillon, 2005]. Este último constitui uma espécie de reservatório onde os elementos contextuais estão diretamente relacionados com o foco, enquanto que o primeiro não possui nenhuma relação com o foco em questão. O foco evolui quando ocorre um novo evento ou como resultado de uma decisão tomada na fase anterior ao foco.

O formalismo apresentado por esta proposta e a possibilidade de aplicação de raciocínio sobre esta estrutura é o que compreende a gerência de contexto, base para o desenvolvimento das pesquisas entre grupos da UFRJ, UNIRIO, UFPE e UFBA, em

cooperação com LIP 6, da Université Pierre Marie Curie (Paris VI). Neste artigo são apresentados os trabalhos desenvolvidos por estes grupos na aplicação da gerência de contexto na gestão de conhecimento em organizações, no desenvolvimento de aplicações sensíveis ao contexto e na gestão de emergências.

2. Trabalhos desenvolvidos nas diversas áreas

Um dos maiores desafios da gestão de conhecimento em organizações não reside somente na falta de informação e sim na sua qualidade e raciocínio. As organizações têm se defrontado com a necessidade de conferir relevância à informação, mais precisamente para os diversos contextos em que cada grupo de trabalho dentro da organização está situado. Uma forma de tornar a informação relevante é associá-la ao contexto na qual foi produzida. Procedimentos não podem ser dissociados da forma como foram cumpridos na prática e das razões por trás das escolhas de acordo com as circunstâncias específicas [Santoro e Brézillon, 2005].

A transferência de conhecimento entre atores só pode ter êxito se um foco comum e seu contexto são criados e, sobretudo, compartilhados (a partir de um foco comum, participantes concordam com contexto estabelecido e compartilhado) [Araujo e Brézillon, 2005]. As informações sobre o contexto de atividades passadas pode ajudar a compreender melhor situações presentes. O entendimento das informações de contexto e como classificá-las pode ser feito a partir de um *framework* para este fim [Borges et al., 2008]. Entretanto, extrair o conhecimento contextualizado de equipes e torná-lo explícito não é uma tarefa fácil. Perret et al. [2004] e Santoro e Brézillon [2005] sugerem que uma história é uma possibilidade de registrar contexto de forma coletiva. A técnica *group storytelling* aliada a um *groupware* com funcionalidades específicas é a proposta para ajudar a re-construção de conhecimento contextualizado de um grupo.

Para apoiar o processo de transferência de informação contextual, Nunes et al. [2006] propõem um modelo para o ciclo de criação, armazenamento e reutilização de conhecimento contextual. A proposta tem como foco a atividade do processo de trabalho e o seu objetivo é criar um ambiente onde os executores destas atividades possam obter ajuda na realização das mesmas. Um repositório é organizado em uma estrutura definida que mapeia os contextos existentes nos objetos de conhecimento, como documentos, bancos de dados, endereços eletrônicos, indivíduos, grupos, e outros.

A gerência de contexto apresenta-se também como uma ferramenta importante no apoio à aprendizagem virtual colaborativa. Esse apoio pode ser oferecido por meio da análise de interações passadas, armazenadas em uma memória do grupo. Porém, sem o conhecimento do contexto em que ocorreram as interações, essa informação pode não ter muita utilidade para o aluno. Siebra et al. [2005] propõem o uso de contexto para enriquecer e qualificar o conhecimento armazenado em uma memória de interações em ambientes de aprendizagem colaborativa.

O tratamento de contexto é essencial na área de gestão de emergências. Sem contexto, a informação pode ser descartada ou mal utilizada, dado seu aspecto heterogêneo, com elementos de contexto provenientes de muitas fontes (a situação, o tipo de emergência, os atores etc). Além disso, nestes ambientes o tempo é muito curto, fazendo com que o tratamento do contexto tenha que ser feito de forma muito rápida, de preferência automática. A informação sem o contexto correspondente ou a associação da informação a contextos equivocados podem alterar as decisões e, portanto, o

processo de resposta, causando perdas humanas e materiais [Diniz et al., 2008]. Poucos sistemas, entretanto, tratam de forma sistemática a informação de contexto, tais como explorados por Brézillon et al. [2008] e Ochoa et al. [2006].

Aplicações sensíveis ao contexto são aquelas que utilizam contexto para prover informações ou serviços relevantes para a execução de uma tarefa. No entanto, compreender e identificar o contexto, e executar ações de acordo com mudanças nesse contexto não é uma tarefa trivial. É preciso lidar com questões como: que tipo de informação considerar como contexto, como representá-las, como podem ser adquiridas e gerenciadas e como serão efetivamente utilizadas para alterar o comportamento do sistema. Para resolver alguns destes desafios, Vieira et al. [2007] propõem uma infraestrutura de apoio à modelagem e gerenciamento de contexto, fundamentada no formalismo de um metamodelo de contexto [Vieira et al., 2008].

3. Perspectivas das pesquisas em contexto

Mesmo com as muitas pesquisas realizadas até o momento na área de contexto, há muito ainda a ser explorado. A infraestrutura, a inserção da informação de contexto no conhecimento e nas interações, assim como o tratamento do contexto em domínios de aplicações são algumas destas áreas. A base já consolidada permite alcançar resultados mais relevantes e de forma mais rápida.

A partir do conhecimento adquirido sobre contexto e aplicações sensíveis ao contexto, Vieira et al. [2009] propõem a formalização de atividades relacionadas à especificação, gerenciamento e uso do contexto sob a forma de um processo de software. Esse processo visa apoiar os desenvolvedores com a definição de tarefas sistemáticas, modelos e padrões. Os estudos sobre metamodelos de contexto iniciados por Brézillon [2005] e aprofundados por Vieira et al. [2008] constituem uma importante base para definir um gerenciador de informações contextuais. Esta base poderá ser utilizada em muitos tipos de aplicações para reduzir o esforço de implementação.

A área de suporte por computador ao trabalho (e aprendizado) cooperativo já demonstrou que tem muito a ganhar com o tratamento explícito do contexto nas interações sociais. Esta área é, certamente, um campo vasto para pesquisas avançadas. As redes sociais e o conhecimento coletivo são dois exemplos desta associação onde a interação entre grupos franceses e brasileiros já demonstrou avanços importantes. As áreas de suporte ao desenvolvimento de software, o desenvolvimento de software ubíquo, particularmente as aplicações que fazem uso de dispositivos móveis, são fontes de inspiração para pesquisas futuras. O domínio de gestão de emergências, um campo bastante aberto e atual, é outro exemplo em que o tratamento do contexto pode contribuir significativamente para o desenvolvimento de soluções mais eficazes.

4. Conclusão

As atividades de pesquisa e resultados apresentados neste artigo têm sido desenvolvidas como parcerias entre universidades brasileiras e o LIP6-UPMC, em caráter de estágios de doutorado, pós-doutorado e visitas temporárias de pesquisadores brasileiros à instituição francesa. Os resultados destas pesquisas têm configurado um grupo de interesse no tema específico de gestão de informação de contexto, com reflexos em outras comunidades de interesse onde os pesquisadores aqui apresentados atuam conjuntamente, como a área de Sistemas Colaborativos, bem como em áreas de

interesse individuais como Sistemas de Informação, Engenharia de Software e Banco de Dados. Os trabalhos iniciados por estes grupos têm encaminhamentos na forma de projetos de mestrado e doutorado e projetos de pesquisa nas instituições envolvidas.

Referências

Araujo, R. M., Brézillon, P. Reinforcing Shared Context to Improve Collaboration, Revue d´Intelligence Artificielle, França, 19(3), p. 537-556, 2005.

Borges, M. R. S., Brézillon, P., Pino, J. A., Pomerol, J-Ch. Groupware System Design and the Context Concept. In: 9th International Conference on Computer Supported Cooperative Work in Design (CSWD'2005), p. 45-54, Coventry, UK, 2005.

Borges, M. R. S., Brézillon, P., Pino, J. A., Pomerol, J-Ch. Dealing with the Effects of Context Mismatch in Group Work. Decision Support Systems, v. 43, p. 1692-1706, 2007.

Brézillon, P., Borges, M. R. S., Pino, J.A., Pomerol, J-Ch. Context and Awareness in Group Work: Lessons Learned from Three Case Studies. Journal of Decision Systems, v. 17, p. 27-40, 2008.

Brézillon, P. Modeling users' practices in contextual graphs. In: A. David (Ed.) Organisation des Connaissances dans les Systèmes d'Informations Orientés Utilisation. Contexte de Veille et d'Intelligence Economique (ISKO-2005). Presse Universitaire de Nancy, pp. 77-96, 2005.

Brézillon P., Pomerol J-Ch. Contextual knowledge sharing and cooperation in intelligent assistant systems. Le Travail Humain, 62(3): 223-246, 1999.

Diniz, V. B., Borges, M. R. S., Gomes, J.O., Canós, J.H. Decision Making Support in Emergency Response. In: Frédéric Adam; Patrick Humphreys. (Org.). Encyclopedia of Decision Making and Decision Support Technologies. 1 ed. London: Information Science Reference (an imprint of IGI Global), v. 1, p. 184-191, 2008.

Nunes, V.T., Santoro, F. M., Borges, M. R. S., Context in Knowledge-Intensive Collaborative Work. In: 10th International Conference on Computer Supported Cooperative Work in Design (CSCWD'2006), Nanjing, China, v.2, p. 793-798, 2006.

Ochoa, S.F., Pino, J.A., Neyem, A., Borges, M. R. S. Using Context to Support Group Decision Making in Disasters Affecting Urban Areas. In: IFIP International Conference on Decision Support Systems, Londres, v.2, p. 546-561, 2006.

Perret, R., Borges, M.R.S., Santoro, F.M. Applying Group Storytelling in Knowledge Management. In: International Workshop on Groupware, San Carlos, Costa Rica, v.1, p. 141-148, 2004.

Santoro, F. M., Brezillon P. Towards the Narrative Approach to Collect Group Knowledge and Context, In: Workshop Context and Groupware in conjunction with the Fifth International and Interdisciplinary Conference on Modeling and Using Context, France, 2005.

Siebra, S.A.; Salgado, A.C.; Tedesco, P.; Brézillon, P. Identifying the Interaction Context in CSCLE. In: Fifth International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT-05), LNCS 3554, p. 464-475, Paris, France, 2005.

Vieira, V., Brézillon, P., Salgado, A. C., Tedesco, P. A Context-Oriented Model for Domain-Independent Context Management, Revue d'Intelligence Artificiel, 22(5), p. 609-627, 2008.

Vieira, V., Tedesco, P., Salgado, A. C. A Process for the Design of Context-Sensitive Systems, In: 13th International Conference on Computer Supported Cooperative Work in Design (CSCWD'2009), Santiago, Chile, 2009 (a ser publicado).

Vieira, V., Tedesco, P., Salgado, A. C., Brézillon, P. Investigating the Specificities of Contextual Elements Management: The CEManTIKA Approach. In: Fifth International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT-07), LNAI 4635, p. 493-506, Roskilde, Denmark, 2007.

Integração e Interoperabilidade de Conteúdos em Portais Semânticos

Ana Maria de C. Moura[1] e Maria Cláudia Cavalcanti[2]

[1]Coordenação de Ciência da Computação
Laboratório Nacional de Computação Científica
Av. Getúlio Vargas 333 – Quitandinha - Rio de Janeiro

[2]Departamento de Engenharia de Computação
Instituto Militar de Engenharia
Praça General Tibúrcio 80 - Praia Vermelha - Rio de Janeiro
{anamoura, yoko}@ime.eb.br

Abstract

E-gov portals are a practical and effective way to disseminate information on the Web. However, most portals are built as independent sites, lacking support to deal efficiently with processes and services, and to provide an integrated view of similar and complementary information extracted from other institutional portals. From the citizen's side, searching information in these portals may become a hard task, since it is spread out along innumerous sites. This work proposes the development of enriched e-gov semantic portals, in order to facilitate citizen´s navigation, by integrating content extracted from other sites, considering similar or complementary domains. This work focuses on interdisciplinary work involving semantic web, and will certainly benefit from a research cooperation with INRIA, since this institute develops important projects on semantic web and maintains effective collaboration with the W3C working group.

1. INTRODUÇÃO

Nos últimos anos as organizações corporativas e setores públicos de vários países vêm sofrendo fortes pressões quanto à necessidade de revitalizarem suas administrações, tornando-as mais pró-ativas, integradas, eficientes, transparentes e, especialmente, mais orientadas a serviços. Apesar do grande desafio que essa mudança de paradigma representa, estes setores vêm introduzindo gradativamente inovações em sua estrutura organizacional de forma a mobilizar, implantar e utilizar o capital humano e de informação, recursos tecnológicos e financeiros para a entrega de serviços aos cidadãos em todo o mundo.

Nesse contexto, as iniciativas de governo eletrônico (e-Gov) referem-se ao uso da Web e outras tecnologias de informação pelo corpo governamental para interação com seus cidadãos. A prática de gestão do conhecimento aliada à adequada utilização das Tecnologias de Informação e Comunicação (TIC) desempenha um papel crucial em iniciativas privadas e públicas, tornando-as mais eficientes e competitivas. Um ótimo exemplo de iniciativa pública no Brasil é a arquitetura do e-Ping[1] (Padrões de Interoperabilidade de Governo Eletrônico), cujo objetivo é definir um conjunto mínimo de premissas, políticas e especificações técnicas que regulamentam a utilização

[1]http://www.governoeletronico.gov.br/acoes-e-projetos/e-ping-padroes-deinteroperabilidade.

de TIC no governo federal, estabelecendo as condições de interação com os demais poderes e esferas de governo e com a sociedade em geral. Nesta arquitetura, fica evidenciada a forte preocupação em oferecer portais intuitivos e de fácil acesso ao cidadão em todas as esferas do governo. Os portais oferecidos pelo governo, podem ser categorizados como: Governo para Cidadãos (G2C), Governo para Negócios (G2B), Governo para Empregados (G2E) e Governo para Governo (G2G) [13].

Apesar do grande avanço das iniciativas públicas e privadas no uso de várias modalidades das TICs na construção de seus artefatos, raras são as iniciativas que consideram a Web Semântica como uma forte tecnologia de integração e reuso de recursos. Grande parte dos portais governamentais disponibilizados até o momento não consideram o uso da Web semântica.

Com relação às iniciativas das diferentes instituições governamentais no Brasil observa-se que: nos diferentes níveis de governo, existe uma grande disparidade no grau de informatização dos serviços oferecidos, incluindo-se órgãos com apenas uma tênue ou nenhuma referência na Internet; existe, até o momento, uma grande independência e diversidade entre os órgãos que oferecem informações e serviços via Web; muito pouco existe em termos de suporte a processos ou serviços envolvendo mais de uma instituição. O mais usual é que estes se limitem às fronteiras de um único órgão governamental; em geral, cria-se um portal de referência às informações dos diversos órgãos do governo, mas que, longe de oferecer uma visão integrada destas informações, apenas se apresenta como um serviço de páginas amarelas de instituições[2], cabendo ao usuário navegar por estas instituições em busca das informações de seu interesse; na maior parte dos sites governamentais, não é fácil a localização de informações, sendo usual o usuário percorrer um longo caminho até encontrar o que deseja; com freqüência, as informações são disponibilizadas segundo níveis de detalhe, unidades e formatos bastante diversos, existindo, em geral, pouca ou nenhuma facilidade para que o usuário possa adequar o resultado de sua busca às suas necessidades.

Neste contexto, uma das poucas iniciativas no sentido de oferecer portais semânticos do tipo G2C, encontra-se em andamento no IME-RJ, um projeto de pesquisa que tem como objetivo demonstrar que o uso de tecnologias da Web Semântica contribui sobremaneira para uma melhoria na organização, integração e gestão das informações em portais. Em [7] [8] foi desenvolvida uma arquitetura que serviu como infra-estrutura básica para a construção de portais semânticos, capaz de integrar e instanciar informações a partir do uso intensivo de ontologias. Essa arquitetura permite que informações distribuídas na Web sejam recuperadas com base em uma ontologia de domínio e, através de mapeamentos definidos entre essas informações e a ontologia de domínio, novos conteúdos sejam dinamicamente categorizados e adicionados ao portal. De forma a evidenciar a aplicabilidade dessa arquitetura, foi desenvolvido o POrtal SEmântico EDUcacional (POSEDU[3]), [9], voltado para o domínio educacional.

No entanto, ao longo desse trabalho, observou-se um número muito restrito de ontologias na Web, mesmo em domínios de grande relevância, a exemplo de educação. A grande maioria dos portais educacionais das universidades e centros de pesquisa tanto no Brasil quanto no exterior não são portais semânticos. Verificou-se também que algumas instituições de ensino, a exemplo das universidades de Lehigh[4] (Pensilvânia, Estados Unidos) e Munique[5] (Alemanha), apresentam suas estruturas organizacionais através de taxonomias representadas em OWL, porém com poucas ou quase nenhuma instância. Portanto, observa-se que o uso de ontologias e ferramentas de anotação ainda é bastante insipiente.

Além disso, a informação oriunda da Web profunda não é contemplada pelos serviços de busca oferecidos pelos portais, uma vez que estes conteúdos não detectados por robôs, a exemplo de formulários HTML ou banco de dados. Diversos trabalhos têm sido realizados com o objetivo de definir mecanismos e algoritmos visando extrair informações (estruturadas ou não) relevantes da Web profunda e de sites sociais [1] [3].

2. PROPOSTA E RELEVÂNCIA DO PROJETO

O presente projeto se propõe a explorar e desenvolver um conjunto de ferramentas que permitam integrar conteúdos de portais, de modo a facilitar a navegação entre portais, e com isso contribuir para melhorar a qualidade de buscas e serviços em atendimento às demandas do cidadão, levando-se em conta a sua diversidade e diferenças culturais, regionais e sócio-econômicas.

A partir da experiência obtida em [8], é possível destacar algumas questões importantes, ainda em aberto, no cenário atual de pesquisa em portais semânticos, que o projeto em questão pretende responder:

[2] www.redegoverno.gov.br

[3] www.comp.ime.eb.br/~posedu

[4] http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl

[5] http://wwwradig.in.tum.de/ontology/organisation

- Como integrar conteúdos em portais G2C sobre vários domínios do conhecimento, a partir de sites e portais tradicionais desprovidos de metadados ou anotações semânticas?
- Como anexar um mecanismo de busca ao portal de forma a integrar conteúdos extraídos da Web profunda?
- Como construir ontologias de domínio a partir da organização e/ou estruturação de sites convencionais?
- Os algoritmos existentes na literatura para realização de casamento (*matching*) entre ontologias são eficientes para selecionar e classificar conteúdos em portais? Em [4] é feita uma comparação entre essas ferramentas, porém em aplicações e uso distintos dos pretendidos no escopo do trabalho em questão;
- Seria viável a construção de um repositório público de metadados com a descrição e localização de ontologias em diversos domínios, para uso em portais semânticos?

A relevância da pesquisa vislumbrada por essa proposta é justificada pelos seguintes fatores:

- Considerando o volume crescente de informações sendo disponibilizadas via Web pelas diversas instituições governamentais, é imperioso que se invista em iniciativas que tenham por objetivo a gerência e integração de informações provenientes de diferentes fontes, de modo a agilizar a utilização deste vasto acervo;
- Este trabalho está em consonância com dois dos Grandes Desafios da Pesquisa em Computação no Brasil[6]: gestão de informação em grandes volumes de dados multimídia distribuídos e acesso participativo e universal do cidadão brasileiro ao conhecimento;
- As questões abordadas por esse projeto são também alvo de pesquisa de equipes do INRIA, notadamente dos projetos WAM[7], GEMO[8] e EDELWEISS[9]. O primeiro foi responsável pelo desenvolvimento da ferramenta de anotação Amaya, em colaboração com a equipe W3C. O segundo estuda problemas ligados à mediação e integração de dados heterogêneos representados em XML, linguagem padrão da Web. Já o terceiro está ligado diretamente ao estudo de problemas ligados à interoperabilidade e contextualização semântica de informações, anotações semânticas de recursos de informação e interfaces Web baseadas em ontologias. Uma parceria internacional com um desses grupos seria de grande interesse científico para esse projeto, e permitiria ampliar e testar as técnicas desenvolvidas na equipe brasileira num outro contexto social.

[6] http://www.sbc.org.br/)
[7] http://www.inria.fr/recherche/equipes/wam.fr.html
[8] http://www.inria.fr/recherche/equipes/gemo.fr.html
[9] http://www.inria.fr/recherche/equipes/edelweiss.fr.html

3. OBJETIVOS E PLANO DE TRABALHO

Este projeto visa contribuir com a gestão dinâmica de informações em portais semânticos, por meio de técnicas e recursos da Web Semântica, notadamente no uso de ontologias para fins de descoberta, extração, classificação e integração de conteúdos de outros sites. Também é objetivo estratégico deste projeto integrar as proponentes a projetos de cooperação de natureza similar ou complementar com outras instituições.

Esse projeto será desenvolvido no período de dois anos, a partir de tarefas específicas, descritas a seguir.

i) Realizar o levantamento detalhado de requisitos do projeto, e definir o seu escopo e domínio de aplicação;
ii) Estudar e definir algoritmos e ferramentas específicos que poderão ser úteis para responder às questões identificadas na seção 2. Essa etapa deverá ser acompanhada de testes que comprovem a eficácia dos mecanismos e como poderão ser aplicados ao trabalho proposto;
iii) Desenvolver mecanismos para compor um portal semântico no domínio de aplicação escolhido, de modo a prover todas as facilidades especificadas na seção 2;
iv) Testar e validar as ferramentas e técnicas desenvolvidas, através da disponiblização de um portal semântico do tipo G2C.

5. EXPERIÊNCIA ANTERIOR

As professoras proponentes têm se dedicado ativamente no desenvolvimento de atividades de pesquisa e à preparação de material didático na área de Web Semântica.

A professora Ana Maria, nos últimos 10 anos de sua atuação no IME, foi responsável pela orientação de mais de 15 dissertações de Mestrado, com resultados publicados em periódicos, conferências nacionais e internacionais, além de ter participado e coordenado projetos de pesquisa na área. Dentre esses, destacam-se as teses orientadas e trabalhos realizados na linha de portais semânticos, a saber [12], [6], [5]. Trabalhou também durante três anos junto ao Proderj (Centro de Tecnologia da Informação e Comunicação do Estado do Rio de Janeiro), onde realizou atividades de consultoria e planejamento em governo eletrônico, e atualmente é pesquisadora junto ao LNCC.

A Profª Maria Cláudia vem trabalhando em conjunto com a Profª Ana Maria há mais de 4 anos, tendo participado das atividades e orientações relacionadas ao tema proposto. Além dos trabalhos já citados, vale destacar ainda o trabalho na linha de integração de ontologias [2]. Junto a grupos de pesquisa da Fiocruz, tem participado de outras iniciativas, envolvendo o uso, mapeamento e evolução de ontologias na área de Bioinformática.

Além disso, as proponentes participaram de experiências anteriores bem sucedidas com a França, através dos projetos ECOBASE[10] e KIWI[11], envolvendo o CNPq e INRIA. O projeto Ecobase teve como foco central o desenvolvimento de tecnologias para a integração de informações em sistemas ambientais. Já o projeto KIWI aproveitou a experiência adquirida anteriormente para definir e desenvolver ferramentas para a integração de informações na Web, cuja tecnologia envolveu o uso intensivo de metadados, gerência de workflows, gerência de dados semi-estruturados, técnicas de replicação e de processamento de consultas. Estas parcerias resultaram numa troca frutífera de conhecimento, com o intercâmbio de alunos e publicações [10] [11].

Portanto, os resultados obtidos ao longo da vida acadêmica e de pesquisa das proponentes demonstram a experiência e o amadurecimento requeridos para a realização de um projeto dessa natureza.

6. RESULTADOS ESPERADOS

Estes resultados de pesquisa serão produzidos no contexto de teses de mestrado e trabalhos de final de curso, que poderão ser orientadas conjuntamente por pesquisadores brasileiros e franceses. Adicionalmente, planeja-se produzir artigos de referência ao longo da pesquisa, publicados em periódicos e conferências internacionais. Os protótipos e técnicas desenvolvidas poderão ser utilizados pela comunidade acadêmica e instituições governamentais em geral, no sentido de facilitar o desenvolvimento de aplicações de governo eletrônico. Por fim, uma contribuição adicional deste projeto é a criação de um repositório de metadados sobre os sites e portais do domínio de aplicação estudados.

REFERENCES

[1] S. Amer-Yahia, L. Lakshmanan, C. Yu. SocialScope: Enabling Information Discovery on Social Content Sites (2009). 4th Bienal Conference on innovative data systems (CIDR), Asilomar, CA, EU, jan.

[2] H. Camargo, AM Moura, M C Cavalcanti. Ontologias Emergentes: Uma Nova Abordagem para Integração de Ontologias. XXIII SBBD, Campinas, S.P, out 2008

[3] M. Cafarella (2009). Extracting and Querying a Comprehensive web Database. 4th Bienal Conference on innovative data systems (CIDR), CA, EU, jan.

[4] J. Euzenat and P. Shvaiko. Ontology Matching. Springer-Verlag, Berlin Heidelberg, 2007.

[5] Ferreira e Gama (2008). Organização Automática de Páginas Web para Exibição em Portais Semânticos. Projeto de final de curso em Eng. da Computação, IME, agosto.

[6] F. A. Lachtim. Organization and Instantiation of Content in Semantic Portals (in Portuguese). *Master thesis*, IME, Rio de Janeiro, Brazil, May 2008.

[7] F. A. Lachtim, A. M. C. Moura, M. C. Cavalcanti. An Architecture for Dynamic Organization and Publication in Semantic Portals. *10th International Conference on Information Integration and Web-Based Applications and Services (IIWAS)*, Linz, Austria, Nov. 2008.

[8] F. Lachtim, A. M. Moura, M. C. Cavalcanti (2009). "Ontology Matching for Dynamic Publication into Semantic Portals," to appear.

[9] F. A. Lachtim, G. Ferreira, R. Gama, A. M. C. Moura, M. C. Cavalcanti (2008). POSEDU: um Portal Semântico Educacional *2nd Brazilian Workshop Semantic Web and Education*, Fortaleza, Brazil, Nov.

[10] L. Bouganim, M. C. Cavalcanti A. M. Moura, et al. (2000). The Ecobase environmental information system: applications, architecture and open issues. Network and Information Systems Journal (NISJ), Hermes, ISSN 1290-2926, Vol 3, No 5.

[11] L. Bouganim, M. C. Cavalcanti A. M. Moura, et al. (2001).The Ecobase Project: Database and Web Technologies for Environmental Information Systems. ACM Sigmod Record, 30(3), ISSN: 0163-5808, set.

[12] W. A. Pinheiro, Moura A.M.C. An Ontology Based-Approach for Semantic Search in Portals. *Proc of the WEBS* 2004, Saragoza – Spain, Sept.

[13] A. Saldhana (2007). Secure E-Government Portals - Building a web of trust and convenience for global citizens. *W3C Workshop on e-Gov and the Web, National Academy of Sciences, Washington DC.*

[10] Ecobase – "Database and Web Technologies for Environmental Information Systems", CNPq/INRIA Research Project, 1998.

[11] KIWI: Key Technologies for the Integration of Web Information, CNPq/INRIA Research Project, 2000.

Vote électronique: informatiser au service des électeurs

Chantal Enguehard[1], Walter de Abreu Cybis[2], Gabriel Michel[3]

[1]LINA - UMR CNRS 6241
2, rue de la Houssinière, BP 92208, 44322 Nantes Cedex 03 – France

[2]Ecole Polytechnique de Montréal
2500, chemin de Polytechnique, C.P. 6079, succ. Centre-ville - Montréal - Canada.

[3]ETIC - Université Paul Verlaine.
Ile du Saulcy, 57045, Metz Cedex01– France

chantal.Enguehard@univ-nantes.fr, walter.cybis@polymtl.ca,
gabriel.michel@univ-metz.fr

Abstract. *Based on the results of research activities accomplished in France and in Brazil, this paper proposes some reflections about the overall electoral process automation. Poll offices organization, ergonomics and accessibility, counting process's transparency are important research directions to follow in order to get electoral automation profitable to voters. This paper offers some advising associated to these points.*

Résumé. *À partir des résultats d'activités de recherche menées en France et au Brésil, cet article propose quelques réflexions sur l'informatisation du processus électoral. L'organisation des bureaux de votes, leurs ergonomie et accessibilité, la transparence du processus de comptage des votes sont des axes de recherche peu explorés pour que l'informatisation électorale soit au service des électeurs. Cet article avance quelques pistes associées à ces axes de recherche.*

1. Introduction

Le vote électronique se présente sous différentes formes selon qu'il s'agit d'un vote en milieu contrôlé ou par correspondance, et selon que les dispositifs sont reliés ou pas à un réseau. Nous traitons ici des ordinateurs de vote de type DRE (Direct-Recording Electronic) non connectés à un réseau et équipant les bureaux de vote.

Tous les constructeurs de ces systèmes affirment que ces dispositifs facilitent l'exercice du droit de vote, favorisent le vote des personnes exclues des technologies (personnes handicapées, seniors et plus globalement celles qui ont des problèmes avec les technologies) et fournissent des résultats précis et justes. Nous examinerons ces questions.

2. L'évaluation de la justesse

Le caractère démocratique d'une élection peut être établi selon le respect de plusieurs principes, en voici quelques-uns : chaque électeur doit pouvoir s'exprimer sans subir de pressions (confidentialité), son vote doit rester secret (anonymat), il doit pouvoir exercer un contrôle direct sur le processus, depuis la collecte des voix jusqu'à la proclamation des résultats, afin de lui accorder sa confiance (transparence).

Lors d'un vote électronique, les choix des électeurs sont confiés à un processus

automatisé qui fournit des résultats à la clôture de l'élection. Ce processus est, par nature, opaque : le choix de chaque électeur, initialement exprimé par une pression de quelques joules sur un bouton, est transformé en une impulsion électrique, puis codé en une série de bits qui sera ajoutée à d'autres choix ainsi transformés. La nature électronique de ces multiples mutations les rend inobservables pour l'œil humain.

Il ne peut être exclu qu'une erreur d'exécution des programmes, ou une fraude, modifie des votes lors de leur enregistrement, d'un traitement intermédiaire, ou de leur addition. La vérification de la justesse des résultats énoncés par le dispositif électronique apparaît indispensable pour entretenir la confiance des électeurs.

Il existe trois voies d'évaluation de la justesse des résultats :

— *la vérification par approximation* : Les résultats sont comparés à des estimations statistiques concurrentes au vote (sondages de sortie d'urnes). Il est alors demandé à un échantillon de l'ensemble des électeurs de révéler son choix. Cette procédure présente des biais : l'échantillon, de taille réduite, ne représente pas tous les segments de la population ; le vote étant révélé, des personnes peuvent ne pas répondre à la question posée, ou mentir ; etc.

Les résultats peuvent aussi être comparés à des estimations statistiques antérieures au vote (intentions de vote pondérées, résultats lors de précédentes élections). Les intentions de vote pondérées connaissent les mêmes biais que les sondages de sorties d'urnes, auxquels il faut ajouter la possibilité que des électeurs changent d'avis entre le moment de sondage et le jour de l'élection. La conformité des résultats d'un vote avec les résultats obtenus lors d'une précédente élection ne peut pas non plus constituer une procédure de vérification acceptable car elle nie la possibilité d'alternance politique qui est le fondement de toute démocratie.

— *les garanties apportées par les traitements* : Cette démarche exige de prouver que le système réalise correctement les traitements attendus. Il doit donc être immune de faute et inaltérable. Il apparaît difficile de déterminer qu'un système est immune de fautes alors qu'écrire un programme sans aucun bug reste considéré comme un exploit. La démarche de preuve formelle s'attache à résoudre cette question, mais elle reste cantonnée à quelques briques logicielles et ne peut prendre en compte tous les composants intervenant dans un système de vote électronique (drivers, micro-code, compilateurs, etc.). Enfin, un système de vote peut être modifié volontairement de manière très subtile (branchement d'une imprimante par exemple), ou involontairement (erreur harware) sans que ces modifications, pourtant susceptibles d'altérer les traitements appliqués aux choix des électeurs, ne soient remarquées.

— *la preuve de résultat* : Une preuve de résultat est établie en calculant le résultat d'un système à partir de ses entrées, indépendamment du système à évaluer. Cette démarche est impraticable pour un ordinateur de vote de type DRE car les entrées du vote, dématérialisées, ne sont pas collectées indépendamment du système de vote et le respect de l'anonymat interdit de tracer le processus de vote de bout en bout afin de les reconstituer. Il est donc impossible de confronter les résultats énoncés aux votes qui ont été émis et dont la teneur reste inconnue.

Il existe une nouvelle génération d'ordinateurs de vote dits VVAT (Voter Verifiable Audit Trail - utilisés aux Etats Unis et au Venezuela.) dont le fonctionnement prévoit que chaque vote enregistré est également imprimé et collecté dans une urne à des fins d'éventuelles vérifications. On prévoit cependant que seuls quelques dispositifs soient vérifiés afin de valider l'ensemble du parc d'ordinateurs de

vote mis en service. Cette démarche présente plusieurs défauts rédhibitoires : la sélection d'un échantillonnage en vue de la vérification de certains paramètres, courante en contrôle qualité, n'est valide que si les matériels testés ont subi exactement les mêmes traitements, ce qui n'est pas le cas d'ordinateurs de vote ayant servi pendant une ou plusieurs journées de vote ; le décompte différé des bulletins collectés facilite les atteintes à leur intégrité ; juridiquement, prouver le dysfonctionnement des ordinateurs équipant quelques bureaux de vote ne constitue pas une preuve du dysfonctionnement des ordinateurs dont le décompte n'a pas été formellement vérifié (Enguehard 2007). Le raisonnement est le même en ce qui concerne le comptage automatique de bulletins scannés ou munis de codes à barres.

Il apparaît donc que le comptage automatique des voix est un processus dont les résultats ne sont pas vérifiés, même s'ils sont théoriquement vérifiables.

L'utilisation de logiciels libres ne constitue pas une solution : l'ensemble du système de vote (systèmes d'exploitation, compilateur) ne peut être scruté dans sa totalité ; le système n'est pas stable (brancher un périphérique - imprimante, casque audio, etc. - le modifie) ; les électeurs et les partis politiques sont démunis des compétences nécessaires, comme l'ont remarqué les observateurs des élections estoniennes de 2007 [OSCE/ODIHR 2007].

3. L'utilisabilité et l'accessibilité des ordinateurs de vote en France

L'accessibilité caractérise la capacité des personnes handicapées à percevoir une interface, à la comprendre, à l'utiliser pour naviguer et interagir [W3C 2005]. L'utilisabilité est définie par la Norme ISO 9241:11 comme « le degré selon lequel un produit peut être utilisé, par des utilisateurs identifiés, pour atteindre des buts définis avec efficacité, efficience et satisfaction, dans un contexte d'utilisation spécifié ». Dans les cas où cette population est constituée de personnes handicapées, l'utilisabilité et l'accessibilité se confondent.

Une évaluation ergonomique des trois ordinateurs de vote utilisés en France a été effectuée en deux étapes en 2007 : une évaluation de la conformité à des recommandations puis des tests utilisateurs menés auprès de personnes aveugles et de personnes vieillissantes. L'évaluation des interfaces a été réalisée par des experts en ergonomie et en accessibilité sur la base d'un ensemble de critères d'évaluation du vote électronique [MICHEL et al. 2008]. Elle a identifié des problèmes importants de ces ordinateurs de vote tels que des problèmes de compatibilité et de flexibilité modale qui ne permettent pas une accessibilité aux différentes catégories d'exclus, des problèmes de compatibilité avec les connaissances informatiques et les outils technologiques de l'électeur, des problèmes de lisibilité et de correction intuitive des erreurs. Des tests utilisateurs menés sur un de ces ordinateurs ont donné les résultats suivants :

— Sur 7 seniors (faisant partie des catégories socioprofessionnelles favorisées), placés dans des conditions idéales (pas de bruit, de pression, pas de contraintes de temps), seuls 3 ont réussi à voter. Sur les 4 échecs, 3 correspondent à une erreur de vote (vote pour un autre candidat), le quatrième n'ayant pas réussi à voter. La durée moyenne de vote a été 3,5 fois plus longue que pour le groupe de contrôle.

— Sur 5 déficients visuels (dont 3 peu habitués à la technologie) un seul a réussi à voter. Les 4 autres ont dû abandonner car l'interface vocale était trop complexe (le temps de vote moyen était en moyenne dix fois supérieur de celui du groupe de contrôle).

Cette étude a mis en évidence que ces trois ordinateurs de votes constituaient un facteur important de discrimination [Michel et al., 2008].

4. L'utilisabilité et l'accessibilité de l'ordinateur de vote au Brésil

Une série d'études concernant l'ergonomie et l'accessibilité de l'ordinateur de vote utilisé au Brésil a été accomplie en 1997 [Cybis et al., 2009]. L'évaluation heuristique de son interface a mis en évidence des problèmes ergonomiques tels que ; guidage affaibli ; manque de feedback global sur les contenus des deux votes accomplis (conseillers et maire) ; manque de protection contre les erreurs, etc. Les tests d'utilisabilité ont révélé des problèmes d'accessibilité majeurs des groupes sensibles à l'exclusion technologique (aveugles et seniors) : la majorité des participants (13 sur 20) n'a pas conclu les deux votes. Cette série d'études s'est achevée par une analyse comparative des résultats des élections de 1992 et de 1996 pour sept villes de l'état de Santa Catarina au sud du Brésil (900 000 électeurs). Les élections étaient informatisées pour 3 d'entre elles (400 000 électeurs). On a pu constater des tendances opposées en ce qui concerne les votes blancs (-41,7% contre +33,4%) et pour les partis politiques (+65,7% contre -55,5%) entre 1992 et 1996 pour les villes informatisées et non informatisées. Ces mêmes auteurs ont étudié l'évolution de l'interface de ce dispositif. Ils ont jugé timides et inefficaces les modifications mises en place sur l'interface de l'ordinateur de vote brésilien au fil des années. Il semble évident que le niveau d'interférence de ce dispositif sur les intentions de vote des électeurs sensibles aux NTIC a encore augmenté lors des élections présidentielles de 2002 et 2006.

5. Propositions

Nous avons démontré que les résultats des ordinateurs de type DRE sont invérifiables, et que, quand un dispositif de vérification est installé, cette possibilité n'est finalement pas correctement mise en oeuvre. Il apparaît donc que l'utilisation de systèmes de vote électronique effectuant automatiquement le comptage des voix érode la confiance des électeurs et, finalement la légitimité des candidats. Inversement, améliorer la transparence directe du processus électoral permet de renforcer la confiance des électeurs.

Nous proposons donc d'équiper les bureaux de vote de systèmes susceptibles de renforcer la sécurité, l'accessibilité et la transparence directe. Il convient de ne dématérialiser les bulletins de vote (faille bien identifiée des systèmes DRE), de systématiser l'usage d'urnes transparentes et le comptage humain des bulletins. Des progrès sont envisagables en ce qui concerne l'accessibilité.

En France et au Brésil les usages sont différents. En France, on vote en choisissant un bulletin parmi plusieurs. Au Brésil, on vote avec des bulletins australiens (cases à cocher ou saisie d'un identifiant numérique). Ces deux cas seront distingués dans les trois propositions exposées ci-dessous. Ces trois propositions sont le résultat d'une étude préliminaire. Des travaux complémentaires devraient permettre de conceptualiser de nouvelles avancées, en particulier en exploitant les opportunités offertes par la reconnaissance de la parole.

5.1 - Présentation des candidats

En France, les bulletins sont fournis par les candidats. Les candidats non affiliés à de grands partis politiques ne sont pas toujours en mesure de fournir des bulletins dans tous les bureaux de vote du pays, ce qui les désavantage (lors des élections européennes 2009

certains candidats proposaient aux électeurs d'imprimer leur bulletin de vote mis à disposition sur internet).

Les bureaux de vote peuvent être munis d'imprimante permettant d'imprimer quelques exemplaires des bulletins de vote des candidats non soutenus par les partis politiques afin que tous les choix soient proposés dans tous les bureaux de vote.

5.2 - vote des handicapés

a - confidentialité pour les aveugles et mal-voyants

Il existe des dispositifs de lecture capables de lire un texte et de le restituer grâce à la synthèse de la parole.

En France ce dispositif peut être utilisé directement dans l'isoloir (avec un casque audio pour la confidentialité) pour choisir le bulletin de vote qui porte le nom du candidat ou de la liste choisie.

Au Brésil ce dispositif pourrait être adapté pour que les aveugles puissent remplir puis relire leur bulletin de vote seuls et sans assistance.

b - confidentialité pour les handicapés moteurs voyants

Les handicapés moteurs peuvent voter de manière confidentielle en se faisant assister d'une personne rendue aveugle pour l'occasion.

En France, cette personne présente les bulletins aux handicapés, le handicapé signifie son choix simplement en disant : "non, ce n'est pas ce bulletin" ou "oui, c'est le bulletin que je vois que je choisis". Il est en mesure de contrôler que c'est bien le bulletin de son choix qui est glissé dans l'enveloppe ensuite déposée sur ses genoux et qui finira dans l'urne.

Au Brésil, à cause de l'usage du bulletin australien, il faut munir les isoloirs d'un dispositif présentant à l'écran les choix dans un ordre aléatoire et permettant de cocher des cases ou de saisir des numéros par un système capable de reconnaissance vocale ou manipulé par une personne rendue aveugle. Le bulletin, imprimé, peut alors être mis sous enveloppe et déposé dans l'urne en toute confidentialité. Ce dispositif doit être utilisé par d'autres électeurs (pas forcément tous) pour que les votes des handicapés ne puissent être systématiquement distingués des autres. Plusieurs bulletins peuvent être imprimés avant le choix définitif

Ces dispositifs pour les handicapés ne sont pas d'un usage obligatoire. Ils ne sont donc pas susceptibles de générer files d'attente et stress peu compatibles avec une sereine prise de décision.

5.3 - centralisation des résultats

Cette proposition vise à rendre transparent le processus de centralisation des résultats.

Les résultats détaillés des élections pourraient être publiés sur les sites des communes en présentant un niveau de détail homogène

— présenter les résultats détaillés par bureaux de vote

— pour chaque bureau de vote, présenter le nombre de votes obtenus par chaque candidat/liste, ainsi que le nombre de votes blancs ou nuls, mais aussi le nombre d'inscrits, le nombre d'émargements et le nombre de procurations.

Des facilités doivent être offertes afin de pouvoir disposer facilement de l'ensemble de tous les résultats détaillés.

6. Conclusions

L'examen des principaux ordinateurs de vote employés dans le monde (Brésil, France, EUA, Inde, Venezuela, Suisse) confirme les analyses théoriques que nous avons menées. Nous constatons que ces dispositifs sont difficiles à utiliser pour de nombreux handicapés, exclus de la technologie et seniors, qu'ils semblent introduire des biais importants dans les résultats des élections et que le contrôle populaire de la sincérité des élections ne peut plus être mis en oeuvre.

Nous proposons une approche nouvelle pour informatiser les bureaux de vote. Il s'agit de conserver les points positifs du vote avec urne et bulletins (en particulier la transparence du décompte manuel des bulletins de vote) et de suivre les recommandations et critères ergonomiques d'accessibilité qui avaient été établis pour les ordinateurs de vote [Cybis et al., 2009] afin d'améliorer les différentes phases : choix de candidat, expression du vote. L'objectif est de permettre à tous les électeurs (y compris les handicapés, les seniors et les exclus de la technologie) de voter en confidentialité et d'améliorer la transparence du système électoral pour tous.

Ce renouveau des bureaux de vote doit prendre en compte les différents contextes législatifs (bulletin unique comme en France ou bulletin australien avec des cases à marquer et des saisies de chiffres) et sociaux (sécurisation des bureaux de vote dans les pays où les élections donnent lieu à des fraudes). Il existe des pistes pour améliorer, pour tous, la sécurité, l'ergonomie, l'accessibilité des bureaux de vote et le droit à voter en confidentialité sans pour autant dégrader la nécessaire transparence des élections sur laquelle se fonde la légitimité des élus.

Des pistes de recherches de partenaires académiques au Brésil sont en ce moment à l'étude.

7. Références

CYBIS, W.A., MICHEL, G. BRANGIER, E.. L'ergonomie dans l'enjeu démocratique du vote électronique des citoyens : éléments d'utilisabilité électorale. In DUBOIS, M., BOBILLIER-CHAUMON M-E. ,RETOUR D.,(eds) Relations de services : Nouveaux usages, nouveaux usagers. Editions De Boeck, 2009 (à paraître).

MICHEL, G., KLEIN, M. Utilisabilité et discrimination : étude préliminaire des machines à voter françaises, IHM 2008, 235-238, Librairie digitale des ACMs, Metz, 2-5 décembre 2008.

ENGUEHARD C. Vote électronique et preuve papier. Actes du 14ème Colloque international "De l'insécurité numérique à la vulnérabilité de la société", Paris, 14 et 15 juin 2007.

OSCE/ODIHR. Republic of Estonia parliamentary elections 4 March 2007 OSCE/ODIHR Election Assessment Mission Report, 28 June 2007.

W3C. Web Accessibility Initiative, September 2005.

MICHEL, G. CYBIS, W.A. BRANGIER, E. Nomenclature de critères ergonomiques pour les urnes de votes électroniques. Revue d'Interaction Homme-machine/Journal of Human-Machine Interaction, Vol 8, n°3, 125-148, 2008.

Uma plataforma livre para padronização do desenvolvimento de sistemas no Governo Federal

Antonio Carlos Tiboni, Flávio Gomes da Silva Lisboa, Luciana Campos Mota

CETEC Curitiba - Serviço Federal de Processamento de Dados
CEP 80520-170 – Curitiba – PR – Brazil

{antonio.tiboni, flavio.lisboa, luciana.mota}@serpro.gov.br

Abstract. *This article presents the Demoiselle Framework project, a platform for Java development that aims to be the standard for all new development hires for the Brazilian federal government. It will ensure interoperability and ease the maintenance of systems in several ministries and local authorities through the adoption of open standards. Built as a free software, it will be maintained in community, benefiting the whole Brazilian society, which can take part in the technological directions of the government as any person on the planet, interested in helping and improving the tool.*

Resumo. *Este artigo apresenta o projeto do Demoiselle Framework, uma plataforma de desenvolvimento em Java que visa ser o padrão para todas as novas contratações de desenvolvimento de programas do governo federal. Ela garantirá interoperabilidade e facilidade de manutenção dos sistemas de diferentes ministérios e autarquias graças a adoção de padrões abertos. Construída como software livre, será mantida em comunidade, beneficiando a toda a sociedade brasileira, que poderá participar do direcionamento tecnológico do governo, bem como de qualquer indivíduo do planeta, interessado em contribuir e melhorar a ferramenta.*

1. Introdução

A ampliação do uso de software livre pelo governo federal surgiu de uma necessidade de segurança ampliada, ao utilizar softwares nacionais, da não dependência tecnológica, da redução de custos com pagamento de licenças e, em contrapartida, investimento na inteligência nacional. Daí verificou-se a necessidade da criação de um produto que permitisse o reuso de estruturas pré-existentes, que facilitasse a manutenção das aplicações e, principalmente, que permitisse a padronização dos códigos.

O framework Demoiselle para Java é uma ferramenta de código-aberto e totalmente livre, que visa garantir a interoperabilidade e facilidade de manutenção dos sistemas das diferentes instituições do governo federal. A padronização é o cerne do Demoiselle, pois visa facilitar a integração dos sistemas. A ideia é que, a partir de um framework e de uma arquitetura de referência, um conjunto de requisitos gere uma aplicação que possa ser mantida por qualquer um que conheça os dois primeiros.

Como framework integrador, o Demoiselle constitui-se de uma camada de alto nível denominada Framework Arquitetural, que estabelece interfaces padronizadas para serem usadas pelas aplicações. Nas camadas inferiores encontram-se os frameworks de base e de fundação, amplamente utilizados pelo mercado, e para a aplicação, as mudanças de infraestrutura são completamente transparentes. A ferramenta foi

construída sob as premissas de ser extensível, fácil de usar, estável, configurável, confiável e ter sua documentação publicada.

O Demoiselle pode oferecer uma maior produtividade, redução na curva do aprendizado, bem como a simplificação dos processos. O framework integrador será mantido em comunidade, totalmente aberto e compartilhado, o que permitirá que diferentes entidades e instituições contribuam e sejam beneficiadas pelo reuso das estruturas existentes.

O desenvolvimento orientado a componentes com ciclo de vida independente do framework arquitetural permite que as aplicações não fiquem dependentes dos módulos, que podem ser construídos colaborativamente.

2. Objetivos

Todos os sistemas especialmente desenvolvidos para órgãos federais deverão utilizar a plataforma Demoiselle. As linguagens e outros aspectos que balizaram a elaboração da plataforma são constituídas de padrões, o que facilita sua adoção, não só por parte dos responsáveis pelas áreas de TI nos órgãos federais, como também pelas próprias empresas fornecedoras de soluções para o Governo.

A implementação do Demoiselle pretende automatizar e acelerar a integração de sistemas, aumentar a produtividade e eliminar o retrabalho. Um dos benefícios esperados é a economia financeira, pois não há necessidade de arcar com custos de licenças de software. Outro benefício é o de qualquer pessoa poder baixar a plataforma e usar livremente, fato este que incentiva a utilização de software livre em todas as esferas da sociedade, quer sejam empresas de Governo, quer sejam empresas privadas.

3. Arquitetura

O modelo arquitetural do Demoiselle é apresentado na figura 1. Os componentes na verdade não fazem parte do *framework* arquitetural, pois possuem um ciclo de vida independente. Por este fato, eles não geram dependência obrigatória nas aplicações utilizadoras e podem ser construídos colaborativamente.

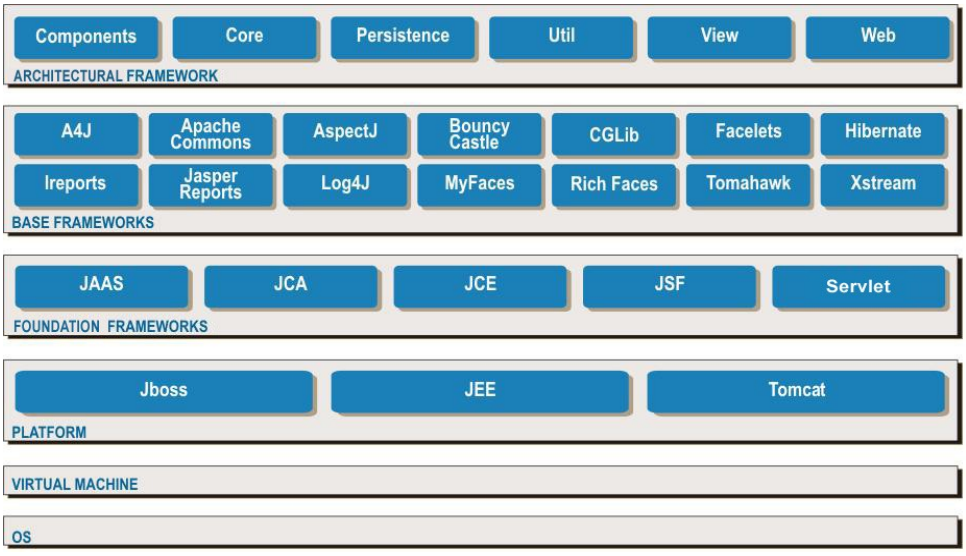


Figura 1. Arquitetura do Demoiselle Framework

A arquitetura de referência proposta para o Demoiselle é baseada em camadas.

Além das camadas clássicas do modelo MVC (Modelo, Visão e Controlador), elas se distinguem como camadas de persistência, transação, segurança, injeção de dependência e mensagem.

3.1. Módulos

- Core: conjunto de especificações que dão base estrutural ao framework possibilitando padronização, extensão e integração entre as camadas das aplicações nele baseadas.
- Persistence: realiza a integração do sistema com outros sistemas gerenciadores de dados garantindo eficiência para recuperar, armazenar e tratar informações.
- Util: contém componentes utilitários que facilitam o trabalho de outras funcionalidades do framework e seus módulos lógicos.
- View: contém implementações de componentes específicos de interface com usuário baseados na especificação JSF.
- Web: implementação do módulo lógico Core para aplicações Web (J2EE), provê utilitários comuns de aplicações web que facilitam tratamento de sessões de usuário e suas requisições.

3.2. Contextos

Enquanto os módulos dentro da arquitetura podem ser vistos como camadas verticais, existem camadas transversais, que são os contextos. Os contextos referem-se a características cuja operação transcende as camadas verticais, como transações, troca de mensagens, segurança da aplicação e aspectos que devem ser injetados no código. A figura 2 ilustra o relacionamento do contexto com a implementação de MVC do Demoiselle.

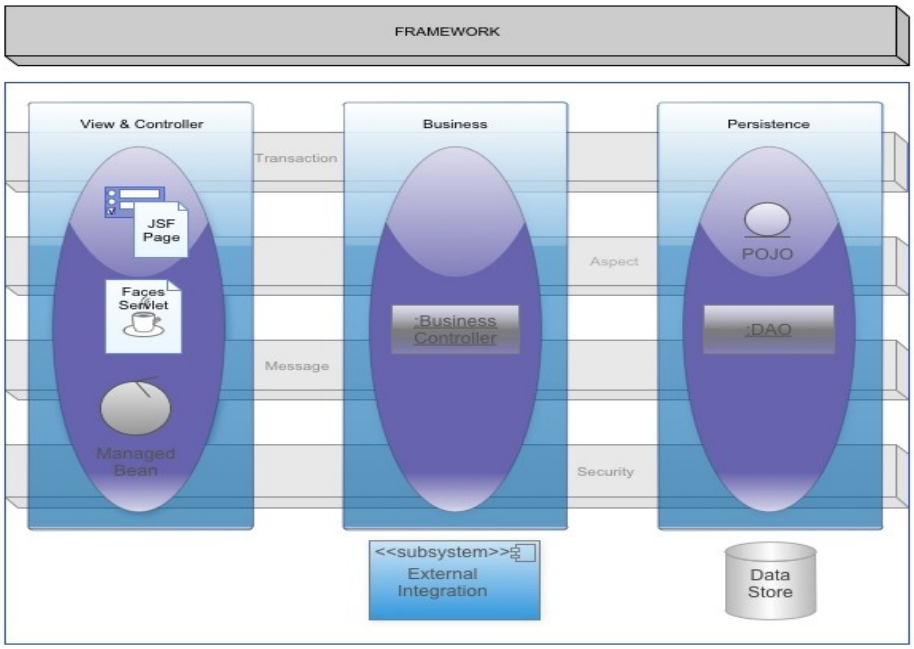


Figura 2. Camadas Verticais e Transversais do Demoiselle

3.3. Integração entre Camadas

Entre os aspectos a destacar sobre a arquitetura, o mais significativo é a utilização de padrões de projeto tais como Factory, Proxy, IoC e injeção de dependências para manter a integração de camadas em um nível de acoplamento baixo afim de garantir uma melhor manutenção e escrita/legibilidade das classes representantes de cada camada.

O mecanismo de integração entre camadas atua na camada de visão injetando objetos de negócio através de uma fábrica do próprio framework ou alguma fábrica definida pela aplicação. Esta fábrica poderá utilizar um proxy, do framework ou da aplicação, para a instanciação do objeto de negócio.

O mecanismo de integração entre camadas atua também na camada de regras de negócio injetando objetos de persistência através de uma fábrica do próprio framework. Esta fábrica poderá utilizar um proxy, do framework ou da aplicação, para a instanciação do objeto de persistência.

3.4. Desenvolvimento Baseado em Componentes

Os módulos citados no item 3.1 constituem o núcleo do framework. A ele podem ser acoplados componentes desenvolvidos por terceiros, desde que usem a interface definida pelo Demoiselle.

Essa arquitetura permite resolver um paradoxo do desenvolvimento de software. Padrões são necessários, para permitir a manutenção e promover o reuso. Padrões são coisas que não devem mudar (com facilidade). Por outro lado, a necessidade exige inovação, e a inovação geralmente provoca a quebra de padrões.

Dessa forma, o Demoiselle Framework faz uso de interfaces de classe em seu núcleo, as quais ditam o padrão de codificação para as aplicações. Por outro lado, a possibilidade de criar componentes e conectá-los ao Demoiselle dá ao desenvolvedor uma grande liberdade, pois ele pode criar qualquer funcionalidade para uma aplicação que use o framework, inclusive uma versão alternativa de uma funcionalidade já existente.

Os componentes tem ciclo de vida independente, não geram dependência obrigatória nas aplicações geradas e podem ser construídos colaborativamente. A facilidade de acoplar e desacoplar componentes, aliada com a padronização permite superar um problema do software desenvolvido em comunidade: o prazo de entrega. A componentização permite que um usuário crie e utilize uma nova funcionalidade para o framework sem ter de esperar que a mesma seja aprovada e incorporada pela comunidade. Mas uma vez que isso ocorra, ele pode trocar sua versão pessoal por aquela testada e melhorada pela comunidade.

4. Inovação e Ineditismo

Demoiselle é disponibilizado sob a licença LGPL 3 e qualquer componente utilizado ou desenvolvido para ele dever ser compatível com essa licença.

4.1. Comunidade

Apesar de ter sido feito para atender ao Governo Federal, o Demoiselle é um software livre, e o seu desenvolvimento é feito de forma colaborativa. Sendo assim, qualquer pessoa pode participar de seu desenvolvimento, bem como na evolução do

framework, quer seja reportando erros, efetuando melhorias de código ou ajudando na confecção da documentação.

4.2. Oportunidades de Negócio

À primeira vista parece ser mais um framework Java, o que pode levantar questões sobre o motivo de ter sido criado, em vez de se adotar um já existente como padrão. A intenção não é criar um produto para competir com outros frameworks, mas estabelecer uma plataforma que implemente o conceito de framework integrador. Ele realiza a integração entre vários frameworks especialistas e garante a evolução, manutenibilidade e compatibilidade entre cada um deles. Sua maior contribuição é dar um direcionamento ao uso das tecnologias.

O direcionamento tecnológico é importante para os prestadores de serviço, pois permite que eles se especializem nas tecnologias que foram definidas como o padrão a ser utilizado. Com a adoção de uma arquitetura de referência e uma plataforma integradora de tecnologias, um órgão do Governo Federal pode contratar uma empresa para desenvolver um sistema e depois ter a segurança de contratar outra para dar manutenção. Ou ainda, se tiver uma área de desenvolvimento de sistemas, pode ele mesmo realizar manutenção. Isso permite contratações mais flexíveis, assim como licitações de ampla concorrência.

Dessa forma, pequenas e médias empresas, ou até desenvolvedores ou consultores independentes que trabalhem como pessoa jurídica, podem participar de concorrências públicas junto com grandes empresas, pois todos terão acesso ao Demoiselle e às tecnologias relacionadas sem qualquer custo. O Governo Federal ganha, pois a ampla concorrência tende a diminuir os custos de manutenção dos sistemas de informação, além de que a própria plataforma é gratuita, o que o desonera em relação a licenças. A iniciativa privada ganha como um todo, pois a padronização dos sistemas democratiza a concorrência, evitando que um grupo de privilegiados tenha o mercado Governo reservado para si. Como diria o matemático John Nash, "*o melhor resultado virá quando todos do grupo fizerem o melhor para si mesmos e também para o grupo como um todo."* (Nasar, 2002)

O Demoiselle extrapola as relações comerciais entre Governo e iniciativa privada. Qualquer empresa pode utilizar a plataforma para desenvolver sistemas para qualquer cliente. E esse cliente terá a segurança de que não dependerá do criador da aplicação para mantê-lo, ele terá liberdade de tomar conta de seus sistemas e de implementar funcionalidades sem ter que esperar por próximas versões. Tudo o que alguém precisa para fazer manutenção em uma aplicação desenvolvida pelo Demoiselle será conhecer os requisitos, o framework e a arquitetura de referência.

A qualidade é claramente afetada pelo estabelecimento de uma comunidade de usuários da plataforma, pois efetiva o reuso dos componentes criados. O Demoiselle oferece uma oportunidade de agregar a experiência de diversas comunidades de software relativas a cada um dos frameworks especialistas e criar uma grande rede colaborativa de desenvolvimento.

5. Ambiente de Desenvolvimento

O ambiente compatível para o desenvolvimento baseado no Demoiselle é composto de JVM, IDE e servidor de aplicação. A JDK requerida é a 1.5.x. Como IDE, recomendamos o Eclipse 3.3.2 JEE Developer (Europa) com o plugin AJDT

1.5.2.200804241330. Neste artigo o Eclipse é utilizado. Os servidores JEE com contêiner web 2.5 são compatíveis com o Demoiselle. Isso inclui Tomcat 6.x e JBoss 4.2.x.

6. O Nome

A nova plataforma é um sistema brasileiro, desenvolvido pelo SERPRO, empresa pública brasileira para uso nacional. A princípio deve causar estranheza que seja nomeado por uma palavra francesa. A explicação é a seguinte: Demoiselle é o nome de um modelo leve e pequeno de avião que o brasileiro Santos Dumont, conhecido como o Pai da Aviação, idealizou e pilotou em 1907, um ano depois do lendário vôo do 14 Bis. Diversas versões do Demoiselle foram testadas pelo aviador de 1907 a 1909. Ao final das anotações e da documentação da construção da aeronave, Santos Dumont permitia a utilização, adaptação e cópia de seu trabalho. Em função dessa perfeita consonância com os conceitos atuais do software livre, Demoiselle foi o nome escolhido para batizar a plataforma de desenvolvimento de software.

Em dado momento de sua vida, Santos Dumont "dizia a todos que estava sem dinheiro. Ninguém acreditava em suas palavras, mas para animá-lo aconselharam-no a patentear o *Demoiselle.* Ele recusou. Era seu presente para a humanidade, disse que preferia terminar seus dias em um asilo de pobres que cobrar aos outros o privilégio de copiar sua invenção e de fazer experimentos aéreos." (Hoffman, 2004)

7. Conclusão

Reusabilidade é palavra de ordem na utilização do Demoiselle Framework. Este artigo apresentou os objetivos, a proposta, a característica de ser software livre, a possibilidade do desenvolvimento colaborativo e, finalmente, o exemplo de criação de uma pequena aplicação web. Obviamente, uma aplicação simples, porém completa no tocante à utilização das camadas propostas pelo framework. Ela é suficiente para dar uma dimensão do que o Demoiselle é capaz, bem como instigar os desenvolvedores a completá-la, levando-os, é claro, a estudar sua documentação.

O Demoiselle apenas está em consonância com o ideal de um mundo de padrões. Padrões são bons, quando bem utilizados pois permitem que as pessoas se comuniquem. E, ainda que componentes sejam fabricados por pessoas diferentes e, mesmo assim, possam ser integrados a um único produto final. Assim, o desenvolvimento de um software que utiliza a mesma plataforma e processo tende a provê-lo de melhorias pela experiência acumulada com o tempo.

Referências

ALUR, Deepak. Core j2eetm patterns: as melhores praticas e estrategias de design. Rio de Janeiro: Campus, 2002. 406 p.

BASS, Len. Software Architecture in Practice. 2 ed. Massachusets: Addison Wesley, 2003. 528 p. ISBN 032115499.

COAD, Peter; YOURDON, Edward. Análise baseada em objetos. 2.ed. Rio de Janeiro: Campus, 1992. 225 p. (Série Yourdon Press) ISBN 8570017006.

DEITEL, H. M. Java: como programar. 6. ed. São Paulo, SP: Pearson Prentice Hall,

2006. 1110 p. ISBN 8576050196.

FAYAD, Mohamed; Schmidt, Douglas. Communications of the ACM, New York, v. 40, n. 10, p. 32-38, Oct. 1997.

FOWLER, M. Patterns of enterprise applications architecture. Prentice Hall, 2002.

GAMMA, Erich; HELM, Richard; JOHNSON, Ralph; VLISSIDES, John. Padrões de projeto: soluções reutilizáveis de software orientado a objetos. Porto Alegre: Bookman, 2000. 364 p. ISBN 8573076100

HOFFMAN, Paul. Asas da Loucura: a Extraordinária Vida de Santos-Dumont. Objetiva, 2004.

JOHNSON, R. E. How to develop frameworks. Disponível em *http://wiki.lassy.uni.lu/se2c-bib_download.php?id=453.* Acesso em 20/03/2009.

LISBOA, F. G. S. Framework Java Serpro. TEMATEC, Ano X, nº XX, 2008.

MENDES, Antonio. Arquitetura de Software: Desenvolvimento orientado para arquitetura. Rio de Janeiro: Campus, 2002. 212 p.

NASAR, Sylvia. Uma Mente Brilhante. Record, 2002.

PRESSMAN, Roger S. Engenharia de Software. 6 ed. São Paulo: Mc Graw Hill, 2006. 720 p. ISBN 8586804576.

SHORE, James; WARDEN, Shane. A Arte do Desenvolvimento Àgil. Rio de Janeiro: Alta Books, 2008. 420 p. ISBN 9788576082033

SOMMERVILLE, Ian. Engenharia de software. 6 ed. Rio de Janeiro: Addison Wesley, 2003. 592p. ISBN 8588639076

VON STAA, Arndt. Engenharia de programas. Rio de Janeiro: , 1987. 296 p.

IV

Computing and Biology, Applications to Health

COLIBRI 2009 • 22-23 de julho de 2009 • ISBN 978-85-7669-245-4

Computational Intelligence for Predicting the Chemotherapy Outcomes in Breast Cancer

René Natowicz[1], Antônio P. Braga[2], Roberto Incitti[3], Marcelo A. Costa[2], Patrick Siarry[4], Arben Çela[1], Euler G. Horta[2], Carmén DM. Pataro[2], Thiago Souza[2], Roman Rouzier[5]

[1] University of Paris-Est, Esiee-Paris - {rene.natowicz, arben.cela}@univ-paris-est.fr

[2]Federal University of Minas Gerais, Belo Horizonte
{apbraga,azevedo,ehorta,cdmp}@ufmg.br, thiago@dcc.ufla.br

[3]INSERM, Mondor Institute for Molecular Medicine, Créteil - roberto.incitti@inserm.fr

[4]University of Paris-Est, LiSSi, Créteil - siarry@univ-paris12.fr

[5]Pierre & Marie Curie University, Tenon Hospital, Paris - roman.rouzier@tnn.aphp.fr

Abstract. *Predicting the outcome of chemotherapy treatments in oncology is an important clinical issue for better allocating the patients to the treatments, and in pharmacology because an accurate characterization of the non responders could be of great help for designing new treatments dedicated to these cases. We present an ongoing franco-brazilian research for the design of efficient multi-genomic predictors in breast cancer. This research is supported by a four years* CAPES-COFECUB *program (2008-2011)*[1].
Keywords: *bioinformatics, computational intelligence, genomic predictors, chemotherapy treatments, breast cancer.*

1. Statement of the problem

Since the advent of high throughput genomic technologies a decade ago, massive information at the genomic level is available, which can be exploited in medical studies. Microarrays allow to measure simultaneously the expression levels of a large fraction of all known genes (ranging from 5 000 to 30 000). Relying on these data, it has become possible to design efficient predictors that significantly outperform the previous clinico-biologic ones. Our research is concerned with pre-operative chemotherapy treatments for breast cancer.

In the process of designing predictors of the outcomes of chemotherapy treatments, the main issues are that of identifying the genes actually involved in the response to the chemotherapy treatment; combining the expression levels of these genes for predicting the response to the treatment; and assessing the robustness of the predictors, i.e. the statistical independence of the method – gene selection and computational model – relatively to the learning set of patient cases. Our dataset comes from clinical trials which were jointly conducted at the Gustave Roussy Institute, Villejuif (France), and at the Anderson

[1]Project #591/08, Multigenic Predictors of the Responses to Preoperative Chemotherapy Treatment in Breast Cancer, shared by : University of Paris-Est, France; University Federal of Minas Gerais in Belo Horizonte, Brazil, School of Engineering, Computational Intelligence Laboratory; University Pierre and Marie Curie, Paris, France, Hospital Tenon, Obstetrics and Gynecology department; INSERM, Créteil, France, Mondor Institute for Molecular Medicine, Genomic Platform.

Cancer Center, Houston-Texas (USA) in 2004 and 2005 [Hess et al. 2006]. For each patient case, the data are the outcome of the pre-operative chemotherapy treatment, either responder or non responder, and the expression level of more than 22 000 genes, measured on the tumor tissues. The gene expression profiling was performed using oligonucleotide microarrays (Affymetrix U133A).

In its essence, designing a predictor is a supervised learning process aiming at recognizing the class of patient cases who are responders to the treatment and that of the non responders. The novelty comes from the nature of the high throughput genomic data and clinical trials. The number n of cases of the clinical trial is less than 200 while the number of variables, p, which are the expression levels measured, is more than 22 000. Otherwise stated, the problem is a two classes supervised learning where p greatly exceeds n [Simon 2003]. Furthermore, the data are very noisy because of the present state of the technology of DNA microarrays and because we are dealing with biological data. Hence, selecting a very small subset of relevant genes is crucial for designing robust predictors.

2. Selecting the genes

The performances of a predictor are measured by its accuracy, sensitivity and specificity. The accuracy, Ac, is the proportion of well predicted cases, regardless of their classes. The sensitivity, Se, is the proportion of well predicted responder cases, and the specificity of the predictor, Sp, is that of the well predicted non responder cases. Because only a very small number of genes can enter the predictor, each of them should give an information about both classes. Such genes are *bi-informative* [Natowicz R. et al. 2008b]. It has appeared to us that this main characteristics of the gene selection had been neglected in the previous works. A lot of methods have been proposed so far for selecting the genes, but the most widely used method still consists in selecting them according to a very basic statistical criterion, namely the the p-value to a t-test. According to this approach, one considers the mean value of the expression levels of each gene[2] measured on each class, then one selects the genes whose mean difference is the most likely not to have been obtained *by chance.* According to this criterion, the expression levels of the most relevant genes are typically as depicted in figure 1, which is the box-plot of the expression levels of the gene MAPT (*microtubule-associated protein tau*, probe 203929_s_at). The gene MAPT is the most relevant gene according to the p-value of a t-test [Rouzier R. et al. 2005]. From this figure one can see that the information given by the gene MAPT is essentially about the non-responder cases. Such genes are *mono-informative*, and almost all the genes selected this way are *mono-informative.*

[2]The expression levels are those of the microarray's DNA probes. A lot of genes are represented by several DNA probes on the Affymetrix microarrays. Hence we should talk of the level of expression of the DNA probes rather than that of a gene. In the following, for making short, we will nevertheless talk of the expression level of a gene.

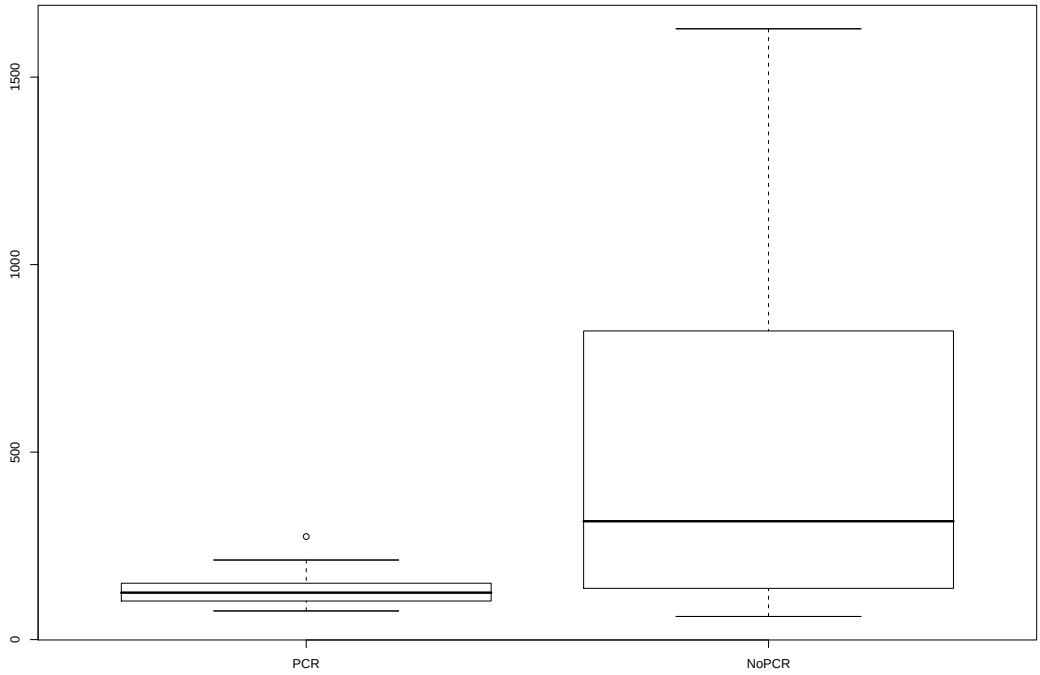


Figure 1. Boxplots of the expression levels of the gene MAPT (probe 203929_s_at). MAPT is a mono-informative gene. Left : expression levels of the responder cases (PCR : pathologic complete response); right : those of the non responder cases (NoPCR).

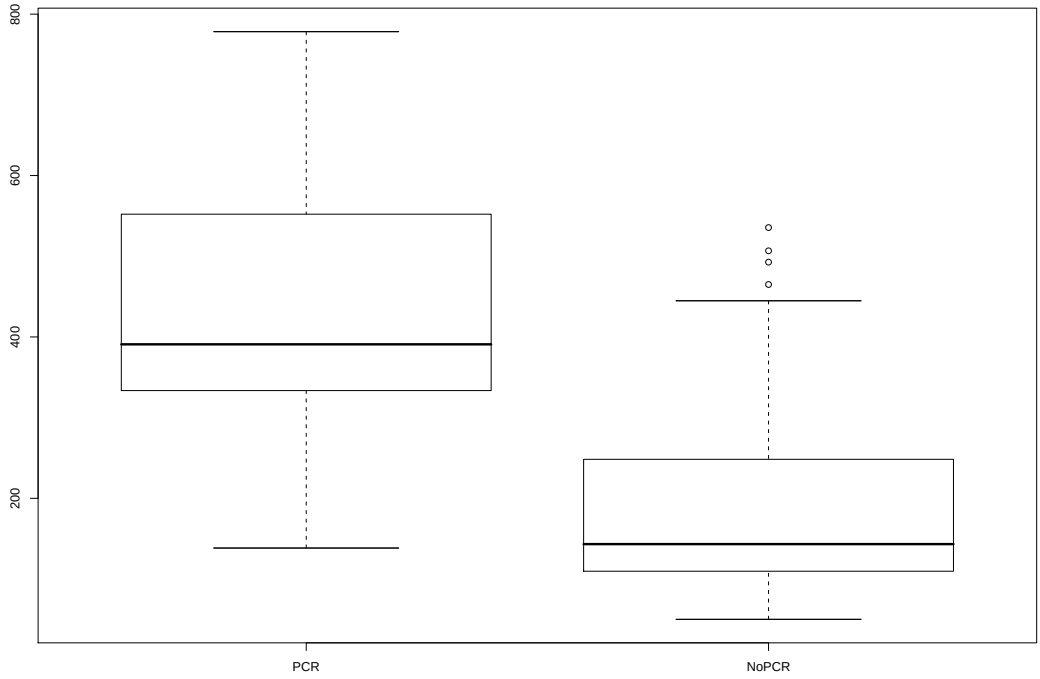


Figure 2. Boxplots of the expression levels of the gene BTG3 (probe 205548_s_at). BTG2 is a bi-informative gene. Left : expression levels of the responder cases; right : those of the non responder cases.

Because we wanted the predictors to be both highly sensitive and highly specific, we have developed a new method for characterizing and selecting *bi-informative genes*. In figure 2, the box-plot is that of the most relevant gene according to the criterion that we have proposed in [Natowicz R. et al. 2008b]. All the genes selected this way, were *bi-informative.* The method that we have developped can be seen as a selection process of the genes according to their individual sensitivities and specificities. We have considered each gene as an elementary predictor or the response to the chemotherapy, the sensitivity and specificity values of which were the proportions of responder and non responder cases of the learning set of cases that were correctly predicted by the gene. This way, each gene could be plotted in the two dimensional space of sensitivity and specificity values. In this space, we have also plotted a hypothetic *ideal* gene, supposed to predict all the cases of the learning set (the sensitivity and specificity values of which were both equal to one). Then, we have selected the genes according to their euclidean distance to the *ideal* one, in the sensitivity-specificity space. In figure 3, all the genes of the DNA microarrays used for the clinical trial have been plotted (more than 20 000 genes), together with the *ideal* gene (at coordinates $(1, 1)$, at the upper right corner of the figure). In figure 4 are the 30 genes which were the closest to the *ideal* gene.

3. Predicting the outcomes of the treatment

A lot of mathematical and computational models have been used so far for combining the expression levels into a prediction function [Natowicz R. et al. 2008a, Rouzier R. et al. 2009]. These models range from very simple linear regression, to highly non linear models, among which various models of neural networks. Whatever the model, the question of the robustness of the predictor is of first importance in the perspective of using it in clinical routine. Thanks to the gene selection step, we could rely on a small number of genes, the order of which was less than thirty in our first publications [Natowicz R. et al. 2008b], and presently less than twenty [Natowicz R. et al. 2008c]. In order to avoid the overfitting of the data, the expression levels of the selected genes were given as input to a multi-objective neural networks, a classifier model developed by Braga & al. [Braga AP. et al. 2006]. The learning process of this classifier consists in minimizing both the classification error and the overfitting of the data. In this approach, the search of for a *global optimum* is substituted by that of *Pareto-optimality*. After optimization, the Pareto-set contains the non-dominated solutions that cannot be improved in one of the objectives without degrading the others. The decision making procedures follow the Pareto-set generation: a solution is selected according to a pre-established criterion. The simplest selection approach is to minimize the error of a validation set. Other selection strategies that explore the Pareto set properties have been applied successfully [Kokshenev and Braga AP. 2008].

4. Results and developments to come

The predictors that we have designed have significantly outperformed the best predictors reported for the same problem and data [Hess et al. 2006]. The performances of our predictors, measured on an independent set of data (data neither used for selecting the genes nor designing the multi-objective classifier) are no less than: Ac=0.86, Se=0.92, Sp=0.84. The statistical validation was done by cross-validations and the stability of the method of gene selection was demonstrated on replicates.

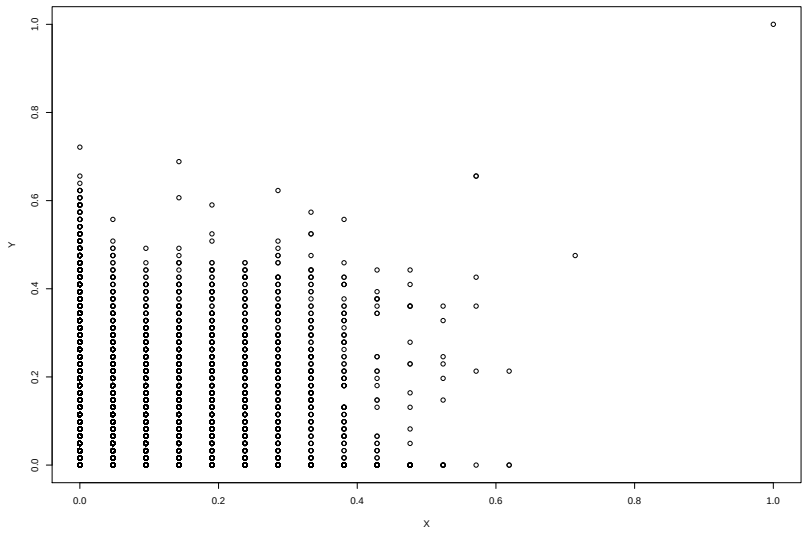


Figure 3. The genes of the DNA microarrays, plotted in the sensitivity-specificity space. X, Y axes: sensitivity and specificity values of the genes. The hypothetic *ideal* gene is at coordinates (1,1).

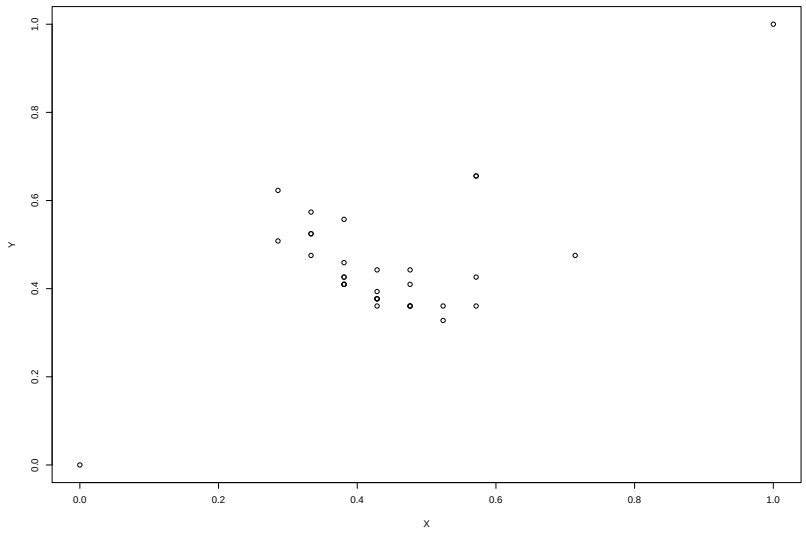


Figure 4. Coordinates of the 30 genes the closest to an *ideal* gene.

But, because our objective is to design efficient predictors to be widely used in clinical routine, we must increase further their sensitivity: the false negatives of a predictor being patient cases wrongly predicted not to benefit from the treatment, these situations should be avoided as much as possible, although the decision of not allocating a patient to the treatment will always belong to the clinicians.

At the present stage of our research, it has appeared to us that we must gain a deeper understanding of the statistical properties of the populations under study in the clinical trials (stratifying the data according to age and genetic criteria is a possible need) and of the statistics of the gene expressions themselves. This is a work in progress. Furthermore,

the biological mechanisms underlying the responses to the chemotherapy treatments are those of gene interaction networks which, up to now, are scarcely known. The gene expressions that one measures are their resulting effects, but these interactions are not taken into account in the gene selection process itself. Hence, we would like to investigate the selection of genes subsets. Because the number of variables is around 20 000, one should rely on (and possibly develop) methods for searching efficiently the huge space of the gene subsets [Siarry P. and Michalewicz 2007]. It is a major development to come for this collaborative research, that motivates our wish to open the project to research teams in metaheuristics for optimization.

References

Braga AP., Takahashi, R., Costa MA., and Teixeira, R. (2006). *Multi-Objective Algorithms for Neural Networks Learning*. Studies in Computational Intelligence. Springer.

Natowicz R., Braga AP., Incitti R., Horta EG., Rouzier R., Rodrigues, T., and Costa MA. (2008a). A new method of dna probes selection and its use with multi-objective neural network for predicting the outcome of breast cancer preoperative chemotherapy. In *ESANN*, pages 71–76.

Natowicz R., Incitti R., Horta EG., Charles, B., Guinot, P., Yan, K., Coutant, C., Andre, F., Pusztai, L., and Rouzier R. (2008b). Prediction of the outcome of preoperative chemotherapy in breast cancer by dna probes that convey information on both complete and non complete responses. *BMC Bioinformatics*, 9:149+.

Natowicz R., Incitti R., Rouzier R., Çela A., Braga AB., Horta EG., Rodrigues, T., and Costa M. (2008c). *Downsizing Multigenic Predictors of the Response to Preoperative Chemotherapy in Breast Cancer*. Lecture Notes in Computer Science. Springer.

Rouzier R., Coutant, C., Lesieur, B., Mazouni, C., Incitti R., Natowicz R., and Pusztai, L. (2009). Direct comparison of logistic regression and recursive partitioning to predict chemotherapy response of breast cancer based on clinical pathological variables. *Breast Cancer Res. Treat.*

Rouzier R., Rajan, R., Wagner, P., Hess, K., Gold, D., Stec, J., Ayers, M., Ross, J., Zhang, P., Buchholz, T., Kuerer, H., Green, M., Arun, B., Hortobagyi, G., Symmans, W., , and Pusztai, L. (2005). Microtubule-associated protein tau: A marker of paclitaxel sensitivity in breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 102(23):8315–8320.

Siarry P. and Michalewicz, Z. (2007). *Advances in Metaheuristics for Hard Optimization*. Natural Computing Series. Springer Berlin Heidelberg.

Hess, K., Anderson, K., Symmans, W., Valero, V., Ibrahim, N., Mejia, J., Booser, D., Theriault, R., Buzdar, A., Dempsey, P., Rouzier R., Sneige, N., Ross, J., Vidaurre, T., Gomez, H., Hortobagyi, G., and Pusztai, L. (2006). Pharmacogenomic Predictor of Sensitivity to Preoperative Chemotherapy With Paclitaxel and Fluorouracil, Doxorubicin, and Cyclophosphamide in Breast Cancer. *J Clin Oncol*, 24(26):4236–4244.

Kokshenev, I. and Braga AP. (2008). A multi-objective approach to rbf network learning. *Neurocomputing*, 71(7-9):1203–1209.

Simon, R. (2003). Supervised analysis when the number of candidate features (p) greatly exceeds the number of cases (n). *SIGKDD Explor. Newsl.*, 5(2):31–36.

Large scale protein function prediction tools

Raquel C. de Melo Minardi[1], François Artiguenave[1], Goran Neshich[2]

[1]Genoscope / Institut de Génomique / CEA
2 rue Gaston Crémieux CP5706 91057 Evry cedex France
[2]Structural Computational Biology Group / CNPTIA / EMBRAPA
Av. André Torsello, 209 – Unicamp, Barão Geraldo – Campinas, SP – Brazil
{raquelcm,artigue}@genoscope.cns.fr, neshich@embrapa.cbi.cnptia.br

Abstract. *In this project we intend to reinforce a recent collaboration established between two French and Brazilian bioinformatics laboratories for the development of a structural genomics approach to functional annotation of proteins. From initial works, we explored a new methodology for protein annotation based on modeled structural data. During this work, we identified that new tools will have to be designed for large scale protein function prediction. In this proposal, we would like to analyze the results of the proposed methodology with proteins of known function as well as analyze the impact of the use of different known techniques for each of its steps (homology modeling, cavity detection, binding / active site prediction and clustering according to sub-family specificity). We will possibly develop new algorithms in order to gain quality in prediction.*

Resumo. *Neste projeto, pretendemos reforçar uma recente colaboração entre dois laboratórios franceses e um brasileiro para o desenvolvimento de uma abordagem de genômica estrutural para anotação funcional de proteínas. Em trabalhos iniciais, exploramos uma nova metodologia para anotação de proteínas baseada em modelos estruturais. Durante este trabalho, identificamos que novas ferramentas são necessárias para anotação de proteínas em larga escala. Nesta proposta, pretendemos analisar os resultados desta metodologia com proteínas de função conhecidas bem como o impacto do uso de diferentes técnicas conhecidas utilizadas em cada uma de suas etapas (modelagem por homologia, detecção de cavidades, sítio de ligação / ativo e agrupamento de acordo com a especificidade de subfamílias). Possivelmente desenvolveremos novos algoritmos com o intuito de melhorar a qualidade das predições.*

1. Objective

The aim of this collaborative project is to develop a methodology and tools for large scale protein function prediction based on structural data. The major goals are:
1. Describe proteins and their surfaces in terms of features that discriminate between target areas which are the molecular interfaces through which proteins perform their function and those that are not.
2. Organize and maintain existing data relevant to broad spectrum of academic and commercial users and generate new data for purposes of identification of protein

function and specific loci responsible for functionality: catalytic site residues and interface forming residues.
3. Offer the newly acquired knowledge about protein substrates interactions to interested partners both from academic and industrial/commercial sector.
4. Fine tune applications for testing in some of the urgent problems in food industry, agro business and pharma, such as new antibiotics, new pesticides, new drugs and new strains.
Our main objective is to raise the level of our understanding in how two molecules (a protein/enzyme from one side and an inhibitor from the other) do communicate in order to engage in close contact which results in some kind of functional modification of (in this case) protein molecule, an important element of metabolic pathways which are responsible for specific disease or economically interesting traits.

2. Motivation

Both the Brazilian and French agriculture, livestock, food and pharmaceutical industries are challenged not only by the upcoming industries in Asia and established ones in North America and Australia, but also internally regarding demands for higher productivity and healthier product output with concomitant environmental protection. It is clear that this sector may add to its competitiveness by new product discovery, by process improvements, by betterment of used strains, and by the introduction of new biotechnological processes especially those referred to us as the "green routes" for the fermentative production of rare chemicals (harvesting chemicals instead of seeds/fruits). In order to promote a progress in this area, and targeting a list of some of the most important challenges facing research and development in both drug and food industries, we may cite problems which need to be addressed, such as: a) how to identify in silico function of proteins annotated in newly sequenced genomes, b) how to recognize and identify a network of interacting proteins within metabolic pathways, c) how to predict interactive part of the protein surface through which protein will communicate with other biomolecules and also perform its function and d) how to predict the effects of drugs and food (for both humans and animal organisms) by specifically keeping costs at lowest possible levels and increasing R&D process output.
This particular collaboration aims to contribute to those strategic goals by jointly reinforcing the competitiveness of the French and Brazilian pharma and agro (food) industries by enhancing research effort through joint work where complementary expertise of French and Brazilian partners will synergistically act in order to offer some improvement for the processes such as the analysis of protein sequences and structures as well as how specific amino acids mediate the interactions of these proteins with their substrates. Noteworthy is to mention that these tools may be used by scientists to collaboratively compile their own data on, for example, the specific anomalies associated with any given genomic locus and to further analyze their new data on the basis of already existing knowledge.

3. Partners

3.1. Genoscope contribution

The Genoscope (the French National Sequencing Center), has participated in the Human Genome Project and in international consortium in the domain of plant genomes and has revealed major events in the evolution of eukaryote genomes (vertebrates, ciliates, plants). The Genoscope has recently joined the CEA (Comissariat à l'Energie Atomique) and as so, has integrated in its main priorities, energy and environment issues. Consequently, the Genoscope is now enlarging the field of analysis of sequence data in extending the analysis to the experimental identification of biological functions. Internal projects are mainly in the domains of the environment and biodiversity and will open up perspectives for the development in the biotechnology industry and sustainable development. We started applying structural bioinformatics techniques in enzyme function predictions and are developing a methodology which is based on homology modeling [Tramontano et al. 2001], cavity analysis [Dundas et al. 2006], Hidden Markov Models (HHMs), conceptual clustering [Fisher 1987] and molecular docking of metabolites. We already have some interesting results that are being experimentally tested and show the applicability of such type of methods. For specific protein families, we intend to test the bioinformatics predictions using the high trough put screening platform setup at Genoscope, which will give experimental proofs of bioinformatics prediction and will allow to annotate new protein families.

3.2. Embrapa contribution

Embrapa will offer STING platform [Neshich et al., 2006], Interface, STING_DB and STING_RDB and also work in the development of new algorithms and tools to predict binding and catalytic sites. STING_DB as a source of structure/function descriptors which are to be used as vectors in 1D representation of 3D proteins. The Sting database operates with a collection of both publicly available data (e.g., PDB [Berman et al., 2000], HSSP [Schneider and Sander, 1996; Schneider et al., 1997], Prosite [Hulo et al., 2006], and UniProt [Apweiler et al., 2004]) and its proprietary protein sequence and structure (PSS) descriptors, such as geometric parameters (e.g., cavity, curvature), physic-chemical parameters (e.g., electrostatic potential), and conservation related parameters (e.g., SH2Qs, evolutionary pressure). The data consolidated and integrated into STING_RDB (STING relational database) make this database one of the most comprehensive databases available for analysis of protein sequence, structure and function and is updated on a weekly basis.

7. Planning

The project has well defined tasks, which ultimately lead to the following milestones:
1. Within one year, the clear guidelines for construction of an algorithm and building of user interface for automatic determination of catalytic site and interface forming amino acids, based on defined value intervals for sequence, structure, function and stability

descriptors, which by themselves have already been recorded within STING_DB or are still to be added as the new ones), would have been established.
2. In second year, the web based program for identification of catalytic site and interface forming AA from the sequence and structural data will be established.
3. In third year, the automatic targeting system for determination of function and function modification from the genomic data, based on both sequence and structure (homology models could be used if experimental structure information available is not sufficient [Tramontano et al., 2001]) will be established.
4. At the end of fourth year, the fully developed automatic system for identification of protein catalytic site, structural alignment and interface area prediction will be available to users, based on data analysis and data mining of the STING platform, which offers the largest collection of physicochemical parameters describing proteins structure, stability, function and interaction with other macromolecules.

References

[Apweiler et al. 2004] Apweiler, R., Bairoch, A., Wu, C., Barker, W., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M., Natale, D., O'Donovan, C., Redaschi, N., Yeh, L. (2004). UniProt: the universal protein knowledgebase. Nucleid Acids Res., 32:(Database issue):D115D119.
[Berman et al.] Berman, H., Bhat, T., Bourne, P., Feng, Z., Gilliland, G., Weissig, H., Westbrook, J. (2000). The Protein Data Bank and the challenge of structural genomics. Nat. Strcut. Biol., Suppl.9579.
[Dundas et al. 2006] Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., and Liang, J. (2006). CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acid Res., 34:W116W118.
[Fisher 1987] Fisher, D. (1987). Knowledge acquisition via incremental conceptual clustering. Machine Learning, 2:139–172.
[Hulo et al. 2006] Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., LangendijkGenevaux, P., Pagni, M., Sigrist, C. (2006). The PROSITE database. Nucleic Acids Res., 34(Database issue):D227D230.
[Neshich et al. 2006] Neshich, G., Mazoni, I., Oliveira, S., Yamagishi, M., KuserFalcão, P., Borro, L., Morita, D., Souza, K., Almeida, G., Rodrigues, D., Jardine, J., Togawa, R., Mancini, A., Higa, R., Cruz, S., Vieira, F., Santos, E., Melo, R., and Santoro, M. (2006). The star sting server: a multiplatform environment for protein structure analysis. Genet. Mol. Res., 5(4):717–726.
[Schneider and Sander 1996] Schenider, R., and Sander, C. (1996). The HSSP database of protein structuresequence alignments. Nucleic Acids Res., 24(1):201205.
[Schneider et al. 1997] Schneider, R., Daruvar, A., and Sander, C. (1997). The HSSP database of protein structure sequence alignments.
[Tramontano et al. 2001] Tramontano, A., Leplae, R., and Morea, V. (2001). Analysis and assessment of comparative modeling predictions in casp4. Proteins, Suppl. 5:22–38.

Developing virtual environments with intelligent interface agents to support virtual treatments

Rosa Maria E. M. da Costa[1], Luis Alfredo V. de Carvalho[2], Lídia S. Cardoso[3], Elie Cheniaux[4], Vera Maria B. Werneck[1]

Universidade do Estado do Rio de Janeiro – UERJ,
[1]IME – Dept de Informática e Ciência da Computação

[4]Faculdade de Ciências Médicas

rcosta@ime.uerj.br, echeniaux@gmail.com, vera@ime.uerj.br

Universidade Federal do Rio de Janeiro – UFRJ,
[2]COPPE - Programa de Eng. de Sistemas e Computação
[3]Laboratório de Neuropsicologia e Cognição – FM/HUCFF
luisalfredo@.ufrj.br, lidiacardoso@hucff.ufrj.br

Abstract. *In general, people who have suffered brain damage, have to perform physical therapies in order to recover lost physical movements. These patients rarely follow treatments associated with cognitive stimulation practices. This paper proposes the integration of inter-institutional collaborative groups to construct and use 3D virtual environments to support cognitive rehabilitation of people with different types of brain damage. These environments may contain intelligent interface agents that help the patient perform the proposed tasks. As there are certain groups in the context of the INRIA working in this field, the integration of researchers from Brazil and France could open new perspectives for the development of new technological applications and new ways of using these environments.*

1. Introduction

The technology advances are promoting changes in the activities associated with surgical and therapeutic training or professional updating. Such applications require new approaches towards study, design, construct, and use of computer systems in an efficient way.

Nowadays, medicine and technology are investigating new applications that extend the possibilities for treatment of various disorders and diseases, exploring the new information and communication technologies. Recently, we had some interesting results that are relevant in opening new possibilities to the Virtual Reality (VR) application in treatment of brain damage patients. Alternatively, the VR applications are trying to integrate the new tendencies of Artificial Intelligence to increase the human-computer interaction.

In 2000, the Census showed that the number of people who declared some disabilities is around 24.5 million, approximately 14.5% of the Brazilian population. This percentage is divided into: 8.3% of mentally handicapped, 4.1% of physical disabled, 22.9% of motor disabled, 48.1% of blindness and 16.7% of deafness [Sarraf, 2009]. These high percentages require an expansion of activities for efficient treatment and social reintegration for these people.

Therefore, this project proposes the creation of an inter-institutional group dedicated to the creation of virtual environments that exploit the new technologies of Virtual Reality (VR). It also integrates Intelligent Interface Agents to create sociable tools to support the treatment of patients who have suffered brain damage resulting in cognitive disabilities.

The target of this project is to promote a collaborative network integrating groups from two specific areas: the Computer, by researchers interested in Virtual Reality and Artificial Intelligence; and another group of physicians interested in testing these environments with patients with cognitive disabilities resulting from different brain accidents and diseases. The virtual environments can stimulate the achievement of daily activities associated to executive functions, readapting patients to social living.

In the next sections we will briefly present the basic theories related to the research, some preliminary results that were obtained from previous work, and the possible multidisciplinary inter-institutional groups that have works on related topics, in the context of INRIA (Institut National de Recherche en Informatique et Automatique).

2. The Intelligent Interface Agents

The concept of intelligent agents emerged from the Artificial Intelligence (AI). Agents are the main tool of a futuristic interface, where computers communicate with users exploring similar expressions used in the human communication. The challenge of the complex systems construction is dealt with the systems control. An agent oriented paradigm reinforces the software flexibility and the agents´ social possibilities. The interface agents act as assistants, supporting users to achieve tasks. They can be used to help instruction and education processes, replacing humans in software applications, or assisting users to navigate in complex environments. These animated characters can be employed as a natural way to provide users with additional information.

3. Virtual Reality

Virtual Reality includes advanced technologies of interface, immersing the user in environments that can be actively interacted with and explored. Moreover, the user is able to accomplish navigation and interaction in a three-dimensional synthetic environment generated by computer, using multi-sensory channels. In this case, diverse types of stimuli can be transmitted by specific devices and perceived by one or more user's senses [Burdea,2003].

VR applications are being explored in different areas through the development of pilot projects that aim mainly to discuss and experience the possibilities offered by this technology.

The integration of VR with the intelligent agents was the subject of research made by Bouchard et al. (2006), Araújo et al. (2009) and Yang et al. (2008). These studies used the interface agents to increase the user involvement with the activities in virtual environments and discussed the technology limitations and advantages.

4. Cognitive Rehabilitation through Virtual Environments

The introduction of VR is being considered as a new aid to diminish the difficulties involved in the cognitive rehabilitation (CR) process, making these activities more friendly and fun.

Recently, several experiments and described case studies have been performed, yielding positive results. These researches as presented by Dawson et al. [Dawson,

2008] and Broeren et al.[Breoeren, 2008] encourage the studies for the integration of these fields. Corroborating these results, the experiments developed in recent years by our inter-institutional research group with patients with neuropsychiatric disabilities [Costa, 2004], [Cardoso, 2006], [Piovesana Neto, 2007] are also showing very interesting results. These pilot studies sought to discuss and experiment with the possibilities offered by this technology. In these contexts, VR is allowing therapeutic practitioners to help their patients in a number of innovative ways, offering new approaches to old questions and augmenting the effectiveness of consolidated methodologies. The use of VR technology, with an interface nearer to reality, could reduce the gap between patients and daily life tasks, decreasing fear of errors.

The potential of the interface agents to facilitate human-computer interaction, stimulated the development of an environment where these agents support the activities carried out by patients and may generate a higher level of motivation to perform the tasks. The environment is a house divided into rooms (Figure 1a); each room provides activities for cognitive stimulation. The interface agent is in the kitchen (Figure 1b), where he supports the execution of tasks associated with daily activities.

The test's first results demonstrated that we have some technical problems associated with the integration of this reactive agent in this environment and its communication with the users. In that case, we need to define new intelligent relationships and include new criteria to control the agent behavior.

(a) **(b)**

Figure 1. Façade of the house and the main entrance of the kitchen with the agent image [Piovesana Neto, 2007].

5. Integration possibilities with the INRIA groups

Currently, our team is composed of researchers/teachers from the Department of Informatics and Computer Science of the State University of Rio de Janeiro (UERJ), Program of Systems Engineering of COPPE-Federal University of Rio de Janeiro (UFRJ), the Faculty of Medical Sciences from UERJ, and the Laboratory of Neuropsychology, Cognition and Cognitive Rehabilitation of the UFRJ Hospital. Our previous experience in the development of studies in the cognitive rehabilitation, supported by 3D virtual environments allowed us to continue toward the same direction of interdisciplinary researches.

In the context of the INRIA, we verified that some groups have many points of adhesion to this proposal. The “Alcove Project” (INRIA, 2007) works with the

construction of 3D interfaces because of the possibilities for cooperation offered by them and has obtained significant results in the navigation and interaction in 3D environments. This group also developed studies in the Cognitive Rehabilitation area: the "Reactive ANR TecSan Project", which has been developed in conjunction with the Foundation HOPALE (Medical Institute specialized in rehabilitation). This project also involves other partners in the CEA, as the Idees3com Company.

Another group that stands out at INRIA in the context of our line of interest is the "Bunraku Group" (INRIA, 2007), which examines the possibilities for interaction in 3D virtual environments and also studies the impacts of the inclusion of intelligent agents in virtual environments.

6. Final Comments

This paper outlined the base theoretical aspects of a project that is being developed by researchers from two public universities in Rio de Janeiro and its promising results.

This proposal has two approaches; the first one is associated with the technical questions, studies and the application of intelligent agents in 3D virtual environments. The other assesses the impact of these agents in the interaction of users within these environments.

Some groups that developed studies in the context of INRIA have good prospects of integration with our group. This collaborative association can open new possibilities to enlarge the development and testing of 3D virtual environments in the cognitive rehabilitation field.

References

Araújo, A. S. ; Carvalho, L.A.; Costa, R. M. E. M. (2009). "The influence of intelligent characters on the users' navigation through a three-dimensional Virtual Environment". *Presence,* (submitted for publication).

Bouchard, S., Dumoulin, S., Chartrand-Labonté, G., Robillard, G., Renaud, P., (2006). Perceived realism has a significant impact on presence, *Cybertherapy Conference,* Quebec.

Broeren, J., Bellner, A-L., Fogelberg, M., Göransson,O., Goude,D., Johansson, B., Larsson, P., Pettersson, K., Rydmark, R., *(2008),* "Exploration of computer games in rehabilitation for brain damage", *Seventh International Conference on Disability, Virtual Reality and Associated Technologies*, Maia, in: http://www.icdvrat.reading.ac.uk/2008/index.htm#abstracts-section-i, visited on 2009-March.

Burdea, G. C., Coiffet, P., (2003), *Virtual Reality Technology*, J. Wiley & Sons Inc.

Cardoso, L., Costa, R. M., Piovesana, A., Costa, M., Penna, L. (2006), "Using Virtual Environments for Stroke Rehabilitation", *IEEE- 5th International Workshop on Virtual Rehabilitation*, New York, p. 1-5.

Costa R.M., Carvalho, L. A. (2004). "The Acceptance of Virtual Reality Devices for Cognitive Rehabilitation: a report of positive results with schizophrenia", *Computer Methods and Programs in Biomedicine*, Elsevier Science, 73(3), p.173-182.

Dawson, A., Buxbaum, L., Rizzo, A., (2008), "The Virtual Reality Lateralized Attention Test: Sensitivity and validity of a new clinical tool for assessing hemispatial neglect", *Virtual Rehabilitation 2008*, Vancouver, in: http://www.virtual-rehab.org/2008/advanced_program.html, visited on 2009-March.

INRIA, (2007) "Scientific Activity Reports", INRIA.

Piovesana Neto, A. (2007), "O desenvolvimento de ambientes virtuais para a Reabilitação Cognitiva", *Workshop de Iniciação Científica do IX Symposium on Virtual Reality-SVR2007*, Petrópolis, (in Portuguese).

Sarraf, V. P. (2009) "A inclusão dos deficientes visuais nos museus", in: http://museuacessivel.incubadora.fapesp.br/portal/publicacoes/artigos/inclusao%20dos%20deficientes%20visuais%20nos%20museus.doc, visited on 2009-March.

Yang, H., Pan, Z., Zhang, M., Ju, C., (2008). "Modeling emotional action for social characters", *The Knowledge Engineering Review*, Cambridge University Press, 23(4), 321-337.

Modélisation et analyse par vérification formelle et simulation conjointe des réseaux de régulation géniques

Olivier Roux
IRCCyN, (UMR CNRS 6597)
1 rue de la Noë, BP 92101 F-44321 Nantes Cedex 3 (FRANCE)
roux@irccyn.ec-nantes.fr

Résumé Pour comprendre les évolutions dans les systèmes vivants, les contrôler et éventuellement éviter ou remédier à certains comportements mauvais, la biologie synthétique s'attache à modéliser ces processus dynamiques afin de les analyser précisément. Plusieurs types de modélisation sont mis en œuvre, parmi lesquels notre démarche se caractérise par une *approche hybride* (mélangeant des aspects discrets et continus) basée sur la formalisation discrète de René Thomas, et elle s'exprime en utilisant l'approche algébrique et temporelle du π*-calcul stochastique*. Nous illustrons succinctement les résultats obtenus sur un exemple simple et nous montrons les bénéfices que nous pouvons tirer de cette nouvelle formulation.

Abstract. In order to understand development of living organisms, and to control them and possibly to avoid or remedy some unexpected behaviours, synthetic biology intends to build models of these dynamical processes to accurately analyse them. Several modelling frameworks were designed for such a purpose, among which the characteristic of our approach is that it leads to *hybrid models* (mixing discrete and continuous features) which are founded upon the discrete formalism of René Thomas. Furthermore, it is algebraically expressed thanks to the *stochastic* π*-calculus* including the temporal aspects. We shortly highlight the results achieved from a simple example and we show the benefits that we brought from this new method.

1 Introduction

Que ce soit au niveau intra-cellulaire, ou à d'autres échelles, les phénomènes biologiques révèlent des comportements très variables d'entités dynamiques interagissant de façon complexe. Pour comprendre les évolutions des systèmes vivants, les contrôler et éventuellement éviter ou remédier à certains comportements mauvais, la biologie synthétique s'attache à modéliser ces processus dynamiques afin de les analyser précisément. Ces analyses consistent à imaginer et mettre en œuvre les prévisions d'évolutions du système en fonction de nombreux paramètres de natures très variées, à la fois internes au système considéré et révélateurs de l'environnement de ce système. La complexité vient notamment de la diversité des actions/réactions mutuelles des entités et de leur multiplicité, ainsi que du très grand nombre de paramètres dont la connaissance des valeurs est difficile à obtenir mais essentielle pour mener à bien les études envisagées.

Plusieurs démarches de modélisation sont mises en application, parmi lesquelles les équations différentielles constituent une approche classique dans ce type de recherche, mais qui se heurte, de façon plus ou moins rédhibitoire, au délicat problème

d'estimation d'un nombre important de paramètres. Une autre approche, que nous privilégions, consiste à s'attacher à une représentation par des systèmes états-transitions. Ce mécanisme se prête assez bien à des extensions parmi lesquelles celle qui prend en compte la dimension temporelle des phénomènes — non seulement par les aspects qualitatifs (chronologiques), mais aussi quantitatifs (chronométriques) — nous paraît particulièrement adaptée.

Sans en constituer une synthèse, les idées prospectives, originales et pas encore publiées, que nous présentons ici se basent sur une somme importante d'expériences déjà menées ou en cours sur la modélisation et la vérification de ces réseaux génétiques et avec passage à l'échelle.

Nous présentons notre démarche dans la section suivante en décrivant le cadre de modélisation que nous avons établi sur une *approche hybride* (mélangeant des aspects discrets et continus) et par une expression en π*-calcul stochastique* [7]. Puis, nous montrons les résultats de simulation que nous avons obtenus sur un exemple très simple (à partir d'une modélisation à six états de la régulation de deux gènes intervenant de façon déterminante dans les comportements de la bactérie pathogène *Pseudomonas æruginosa*). Avant de conclure, nous indiquons comment tirer profit de cette nouvelle représentation algébrique.

2 Description du modèle

Fondamentalement, notre démarche, qui se place dans la lignée du principe de modélisation discrète de René Thomas [9], renforce les travaux que nous avons effectués et proposés [1,2] sur la modélisation hybride. Comme nous allons le montrer, ce récent travail et les perspectives qu'il ouvre permettent d'affiner certains résultats que nous obtenions dans la première phase de notre démarche. Il se place en outre dans l'esprit des recherches menées actuellement dans notre équipe [3,4,5].

2.1 Motivations

Nous partons donc d'un espace d'états où chaque état est constitué, selon le principe de René Thomas, à partir de l'ensemble des valeurs abstraites qui concernent, pour chaque gène, les seuils (discrets) au delà desquels il commence à exercer une influence positive (activateur) ou négative (inhibiteur) sur un autre gène. Notre originalité consiste à *considérer le temps qui s'écoule dans chacun de ces états* jusqu'au passage dans un autre état au bout d'une durée donnée par des paramètres de délais de l'activation ou de l'inhibition de chaque gène. Dans la modélisation hybride que nous avons ainsi conçue et présentée dans [1,2], le nombre d'états et de paramètres mis en jeu est très important et ceci constitue actuellement un facteur limitatif très contraignant pour la plupart des analyses que nous voudrions pouvoir effectuer sur des applications biologiques réelles. C'est pourquoi nous avons imaginé la traduction dont nous allons présenter ici les principes et qui nous permet, d'une part, d'effectuer des simulations pouvant induire certains résultats et, d'autre part de déboucher, grâce à des transformations algébriques, sur des simplifications intéressantes.

2.2 Principe de la traduction en π-calcul

Le principe de traduction est assez simple et direct puisqu'il se résume à transformer chaque état en un processus du π-calcul stochastique [8,7] qui s'exécute en passant un certain délai avant de se transformer en un autre processus correspondant à un nouvel état vers lequel il était susceptible de transiter. De fait, la transition effectivement choisie dépend des délais de transitions qui se trouvent donc être des paramètres comme dans la modélisation hybride. Pourtant, nous en tirons profit des façons suivantes.

2.3 Résultats de simulation

L'application de la démarche esquissée ci-dessus, en utilisant l'outil SPiM [6], donne des résultats de simulation (ici, par exemple, sur cent exécutions) qui sont illustrés par la figure 1. Deux états particuliers du système ont été sélectionnés : d'une part, un état, choisi comme état initial, et appartenant à un cycle potentiel d'exécution ; et d'autre part, un état puits (à partir duquel il n'y a pas de transition vers un autre état). En fonction du temps, figure le nombre d'exécutions se trouvant dans l'un des deux états considérés[1]. Pour rendre les résultats plus lisibles, nous avons exclu les quatre autres états du modèle discret.

Ces résultats correspondent à des valeurs de paramètres de délai fixés plus ou moins arbitrairement mais qui mettent en évidence deux types de comportements fondamentalement différents, c'est-à-dire conduisant dans un cas à une attraction très forte vers un état d'absorption et la diminution du nombre de passages par l'état appartenant au cycle (à gauche), et un équilibre entre les exécutions conduisant vers cet état d'absorption ou à des répétitions sempiternelles de passage par le même état du cycle (à droite).

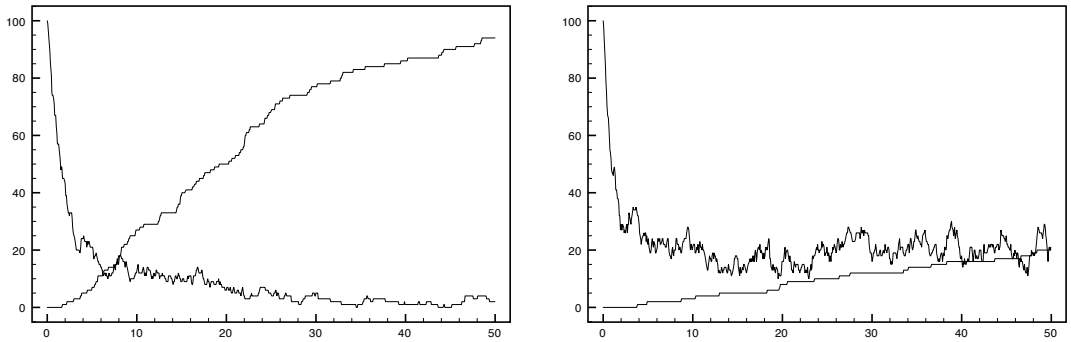


Figure 1 : Résultats obtenus pour différents délais d'activation et inhibition

2.4 Simplifications algébriques

Un avantage de la formulation algébrique (en π-calcul) du comportement global est aussi qu'elle permet de mettre en œuvre des opérateurs pour calculer des équivalences.

[1] A l'instant initial, toutes les exécutions sont donc dans l'état initial.

En particulier, nous pensons que les cycles, qui correspondent à des phénomènes oscillatoires, pourraient être reconnus comme tels et, dans certains cas, être abstraits en des phases dans lesquels on ne s'intéresse qu'au temps qui s'écoule dans ces phases. De même, alors qu'on ne considère actuellement qu'un unique processus, résultant de la mise à plat de l'ensemble des exécutions, on pourrait faire en sorte d'ajouter des mises en parallèle avec des comportements externes de type environnemental ou multi-échelle.

3 Conclusion

Les travaux brièvement présentés ici sont évidemment loin d'être définitivement aboutis, mais ils constituent, au contraire, un état provisoire de recherches en cours et ils se placent naturellement dans une démarche cohérente allant de la modélisation des systèmes complexes de régulation d'après René Thomas à la prise en compte des délais d'interaction jusqu'à une représentation algébrique permettant à la fois des vérifications formelles et des simulations pour la synthèse de paramètres. Précisément, cette cohérence de diverses approches laisse entrevoir des perspectives assez riches, notamment pour arriver à pallier les problèmes d'explosion de l'espace d'analyse, qui sont dûs à la complexité des systèmes considérés.

Références

1. Jamil Ahmad, Gilles Bernot, Jean-Paul Comet, Didier Lime, and Olivier Roux. Hybrid modelling and dynamical analysis of gene regulatory networks with delays. *ComPlexUs*, 3(4) :231–251, October 2007.
2. Jamil Ahmad and Olivier Roux. Invariance kernel of biological regulatory networks. In *IEEE Int. Conference on Bioinformatics and Biomedicine (BIBM 2008)*, pages 135–140, Philadelphia (Penn.), USA, nov 2008. IEEE Computer Society.
3. Jonathan Fromentin, Jean-Paul Comet, Pascale Le Gall, and Olivier Roux. Analysing gene regulatory networks by both constraint programming and model-checking. In *EMBC07, 29th IEEE EMBS Annual International Conference*, pages 4595–4598, August 23–26 2007.
4. Mylène Maurin, Morgan Magnin, and Olivier Roux. Modeling of genetic regulatory network in stochastic pi-calculus. In *International Conference on Bioinformatics and Computational Biology (BICOB 2009)*, LNBI, New Orleans (Louisiana) USA, apr 2009. Springer Verlag. to appear.
5. Loïc Paulevé, Morgan Magnin, and Olivier Roux. Refining dynamics of gene regulatory networks in a stochastic pi-calculus framework, 2009. submitted.
6. Andrew Phillips. SPiM, 2007. `http://research.microsoft.com/~aphillip/spim`.
7. Andrew Phillips and Luca Cardelli. Efficient, correct simulation of biological processes in the stochastic pi-calculus. In *Computational Methods in Systems Biology (CMSB 2007)*, volume 4695 of *LNCS*, pages 184–199. Springer, September 2007.
8. Corrado Priami. Stochastic pi-calculus. *The Computer Journal*, 38(7) :578–589, 1995.
9. René Thomas, Denis Thieffry, and Marceline Kaufman. Dynamical behaviour of biological regulatory networks–i. biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bull Math Biol*, 57(2) :247–276, Mar 1995.

Framework for Interactive Medical Imaging Applications

André Ferreira Bem Silva, Tiago H. C. Nobrega , Diego Dias Bispo Carvalho , Renan Teston Inácio , Aldo von Wangenheim

[1]LAPIX – Laboratory for Image Processing and Computer Graphics
UFSC – Federal University of Santa Catarina
Florianópolis, Santa Catarina, Brazil

{tigarmo,diegodbc,awangenh}@inf.ufsc.br

Abstract. *We propose a set of libraries with the goal of providing a framework to build medical applications. Instead of using the existing packages in libraries to perform the tasks of image and volume handling and visualization, all code has been written for scratch. We expose the reasoning for this decision while presenting a medical workstation and a few physical simulation prototypes as example applications.*

1. Introduction

Medical Computer-Aided Diagnostic tasks are critical in the sense that as the technology in the field advances, fewer are the possibly vital errors in medical judgement. We present the design and implementation of a radiological 3D workstation (abbreviated as "WS") together with a set of libraries for basic algorithm and rendering tasks. The set was developed in-house in its entirety instead of adopting existing solutions, and our reasons for this are presented and discussed.

1.1. Motivation

There are many established libraries and toolkits that aid in the creation of medical software, such as MITK, DCMTK, ITK and VTK and their extensions. In this sense, there should be good reasons for coding teams to implement (and maintain) everything from scratch.

In our case, we needed different, specific functionalities that the available APIs (at the time) didn't provide, such as GPU access for physical simulations (*GPGPU*) and multithreaded implementations of specific algorithms to take advantage of the multiple cores in today's desktops. We needed the understanding of every layer in the set of algorithms, from the low-level data-structure manipulation to the final rendering, in order to identify hot spots and bottlenecks (to obtain high performance). Such knowledge is vital when the time comes to port the application and the libraries to another architecture. Additionally, the comparison between CPU and GPU versions of an algorithm can be easier if the same environment and methodologies were used to implement both.

Finally, when there are different teams in the same group working on different tasks, having a common set of libraries developed in-house has the advantage of better feedback (framework and application developers working together). This way, problems are solved more easily and the learning curve is smaller (easy to share experience).

2. C3DE libraries architecture

In addition to the needs specified in the previous section, the Cyclops 3D Environment libraries are all written in standard C++ using cross-platforms APIs for widgets and 3D rendering. They're currently supported on Linux® and Microsoft Windows®, and are compatible with 32 and 64-bit x86 architectures.

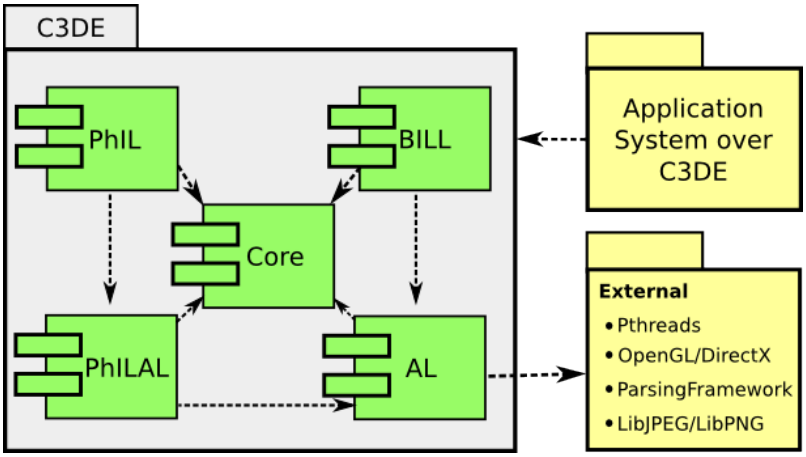


Figure 1. C3DE basic architecture

2.1. Core

The Core's tasked with base mathematical structures such as matrices and vectors, volume abstractions, image definitions, 3D spatial management such as the intersection of arbitrary meshes with planes and lines, and common utilities like smart pointers. For generical data and data structures treatment, the Core's most complex structures make heavy use of template metaprogramming.

2.2. AL

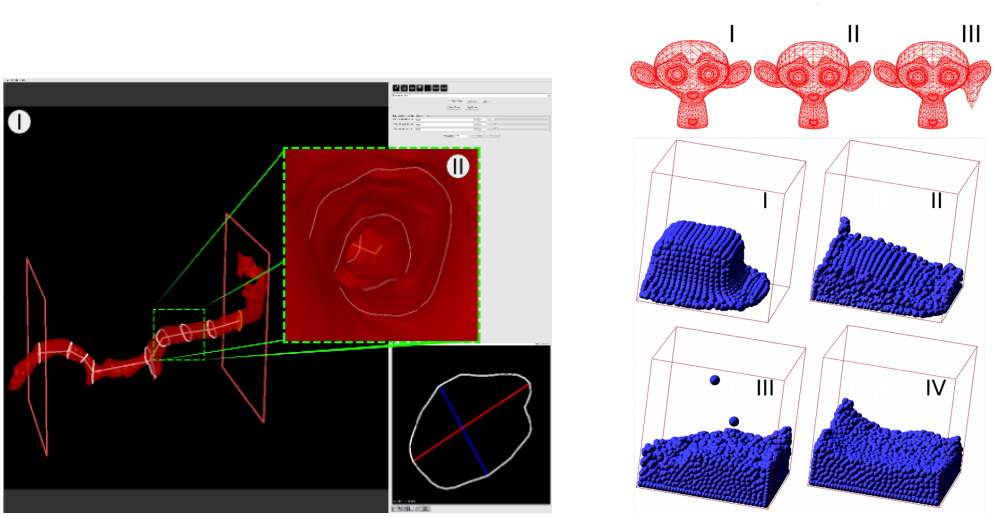
The Abstraction Layer, as the name implies, wraps the Core's complex metaprogramming with the standard Object-Oriented paradigm. It provides factories for images, volumes and three-dimensional reconstructions, drawable objects, shaders and texture handling through OpenGL, and standard multi-planar reconstruction.

2.3. PhIL

PhIL stands for Physical Interaction Layer and provides structures that extend the basic functionalities found in Core with physical properties. For instance, the standard Half-Edge mesh is extended to support deformations using the mass-spring model [Provot 1995]. Currently, the library also supports fluid simulations with the particle-based Smoothed Particle Hydrodynamics (SPH) method [Muller et al. 2003].

2.4. PhILAL

The Physical Interaction Abstraction Layer binds PhIL and AL, with multithreaded and GPU-enabled versions of PhIL algorithms. Like AL, there are also factory methods to quickly prototype and render PhIL structures, like deformable meshes (mass-spring) and liquid simulation (SPH), sawn in steps on 2(b).



(a) (b)

Figure 2. Centerline approximation (a) and physical simulation prototypes (b).

2.5. BILL

The Binding Input Layer Library (BILL) defines interfaces to connect input devices and the applications. Its core purpose is on the translation of events from devices such as a mouse into meaningful actions for the application, through event handlers.

BILL itself depends only on Core for metaprogramming facilities and AL for multithreading — this is because the library does not provide devices or actions, only their interfaces and the basic mechanisms of their integration.

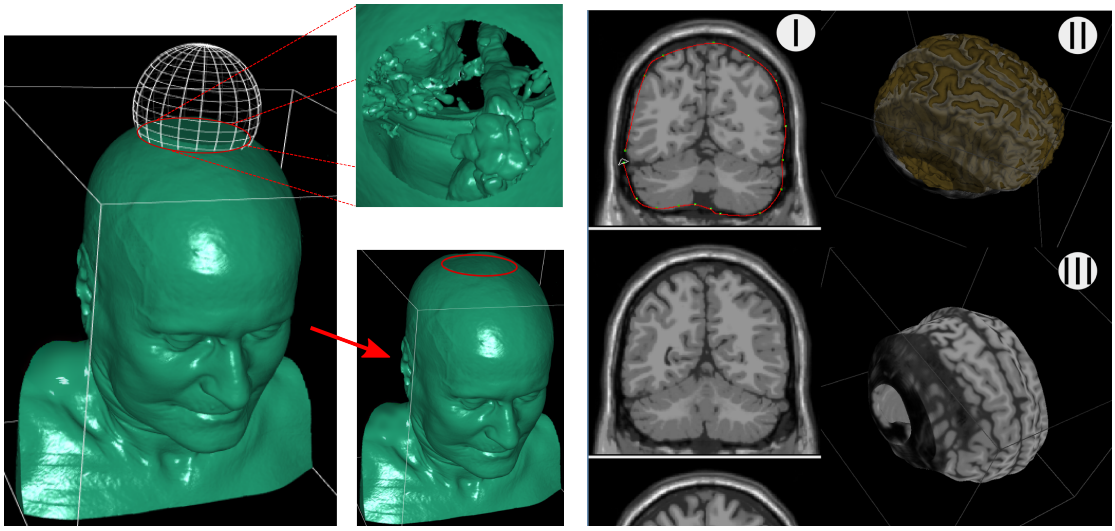
3. Workstation

The WS is built on top of the framework and uses extensively all Core and AL resources. Its features include reconstruction and segmentation, approximate centerline calculation, Volume-Of-Interest oriented reconstruction, Curvilinear reconstructions and fiber tracking.

Given two user-positioned planes of interest, the approximate centerline generation algorithm computes a set of line segments that pass roughly through the center of a structure by firing lines between the planes and computing the lines' intersection with the structure, recursively (Figure 2(a)).

The curvilinear reconstruction [Bastos et al. 1995] aims at reconstructing segments that are difficult to measure and distinguish with simple grayscale thresholding. By placing a set of points on two-dimensional slices (Figure 3(b)) representing regions of interest, the user obtains a curved surface that is interpolated from the points with Catmull-Rom splines.

The Volume-Of-Interest reconstruction works on objects to remove unwanted portions, or segment it manually. Its counterpart is the segmented reconstruction that differentiates distinct objects through thresholding on the CT or MRI volume. The inputs are the density limits (lower and upper) to be used, with optional fields for noise filters for



(a) (b)

Figure 3. VOI-oriented (left) and curvilinear (right) reconstructions.

relatively small objects and minimum and maximum absolute volumes. The VOI feature uses spheres and boxes to indicate the areas of interest (Figure 3(a)).

4. Conclusions and Future work

The libraries can be used for fast prototyping (like the physical simulation ones) while being robust enough to implement a medical workstation with standard features. We plan to extend the framework with new libraries such as a Network Interaction Layer over BILL for collaborative sessions over the internet. Additionally, we plan to build a sandbox prototype with PhIL and PhILAL to easy visualize and simulate interactions of solid and deformable objects and fluids. The WS is always maintained with the latest version of the framework to always maintain compability with it.

There are some interesting related work, like the ones that INRIA group made in medical imaging software. It would be interesting to exchange information and development experience with them and/or other groups related to the area.

References

Bastos, A. C., Korah, I. P., Cendes, F., Melanson, D., Tampieri, D., Peters, T., Dubeau, F., and Andermann, F. (1995). Curvilinear reconstruction of 3d magnetic resonance imaging in patients with partial epilepsy: A pilot study. *Magnetic Resonance Imaging*, 13(8):1107 – 1112. Workshop on Magnetic Resonance Techniques and Epilepsy Research.

Muller, M., Charypar, D., and Gross, M. (2003). Particle-based fluid simulation for interactive applications. In *SCA '03: Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 154–159, Aire-la-Ville, Switzerland, Switzerland. Eurographics Association.

Provot, X. (1995). Deformation constraints in a mass-spring model to describe rigid cloth behavior. In Davis, W. A. and Prusinkiewicz, P., editors, *Graphics Interface '95*, pages 147–154. Canadian Human-Computer Communications Society.

High Performance Reconfigurable Computing for Bioinformatics Applications

Alba M. A. Melo[1], Maria E. T. Walter[1], Carla C. C. Koike[1], Ricardo P. Jacobi[1]

[1]Department of Computer Science – University of Brasilia (UnB)
Campus UNB – ICC Norte – sub-solo - 70.910-900 – Brasilia – DF – Brazil

{albamm, mia, ckoike, rjacobi}@cic.unb.br

Abstract. *Compute and data intensive applications have been usually executed in high performance computing platforms. Nevertheless, with the exponential growth of the amount of data produced in some research domains, even these parallel systems are unable to produce results in reasonable time. Therefore, in order to further accelerate these applications, parts of them are migrated to hardware and implemented, for instance, in reconfigurable hardware such as FPGAs. Many bioinformatics applications are good candidates to explore the benefits of high performance reconfigurable computing platforms, since they are often very compute intensive, dealing with enormous amounts of data. This paper presents the research already conducted at UnB in the area of High Performance Reconfigurable Computing for Bioinformatics, discusses open problems and presents perspectives for future joint international research.*

1. Introduction

Nowadays, biological data is being produced in laboratories all over the world in a rate higher than the speed needed to process them. Public repositories for genomic data, such as the one maintained at NCBI (*National Center for Biotechnology Information*), have attained exponential growth rates. In this scenario, successful biology and medicine research labs are the ones which are able to produce accurate results very rapidly.

Parallel processing can be used to produce results faster, bridging the gap between the generation of genomic data and its analysis. Also, with parallel approaches, exact compute-intensive methods can be executed in reasonable time. However, using parallel processing in Bioinformatics is not straightforward since the problems are often solved by complex methods, with a great amount of data dependency. To further accelerate the production of results, specific hardware can be used in conjunction with parallel processing, generating high performance reconfigurable computing platforms.

The remainder of this paper is organized as follows. In section 2, we briefly describe the research developed by our group in this domain. Section 3 presents ongoing research. In section 4, we discuss the expected scientific impact. Finally, in section 5, we discuss perspectives of new collaborations and conclude the paper.

2. Research Accomplishments

From 2002 to 2009, our research group has been working on parallel algorithms, reconfigurable architectures and load balancing strategies for bioinformatics problems.

The first treated problem was the pairwise sequence alignment (Gusfield 1997), since it is a basic building block for solving more complex problems. A pairwise sequence alignment is defined for two sequences, and is obtained by putting one

sequence above the other, such that their similarities can be easily identified. Studying the SW dynamic programming exact algorithm proposed by (Smith and Waterman, 1981) to solve this problem, it was clear that, besides its high computing power needs, memory requirements were also a huge issue. For instance, to compare 23MBP (Mega Base Pairs) sequences, we would need at least 500 TB of memory.

Concerning the dynamic programming solution, we first proposed, in (Melo et al, 2003), a parallel heuristic variant of SW based on Distributed Shared Memory that executes in linear space. A speedup of 4.58 was achieved in an 8-processor cluster. In (Batista et al., 2004), we modified this heuristic by adding a blocking factor obtaining a speedup of 7.28, in the same platform, for the same sequences. Also, a new exact variant is proposed in (Boukerche, Melo, Ayala-Rincon and Walter, 2007) that produced almost linear speedups. Our first work using FPGA to accelerate SW was presented in (Jacobi et al., 2005). Simulation results for Altera devices produced speed-ups of two ordens of magnitude. In (Boukerche, Melo, Sandes and Ayala-Rincon 2007), we proposed a parallel exact variant of the SW algorithm that reduces memory requirements, which was able to compare 1.6MBP sequences.

Finally, we proposed z-align (Batista, Boukerche and Melo, 2008), a parallel exact variant of the SW algorithm that, based on a new concept called divergence, was able to compare 23MBPx24MBP sequences. As far as we know, this is the first work where sequences longer than 3MBP are compared with an exact affine-gap SW variant. For the 3MBP case, we reduced the execution time from 3 days and 8 hours (1 processor) to less than 3 hours (64 processors).

We also proposed in (Boukerche, Correa, Melo, Jacobi, Rocha 2007a) a hardware accelerator to execute an SW variant. In this case, our simulated FPGA prototype achieved a speedup of 246.9 over the software implementation, in a 100MBP x 100BP sequence comparison.

In order to execute BLAST in grid environments, we proposed PackageBLAST (Sousa and Melo, 2006), which distributes BLAST queries among the grid nodes using multiple task allocation strategies. In this case, a speedup of 11.28 over the best machine was achieved in a 16-machine heterogeneous platform.

In (Boukerche, Correa, Melo, Jacobi, Rocha 2007b) we propose the execution of the DIALIGN (Morgenstern et al. 1998) sequence comparison algorithm in an FPGA-based architecture. A speedup of 383.41 in a 160KBP x 190KBP comparison was obtained, reducing the execution time from more than 3 hours to 28.83s.

Besides, we developed a Peer-to-Peer (P2P) architecture for bioinformatics applications that executed in 13 heterogeneous machines distributed by three geographically distinct research institutions of Brasilia. As a case study we used BLAST (Ribeiro, Walter, Togawa, Costa, Pappas 2008). In the experiments, we obtained a gain performance of 80%, by comparing the time of running BLAST with input files containing at most 800 biological sequences against a big database (NCBI nr), both in a stand alone machine (the best machine) and in our P2P architecture.

3. Ongoing research

We are currently interested in the Multiple Sequence Alignment Problem (MSA), already shown to be NP-Complete. A MSA of $k > 2$ sequences $S = \{S_1, S_2, ..., S_k\}$ is obtained in such a way that chosen spaces (*gaps*) are inserted into each of the k sequences so that the resulting sequences have the same length l. Then, the sequences are arranged in k rows of l columns each, so that each character or space of each sequence is in a unique column (Gusfield 1998). We are currently working on a Parallel Island Injection Genetic Algorithm to solve this problem in a reasonable time, using no

information about the classes of the sequences being compared. Now, we are designing a parallel version of the linear space variant of DIALIGN, that will be used for MSA.

We are also interested in extending our P2P architecture to other research institutions of the MidWest Region of Brazil, and to implement other applications in this framework such as efficient computational methods based on probabilistic theories to identify non-coding RNAs.

Finally, we are implementing a Sequence-Profile alignment solution that uses Hidden Markov Models (HMMs) in FPGA. In this case, a biological sequence is compared to a profile that characterizes a family of sequences.

4. Expected Scientific Impact

Pairwise and Multiple Sequence Alignment are basic operations in Bioinformatics and many sophisticated methods use them as a fundamental building block. Therefore, improving the accuracy of these alignments will certainly improve the accuracy of the methods that rely on them.

Even though the SW algorithm for pairwise sequence alignment has quadratic time and space complexity, the heuristic method BLAST is much more used in practice, since it can produce good average results very quick. However, as long as the sizes of highly similar sequences being compared increase, the results produced by BLAST are not so good. In (Boukerche, Batista and Melo 2009), we presented the case of the Anthrax comparison, where our z-align strategy was able to create an alignment of 5220960 bases between two different *Bacillus anthracis* strains (Ames and Sterne), resulting in a similarity of 98.1%. The same comparison using the BlastN program generated an alignment of size 36159, with a similarity of 0.69%, which does not correspond to the reality. Thus, we claim that the use of exact methods for pairwise sequence comparison can lead to a significant higher accuracy. The benefits of z-align can be better exploited if some parts of it are implemented in hardware, thus reducing drastically its execution time.

Although Multiple Sequence Alignment problem is known to be NP-Complete, the use of high performance reconfigurable computing to solve this problem can accommodate more elaborated methods that potentially could lead to better accuracy.

5. Perspectives of New Collaborations

The research conducted by our group has a multi-disciplinary characteristic, involving 4 research domains: parallel algorithms, high performance computing infrastructure, reconfigurable architectures and bioinformatics.

By now, we have software-only parallel solutions and hardware-only solutions for some bioinformatics problems. Nevertheless, the ideal situation would be a complete hardware/software platform where many bioinformatics problems could be solved in restricted time. In order to attain this complete solution, international collaborations are highly desirable. Research cooperation with groups working on load balancing, reconfigurable architecture design, hardware/software codesign, complex systems modeling, and large data sets could be very interesting to achieve an efficient hardware/software solution.

Another research cooperation area that can be envisaged is reconfigurable supercomputing, where FPGAs are used to accelerate algorithms in a variety of applications domains, such as cryptography, video compression and virtual reality, among others. We are also currently working on H.264/AVC video encoding with FPGAs, which has huge processing requirements.

Probabilistic inference, as employed in Bayesian modeling, is another domain where there is exact and approximated calculations due to big time and space complexity. Nowadays, Bayesian methods, including Bayesian Networks, Markov Chains, Hidden Markov Model, among others, are widely used in several applications, including bioinformatics, and we are highly interested in developing hardware/software platforms to perform these calculations in restricted time.

References

Batista, R.B., Silva, D. N., Melo, A. C. M. A., Weigang, L. (2004) "Using a DSM application to locally align DNA sequences", In: *CCGrid*, Chicago, p.372-378.

Batista, R. B., Boukerche, A., Melo, A. C. M. A. (2008), "A Parallel Strategy for Biological Sequence Alignment in Restricted Memory Space". *Journal of Parallel and Distributed Computing*, v.68, n. 5, p. 548-561.

Batista, R. B., Boukerche, A., Melo, A. C. M. A. (2009) "Exact Pairwise Alignment of Huge Biological Sequences with the Z-Align Parallel Strategy", In*: Workshop NIDISC at IPDPS*, Rome, Italy, May.

Boukerche, A., Melo, A. C. M. A., Ayala-Rincon, M., Walter, M. E. M. T, (2007) "Parallel Strategies for Local Biological Sequence Alignment in a Cluster of Workstations". *Journal of Parallel and Distributed Computing*, v. 67, n. 2, p. 170-185, 2007.

Boukerche, A., Melo, A. C. M. A., Sandes, E., Ayala-Rincon, M. (2007), "An exact parallel algorithm to compare very long biological sequences in clusters of workstations". *Cluster Computing*, v. 10, n. 1, p. 187-202.

Boukerche, A. Correa, J. M., Melo, A. C. M. A, Jacobi, R. P., Rocha, A. F. (2007a) Reconfigurable Architecture for Biological Sequence Comparison in Reduced Memory Space. In: *Workshop NIDISC at IPDPS*, Long Beach, USA, March.

Boukerche, A. Correa, J. M., Melo, A. C. M. A, Jacobi, R. P., Rocha, A. F. (2007b), "An FPGA-based Accelerator for Multiple Biological Sequence Alignment with DIALIGN". In: *Int. Conf. on High Performance Computing*, Goa, India, p. 71-82.

Gusfield, D. (1997), "*Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology*", Cambridge University Press.

Jacobi, R. P., Ayala-Rincon, M., Carvalho, L. G., Quintero, C. H. L. ; Hartenstein, Reiner W., "Reconfigurable systems for sequence alignment and for general dynamic programming". In: *Genetics and Molecular Research*, São Paulo, Brasil, v. 4, n. 3, p. 543-552, 2005.

Melo, R. C. F, Walter, M. E. M. T., Melo, A. C. M. A., Batista, R. B., Santana, M. N. P., Martins, T., Fonseca, T., (2003), "Comparing Two Long Biological Sequences Using a DSM System", In 9th *Euro-Par*, Klagenfurt, Austria, p. 517-524.

Ribeiro, E., Walter, M. E. M. T., Togawa, R. C., Costa, M. M., Pappas, G. (2008) "p2pBIOFOCO: Proposing a Peer-to-Peer System for Distributed BLAST Execution", In 10th *HPCC-08*, DaLian, China. p. 594-601.

Smith, T., Waterman, M. (1981), "Identification of common molecular sub-sequences". *Journal of Molecular Biology*, v.147, 1981, p.195-197.

Sousa, M. S., Melo, A. C. M. A.: PackageBLAST: An Adaptive Multi-Policy Grid Service for Biological Sequence Comparison (2006). In: *ACM SAC*, Dijon, France, v. 1. p. 156-160.

Detecção precoce de epidemias: novas tecnologias

Maria Lucia F. Penna[1], Jair Koiller[2]

[1]Instituto de Medicina Social/UERJ
Rua São Francisco Xavier, 524, pavilhão João Lyra Filho, 7º andar, blocos D e E
Maracanã, Rio de Janeiro CEP 20550-900

[2]Centro de Matemática Aplicada – Fundação Getúlio Vargas/RJ
Praia de Botafogo 190 – 22250-040 – Rio de Janeiro – RR – Brazil

mlfpenna@terra.com.br jkoiller@fgv.br

Abstract. *In spite of good advances in epidemiological surveillance and information systems produced by the Brazilian Health Ministry, one still finds, upon retrospective analysis, surges or epidemics that may have been detected in early stages. It is of strategic interest for Brazil to absorb and take part on the development of new mathematical technique in the theme. In this note we register the interest of CMA/FGV to participate in such research networks both from the theoretical viewpoint and software applications. We also present and outline of the studies done in our seminars since 2008.*

Resumo. *Apesar de bons avanços em vigilância epidemiológica e sistemas de informação, realizados pelo Ministério da Saúde, ainda é frequente encontrar, quando da análise retrospectiva de dados secundários, surtos ou epidemias não registradas oportunamente. É interesse estratégico para o Brasi absorver técnicas matemáticas desenvolvidas recentemente ao nivel internacional.. Registramos nosso interesse em participar de redes de pesquisa, tanto do ponto de vista teórico, como no desenvolvimento de aplicativos. Sumarizamos também os estudos feitos em nosso seminário desde 2008.*

1. Sistemas de "early warning" para surtos epidêmicos.

Melhorar os sistemas de vigilância epidemiológica, em escala mundial, é uma necessidade premente. Em 2003 a Organização Mundial da Saúde (OMS) declarou alerta pandêmico global, com o episódio da SARS, felizmente controlado a tempo. Porém, no momento em que esta comunicação é escrita, infelizmente o risco de uma pandemia de gripe suína é manchete do noticiário em todos os países.

Duas novas estratégias, através de ferramentas de inteligência artificial já tem sido utilizadas internacionalmente. A detecção de surtos e epidemias via busca na Internet, e a detecção automática e em tempo real a partir das bases de dados da área de saúde.

No Brasil, os esforços recentes do Ministério da Saúde tem sido na direção de implantar sistemas de informação, por exemplo com a modernização do processo de entrada de dados, como o cartão do Sistema Único da Saude (SUS). Porém, o alto custo do desenvolvimento deste e de outros sistemas não se justifica, se toda a informação potencial contida nos dados não for inteligentemente analisada, para subsidiar as decisões da política de saúde.

2. Organismos responsáveis no Brasil e na França.

Centro de Informações Estratégicas em Vigilância em Saúde. **Reproduzido do portal do [Ministério da Saúde do Brasil, (2009)].** "O Centro de Informações Estratégicas em Vigilância em Saúde (Cievs) foi inaugurado em março de 2006. O Brasil é um dos cinco países do mundo a possuir uma sala especialmente equipada com os mais modernos recursos tecnológicos para receber informações sobre a ocorrência de surtos e emergências epidemiológicas que coloquem em risco a saúde da população em qualquer local do país... uma equipe especializada faz plantão, 24 horas por dia, todos os dias da semana, para receber notificações e comunicar as autoridades em caso de emergência"

Institut de Veille Sanitaire. **Reproduzido (sem tradução) do portal do [Ministère de la Santé et des Sports, France, 2009]** "Établissement public, placé sous la tutelle du ministère chargé de la Santé, l'Institut de veille sanitaire (InVS) réunit les missions de surveillance, de vigilance et d'alerte dans tous les domaines de la santé publique... l'InVS a vu ses missions complétées et renforcées par la loi du 9 août 2004 relative à la politique de santé publique, afin de répondre aux nouveaux défis révélés par les crises sanitaires récentes et les risques émergents".

3. Novas tecnologias

3.1 Oportunidades de colaboração.

Além dos novos métodos estatísticos udados em aprendizado de máquinas [Smale e Poggio, 2003]), mencionamos o aporte de geometria diferencial na redução dos dados ("manifold learning", [Verna, 2008]), e sistemas dinâmicos (bifurcações e atratores, [Guckenheimer, 2007]) na modelagem e análise qualitativa. As novas técnicas diferem da modelagem estatística tradicional, pela necessidade de robustez no caso de séries temporais curtas geradas por fontes de dados fora das tradicionais, como por exemplo, distribuição de medicação, e incorporar as anomalias como dados relevantes.

Detecção de surtos e epidemias através de pesquisa na Internet. O Ministério da Saúde do Canadá desenvolveu um sistema de busca (Global Public Health Intelligence Network [Mykhalovski, Weir, 2006]) de eventos na Internet em sete línguas: inglês, francês, espanhol, árabe, russo e chinês tradicional e simplificado. Os resultados das buscas, além de usados pelo governo canadense, são vendidos a diversos países e organismos internacionais. Relata-se que o sistema foi responsável pela detecção precoce de 40% das epidemias confirmadas pela OMS. A informação não estruturada é acessada, analisada e transformada, permitindo a geração de informação inteligente em tempo real. A tecnologia envolve a mineração de texto, envolvendo a formação de significados específicos (quantos, quando, onde, quem).

Detecção automática e em tempo real de surtos e epidemias a partir de base de dados: O governo dos USA, através de diversas agências financiadoras (NIH, NSF, DARPA) e a agência da Homeland Security (esta última mais preocupada com bioterrorismo) vem já desde algum tempo estimulando o desenvolvimento de novas técnicas matemáticas para uso na vigilância sindrômica. Algumas destas são descritas em [Banks e Castillo-Chavez, 2003]. As dificuldades são apontadas, por exemplo, em [Drake, 2005)].

3.2. Interesse do CMA/FGV.

Seguindo a tradição da FGV de contribuir no desenvolvimento de áreas de aplicação importantes para o Brasil, na área da saúde pretendemos inicialmente estudar as metodologias de detecção precoce de surtos epidêmicos. Nesta direção gostariamos de contar com colaboradores na rede Colibri/SBC. Foram feitos contatos preliminares junto ao CIEVS/MS, com excelente receptividade. Assim, sugerimos a SBC-Colibri a realização de um workshop sobre o assunto. A seguir relatamos os primeiros esforços no tema, realizados por nosso grupo, em andamento para ser submetidos a publicação.

4. Resultados obtidos por nosso grupo.

4.1.Simulador de surtos epidêmicos em redes - visualização geo-referenciada.

Para simular situações com heterogeneidades espaciais nas infectividades, divide-se uma região como a área metropolitana do Rio de Janeiro em regiões menores (Fig 1a). Em cada nó da rede, utilizamos um modelo SIR (suscetíveis, infectantes e recuperados) adaptado para arboviroses urbanas [Nishiura, 2006]. O subsistemas se acoplam através da matriz de circulação, que pode ser estimada com dados reais.

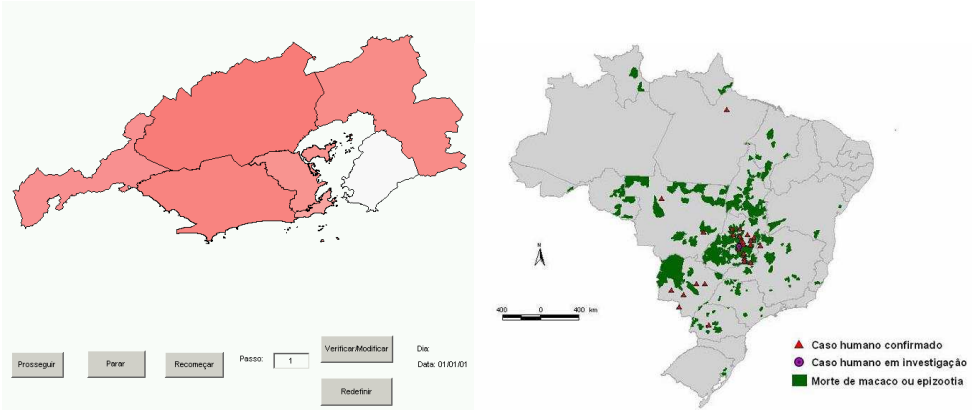


Figura 1. a. Simulação para dengue. b. Corredores ecológicos na febre amarela

O simulador foi desenvolvido por Paulo Cezar P. Carvalho, Moacyr Alvim e Aslá Sá. No momento o simulador está sendo calibrado com dados reais. Os resultados preliminares produziram um "fit" bastante razoável. Cenários para a introdução do Den4 e da febre amarela estão sendo avaliados. A ameaça da febre amarela urbana é claramente visível na Fig.1b.

4.2. Resultados teóricos.

A literatura sobre modelagem epidemiológica é vastíssima. O debate "sangrento" entre os especialistas começou com D'Alembert e Bernoulli. Assim, sem maiores pretensões, mencionamos duas observações obtidas em nossos seminários em 2008. A primeira é a existência de uma janela de oportunidade para intervenção, tipicamente de uma semana. Simulações sugerem que há um atraso no inicio do surto numa determinada região, ainda livre de surto epidêmico, quando regiões vizinhas tem suficiente grau de

comunicação e população infectada. Uma fórmula algébrica, até bastante simples, dá o sinal de alerta. Acreditamos que do ponto de vista teórico, a teoria dos "canards" pode prover uma explicação para o fenômeno [Benoît et al., 1981, Wechselberger, 2007].

A segunda observação é de Max Souza, nosso colaborador do Instituto de Matemática da UFF. Max propõe uma nova classificação metodológica de doenças infecciosas, baseada nas relativas escalas de tempo, das infecciosidades e recuperação [Souza, 2009]. Por exemplo, se o ciclo de vida de um vetor é rápido em relação aos demais parâmetros, o modelo pode ser simplificado, diminuindo a dimensionalidade do modelo. Por exemplo, a dinâmica do mosquito *Aedes* é substituida por uma "virtual transmissão direta" entre humanos. A técnica matemáticas aqui utilizada é a analise multi-escala, obtendo-se fórmulas para novos parâmetros agregados.

Referências

Ministerio da Saúde do Brasil (2009), Portal,
http://portal.saude.gov.br/portal/saude/visualizar_texto.cfm?idtxt=22233
http://portal.saude.gov.br/portal/saude/area.cfm?id_area=962

Ministère de la Santé et des Sports, France, (2009), France http://www.invs.sante.fr/

Mykhalovski, E, Weir, L.(2006), The global public health intelligence network and early warning outbreak detection, Canadian J. Public Health, 97:1, 43-44, ver também
http://www.phac-aspc.gc.ca/media/nr-rp/2004/2004_gphin-rmispbk-eng.php
http://www.who.int/csr/alertresponse/en/

Drake, J.M. (2005), "Fundamental Limits to the Precision of Early Warning Systems for Epidemics of Infectious Diseases", PLoS Med. 2005 May; 2(5): e144. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1140954

Poggio, T., S. Smale (2003), "The Mathematics of Learning: Dealing with Data", Notices of the AMS 50, no. 5, 537-544. Ver também: Poggio et al. (2008) http://www.mit.edu/~9.520/spring08/index.html "Statistical Learning Theory"

Banks H.T., Castillo-Chavez, C., Editors (2003) Bioterrorism: Mathematical Modeling Applications in Homeland Security, SIAM Frontiers in Applied Mathematics 28.

Verna, N. (2008) "Mathematical advances in manifold learning", preprint http://www.cse.ucsd.edu/~naverma/papers/re_manifold.pdf

Guckenheimer, J. (2007) Scholarpedia, 2(6):1517
http://www.scholarpedia.org/article/Bifurcation

Nishiura, H. (2006) "Mathematical and Statistical Analyses of the spread of dengue", Dengue Bulletin 30, 51-67.

Benoît, E., Callot, J. F., Diener F., Diener, M. [1981]; "Chasse au canard"; Collectanea Mathematica, 31-32 (1-3), 37-119. Ver: http://www.scholarpedia.org/article/Canards

Wechselberger, M. (2007), Scholarpedia, 2(4):1356.
http://www.scholarpedia.org/article/Canards

Souza, M (2009) "Global Analysis and Mutliscaling Asymptotics for a Vector-Bourne Epidemic Model", em *Mathematical Methods and Modeling of Biophysical Phenomena*, IMPA. http://www.impa.br/opencms/pt/eventos/store_old/evento_0901

V

Network and Mobile Computing

COLIBRI 2009 • 22-23 de julho de 2009 • ISBN 978-85-7669-245-4

Adding Dynamicity to the Uncertainty that Characterizes Distributed Systems: Challenges Ahead

Raimundo Macêdo[1] **and Michel Raynal**[2]

[1]Distributed Systems Laboratory (LaSiD/DCC)
Federal University of Bahia
Campus de Ondina – Salvador – BA – Brazil

[2]IRISA/INRIA
University of Rennes
Campus de Beulieu – Rennes – France

macedo@ufba.br, raynal@irisa.fr

Abstract. *The uncertainties inherent to distributed systems, such as unpredictable message delays and process failures, gave place to a great research effort in the past, where numerous solutions to fault-tolerance mechanisms have been proposed with a variety of guarantees and underlying system assumptions. The advent of new classes of distributed system applications (such as social networks, security, smart objects sharing etc) and technologies (VANET, WiMax, Airborn Networks, DoD Global Information Grid, P2P) are radically changing the way in which distributed systems are perceived. Such emerging systems have a composition, in terms of processes participating to the system, that is self-defined at run time depending, for example, on their will to belong to such a system, on the geographical distribution of processes etc. In this paper, we point to some of the challenges that have to be addressed by fault-tolerance solutions in the light of such a new dynamicity dimension, and that can motivate future collaborative work.*

1. Introduction

A distributed system is usually characterized by a set $\Pi = \{p_1, p_2, ..., p_n\}$ of processes sited on possibly distinct networked computers and a set $\chi = \{c_1, c_2, ..., c_m\}$ of communication channels. Networked computers form arbitrary network topologies and processes communicate by using a communication protocol that implements process-to-process communication. Such process-to-process communication defines communication channels that may include several intermediate network level communication links. Therefore, a communication channel c_i connecting processes p_i and p_j defines a "is able to communicate"relation between p_i and p_j, rather than a network level link connecting the machines that host p_i and p_j. The actual behavior or properties observed in the distributed system, which defines a distributed system model, results from the way processes and channels are implemented by the underlying operating and communication systems. Whether processes and channels can operate within certain known time bounds, and how and when processes and channels can fail, are the usual uncertainties that designers have to face when solving application problems in distinct models. That is why being able to solve such problems despite failures (i.e., being Fault-Tolerant) has been considered a major research challenge. In this context, the abstractions of distributed consensus

and concurrent registers are important building blocks from which fault-tolerance mechanisms such as state machine replication can be constructed. Basically, a consensus protocol guarantees that distributed processes can unanimously agree on a proposed output value, whereas concurrent registers provide distributed processes with certain guarantees on read/write operations. However, the possibility of solving these problems depends on the uncertainty level assumed in a given system model. For instance, let us consider the synchronous (or time-based) and asynchronous (or time-free) models. Under process crashes and reliable channels, the reliable multicast problem is solvable in both models [Lynch 1996, Hadzilacos and Toueg 1993], whereas distributed consensus can be solved in the synchronous model, but not in the asynchronous model [Fisher et al. 1985].

The advent of new classes of distributed system applications (such as social networks, security, smart objects sharing etc) and technologies (VANET, WiMax, Airborn Networks, DoD Global Information Grid, P2P) are radically changing the way in which distributed systems are perceived, adding even more uncertainties that lead to more dynamic distributed system models.

Such emerging systems have a composition, in terms of processes participating to the system, that is self-defined at run time depending, for example, on their will to belong to such a system, on the geographical distribution of processes etc. Therefore, a common denominator of such emerging systems is the dynamicity dimension related to the processes that actually make the system at a given time, which introduces a new source of unpredictability inside a distributed system. Such a dynamicity reflects also on the available system resources that dynamically changes following system compositions. This in turn requires applications to be adaptive, for instance, to less network bandwidth or degraded Quality-of-Service (QoS). Ideally, in these highly dynamic scenarios, adaptiveness characteristics of applications should be self-managing or autonomic.

In this paper, we point to some of the challenges that need to be addressed by distributed consensus and concurrent register solutions in the light of such a new dynamicity dimension, and should motivate future collaborative work of the Franco-Brazilian community working in distributed computing. The reminder of this extended abstract is structured as follows. In section 2 is presented related work with emphasis on our previous collaborative work on this field. Challenges that should be of primary importance for the Franco-Brazilian research community interested in distributed computing are presented in section 3.

2. Related and Previous Work

Among the faul-tolerant problems, distributed consensus has received a great deal of attention because it can be used as a basic building block to solve several class of problems that includes group membership, atomic commitment, atomic broadcast, among others. Motivated by the consensus impossibility result in asynchronous systems, researchers have proposed a number of partially synchronous distributed system models that introduce different levels of synchrony into the asynchronous system, where the consensus problem is solvable [Dwork et al. 1988, Dolev et al. 1987, Cristian and Fetzer 1999, Lamport 1998, Chandra and Toueg 1996]. Among them, the failure detectors mechanism proposed by Chandra and Toueg[Chandra and Toueg 1996] has received special attention because of its simplicity, encapsulating the synchrony needed to achieve consensus by

defining axiomatic properties associated with different classes of failures detectors.

In [Hurfin et al. 1999] we have defined a general framework that simplifies the system designer work. Such framework can be used to solve a variety of classes of distributed agreement problems based on failure detectors. Following the same paradigm of failure detectors, we defined consensus protocols based on a decentralized communication pattern and observed that a tradeoff has to be found between the number of communication steps and the number of exchanged messages, by automatically switching between a centralized and a decentralized communication pattern [Greve et al. 2000]. The notion of round is at the core of failure detector based Protocols. To be more tolerant with regard to the message transfer delays, we introduced the notion of cycle of K rounds [Hurfin et al. 2001].

In [Gorender et al. 2007] we have explored the notion of a hybrid model to solve the uniform consensus problem, where we assume that the underlying system is capable of providing distinct QoS communication guarantees. Because timely and untimely channels may exist in parallel, the underlying system model can be hybrid in space (in this sense, similar to TCB [Veríssimo and Casimiro 2002]), but, differently from TCB, the nature of such hybridism does not require that all processes are interconnected by timely channels (which would characterize a synchronous wormhole).

In recent work, we have investigated the implementation of regular registers in asynchronous dynamic message-passing systems, where the dynamicity rate (i.e., frequency of joins and leaves of the processes) is constant [Baldoni et al. 2009]. In another recent work, we focused on the support for self-management behavior, that allows an application to dynamically adapt to a given system configuration [Andrade and de Araújo Macêdo 2009]. These research efforts represent a preliminary endeavor towards models and mechanisms that fulfill the requirements of modern dynamic distributed systems.

3. New Challenges

The new classes of highly dynamic distributed system applications and technologies pose a tremendous challenge to fault-tolerant systems. The main difficulty stems from the fact that in distributed systems the implementation of fault-tolerance mechanisms (such as replication) depends on coordinated actions from the system processes. As the system composition changes dynamically, it is difficult to assure that the necessary resources will be kept long enough in order to guarantee the prescribed fault-tolerance properties (for instance, that redundancy level is kept satisfactory).

So, the main challenges consists in defining appropriate *abstractions suited to dynamic systems*, and related self-management mechanisms. Those abstractions should be *locality-based* and should take into account notions such a *presence detector* (similar to failure detectors). From a computational point of view, an insight into population protocols [Angluin et al. 2007] should be of primary importance and could be the starting point of promising research inside the realm of distributed computing.

Referências

Andrade, S. S. and de Araújo Macêdo, R. J. (2009). A non-intrusive component-based approach for deploying unanticipated self-management behaviour. In *Proceedings of*

IEEE ICSE 2009 Workshop Software Engineering for Adaptive and Self-Managing Systems.

Angluin, D., Aspnes, J., Eisenstat, D., and RuppertDwork, E. (2007). The computational power of population protocols. *Distributed Computing*, 20(4):279–304.

Baldoni, R., Bonomi, S., Kermarrec, A.-M., and Raynal, M. (2009). Implementing a register in a dynamic distributed system. In *International Conference on Distributed Computing Systems (ICDCS2́009)*.

Chandra, T. D. and Toueg, S. (1996). Unreliable failure detectors for reliable distributed systems. *Journal of the ACM*, 43(2):225–267.

Cristian, F. and Fetzer, C. (1999). The timed asynchronous distributed system model. *IEEE Transactions on Parallel and Distributed Systems*, 10(6):642–657.

Dolev, D., Dwork, C., and Stockmeyer, L. (1987). On the minimal synchronism needed for distributed consensus. *Journal of the ACM*, 34(1):77–97.

Dwork, C., Lynch, N., and Stockmeyer, L. (1988). Consensus in the presence of partial synchrony. *Journal of the ACM*, 35(2):288–323.

Fisher, M. J., Lynch, N., and Paterson, M. S. (1985). Impossibility of distributed consensus with one faulty process. *Journal of the ACM*, 32(2):374–382.

Gorender, S., de Araújo Macêdo, R., and Raynal, M. (2007). An Adaptive Programming Model for Fault-Tolerant Distributed Computing. *IEEE Transactions on Dependable and Secure Computing*, pages 18–31.

Greve, F., Hurfin, M., Macêdo, R., and Raynal, M. (2000). Consensus based on strong failure detectors : A time and message-efficient protocol. In *IEEE International Workshop on Fault-Tolerant Parallel and Distributed Systems. Lecture Notes in Computer Science 1800. Springer Verlag,pg 1258-1267.*

Hadzilacos, V. and Toueg, S. (1993). *Distributed Systems (Edited by Sape Mullender). Chapter 5. Fault-Tolerant Broadcasts and Related Problems*. Addison-Wesley Pub Co.

Hurfin, M., Macêdo, R., Mostefaoui, A., and Raynal, M. (2001). A consensus protocol based on a weak failure detector and a sliding round window. In *20th IEEE Int. Symposium on Reliable Distributed Systems (SRDS0́1).*

Hurfin, M., Macêdo, R., Raynal, M., and Tronel, F. (1999). A general framework to solve agreement problems. In *18th IEEE Symposium on Reliable Distributed Systems (SRDS9́9).*

Lamport, L. (1998). The part time parliament. *ACM Trans. on Computer Systems*, 16(2):133–169.

Lynch, N. A. (1996). *Distributed Algorithms*. Morgan Kaufmann Publishers, Inc.

Veríssimo, P. and Casimiro, A. (2002). The timely computing base model and architecture. *IEEE Transactions on Computers*, 51(8):916–930.

Systèmes collaboratifs à base d'agents :

Une expérience de collaboration entre LIP6 et PUC-Rio

Jean-Pierre Briot[1], Carlos Lucena[2], Karin Breitman[2] et al.

[1]Laboratoire d'Informatique de Paris 6 (LIP6), Université Pierre et Marie Curie – CNRS, Paris – France

[21]Departamento de Informática – Pontifícia Universidade Católica (PUC-Rio), Rio de Janeiro, RJ – Brésil

jean-pierre.briot@lip6.fr, lucena@inf.puc-rio.br, karin@inf.puc-rio.br

Résumé. *Nous présentons dans ce court papier notre expérience de collaboration de recherche débutée en 2005 entre le LIP6 et PUC-Rio, dans le domaine de l'informatique collaborative à base d'agents. Nous montrons comment, à partir d'une collaboration initiale sur la question de la tolérance aux fautes des systèmes multi-agents, cette collaboration s'est étendue vers des domaines tels que : gouvernance des systèmes multi-agents ouverts, systèmes sensibles au contexte, réseaux, et gestion participative de l'environnement. Ceci est également l'occasion d'analyser notre expérience en matière de programmes de collaboration (tels que : thèses co-dirigées, CAPES-COFECUB, ANR-FINEP).*

Abstract. *In this short paper we present our experience in computer science research collaboration from 2005 between LIP6 and PUC-Rio, about agent-based collaborative systems. We show how, from an initial collaboration on the issue of fault-tolerant multi-agent systems, the collaboration expanded to domains such as: governance of open multi-agent systems, context-aware systems, networks, and participatory management of the environment. This is also the opportunity to analyse our experience in terms of collaboration programs (such as: co-directed theses, CAPES-COFECUB, ANR-FINEP).*

1. Introduction et historique

Notre objectif est de présenter notre expérience de collaboration entre l'équipe Systèmes Multi-Agents (SMA) du Laboratoire d'Informatique de Paris 6 (LIP6), Paris, et le *Laboratório de Engenharia de Software (LES)* du *Departamento de Informática* de la *Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)*. Notre collaboration a véritablement débuté en 2005 par un projet CAPES-COFECUB, terminé en décembre 2008. Mais le premier contact est intervenu à l'occasion de notre participation [Guessoum *et al*. 2003] à la première occurrence d'une série de workshops sur Software Engineering for Large Scale Multi-Agent Systems (SELMAS), organisée par Carlos Lucena et son équipe [Garcia *et al*. 2003]. Nous y avions présenté des travaux sur la tolérance aux fautes de systèmes multi-agents (travaux qui ont par ailleurs donné lieu au projet FACOMA du programme ANR SetIn, incluant LIP6, INRIA – projet REGAL LIP6/INRIA – et LIRMM) [Guessoum *et al*. 2010].

Ce premier contact entre nos deux équipes a permis d'identifier une proximité de culture et d'objectifs scientifiques (au croisement entre programmation, génie logiciel, intelligence artificielle et informatique répartie [Briot 2005]) et a été le déclencheur d'une collaboration maintenant bien établie (entre autres résultats : une quarantaine de publications co-signées en 4 ans, 2 doctorats co-dirigés, 2 masters co-dirigés, un projet CAPES-COFECUB, un projet ARCUS, un projet CNPq et un projet ANR-FINEP). Ce

court texte est ainsi l'occasion de présenter cette expérience de collaboration, la manière dont elle s'est progressivement élargie à différents domaines et équipes et une réflexion sur les programmes de soutien à des collaborations.

Il nous faut mentionner qu'il existait déjà une tradition de collaboration du LIP6 avec le Brésil. Le LIP6 inclut notamment actuellement 3 chercheurs permanents de nationalité brésilienne (2 maîtres de conférences à Paris 6 et un chargé de recherche au CNRS, dans les équipes réseaux et systèmes répartis). De plus, il y a en moyenne entre 10 et 20 doctorants brésiliens en thèse au LIP6. Parmi les collaborations existantes, on peut souligner la collaboration historique entre Guy Pujolle et Serge Fdida et Otto von Duarte de l'UFRJ (Rio) dans le domaine des réseaux. Une autre collaboration s'est créée entre Jean-Gabriel Ganascia et Vincent Corruble et Geber Ramalho de l'UFPE (Recife) dans le domaine de l'apprentissage automatique puis des jeux [Corruble et Ramalho 2009]. Ce deuxième canal de collaboration inclut d'ailleurs également une composante musique informatique, auquel nous avons également participé [Cabral *et al*. 2005]. On peut enfin également mentionner un canal de collaboration entre le projet REGAL et l'UFBA (Salvador) dans le domaine des systèmes répartis ainsi que des canaux de collaboration entre l'équipe Décision et l'UFPE ainsi qu'avec l'UFRJ dans le domaine de l'aide à la décision.

Notre collaboration, nouveau canal LIP6-Brésil depuis 2005, s'est d'ailleurs amplifiée et s'est interconnectée avec plusieurs de ces collaborations historiques, puisque nous avons récemment coordonné la composante informatique d'un projet ARCUS (2006-2008) porté par Paris 6 et l'Ile de France avec PUC-Rio et UFPE et portant sur les systèmes collaboratifs à base d'agents tels que jeux vidéo et jeux sérieux. Nous participons également avec plusieurs autres partenaires au tout récent projet HORIZON du programme ANR-FINEP réseaux du futur, sur l'utilisation d'une approche agent pour le pilotage de réseaux virtualisés.

2. Thématiques abordées

2.1. Systèmes multi-agents tolérants aux fautes

Cette facette initiale de notre collaboration a porté sur des mécanismes de fiabilisation (tolérance aux fautes) de systèmes multi-agents à large échelle, par des techniques de réplication adaptative [Guessoum *et al*. 2003], développés en collaboration avec le projet REGAL de l'INRIA/LIP6. Le contrôle de la réplication (quels agents, quelle politique de réplication, combien de réplicas…) est pris en charge par une couche de contrôle en fonction d'une estimation de la criticité des agents à partir de différents types d'informations [Briot *et al*. 2006] : informations statistiques (messages échangés, charge des processeurs…) et sémantiques (rôles [Guessoum *et al*. 2010], plans [Almeida *et al*. 08], normes [Gatti *et al*. 2007]).

2.2. Gouvernance de systèmes multi-agents ouverts

Cette deuxième facette a porté sur la gouvernance de systèmes multi-agents ouverts. L'idée est, à partir d'une ingénierie des besoins d'une application, de spécifier différents types de lignes de conduite et de régulation, sous la forme de normes (interdictions, obligations, permissions), et qui pourront être utilisées aussi bien par les agents dans leurs processus de décision interne [Felicíssimo *et al*. 2007], que par des mécanismes de contrôle – cette fois externes – de leurs actions et communication [Paes *et al*. 2007] [Carvalho *et al*. 2007]. La représentation et la structuration de normes de gouvernance repose sur l'utilisation d'ontologies, des règles permettant également d'exprimer de manière déclarative et dynamique leur utilisation et composition [Felicíssimo *et al*. 2007]. Des travaux sur la génération de monitoring de systèmes multi-agents, à partir de normes exprimées dans une logique déontique, ont fourni un pont naturel entre spécification de

normes et architectures de contrôle [Felicíssimo *et al*. 08] via la collaboration entre 2 thèses menées à PUC-Rio (Carolina Felicíssimo) et au LIP6 (Caroline Chopinaud).

2.3. Systèmes collaboratifs sensibles au contexte

Une extension naturelle de ce travail sur la gouvernance a alors commencé avec Markus Endler et Karin Breitman (également du département d'informatique de la PUC-Rio) sur l'utilisation de mécanismes de manipulation dynamique de normes et de gouvernance [Viterbo *et al*. 2008] pour contrôler l'adaptation au contexte (« context aware ») d'applications collaboratives nomades [Viterbo *et al*. 2009]. L'objectif est d'explorer la représentation d'informations (sur le contexte) et de normes (contraintes sur les services offerts) sous la forme d'ontologies, ainsi que des mécanismes les manipulant (alignement, sélection…) pour adapter et contrôler les services offerts. Une soumission d'un projet (CAMPUS) sur cette thématique, avec des partenaires du Chili et du Pérou, a été faite cette année au programme STICAmSud.

2.4. Pilotage de réseaux

Le projet HORIZON représente l'exploration d'approches et mécanismes multi-agents pour la conception de réseaux du futur (programme ANR), plus dynamiques et autonomes. Le projet explore un pilotage autonomique décentralisé et collaboratif de la gestion de la virtualisation des réseaux (par exemple, pour assurer la création ou/et la clôture dynamique de réseaux, ou la gestion du nomadisme dans des réseaux WiFi). Elle représente une collaboration interdisciplinaire entre LIP6 (équipes réseaux et agents), PUC-Rio (agents et génie logiciel) et UFRJ (réseaux), ainsi que plusieurs autres partenaires (UNICAMP, Télécom Sud Paris…), et ainsi élargit notre palette d'étude de mécanismes multi-agents pour la conception et le contrôle d'applications collaboratives décentralisées.

2.5. Gestion environnementale assistée par ordinateur

La retombée la plus interdisciplinaire de notre collaboration porte sur le développement de méthodologies collaboratives informatisées pour la gestion de l'environnement (en particulier des espaces protégés tels que des parcs). L'approche se base sur des jeux de rôles informatisés distribués (jeux sérieux) [Briot *et al*. 2008], une aide à la négociation [Vasconcelos *et al*. 2009], et à la décision, et l'insertion d'agents artificiels joueurs et décideurs [Briot *et al*. 2009]. Ce projet, nommé SimParc, basé sur une collaboration LIP6, PUC-Rio et UFRJ (gestion environnementale) est progressivement monté en puissance et comprend actuellement une dizaine de chercheurs, dont 4 doctorants. Un workshop bilatéral sur ce thème a également été organisé en novembre 2007 à Rio de Janeiro et a réuni une vingtaine de participants de différents organismes (CIRAD, USP, UNIFOR…). Le projet SimParc a reçu le soutien du programme ARCUS, qui a permis son émergence en tant que projet bilatéral et interdisciplinaire. SimParc reçoit actuellement le soutien du programme Grandes Desafios du MCT-CNPq-INFO. Un projet issu de SimParc, incluant également le LISC du CEMAGREF, et qui explore la modélisation de la viabilité des décisions de gestion à partir de la théorie de la viabilité, a été récemment accepté dans le cadre du programme Ingénierie Ecologique du CNRS-Cemagref (projet ViabilitéSimParc). Un projet plus ample (avec également participation de l'équipe Décision du LIP6 et d'un Parc National) a également été récemment soumis au programme ANR-AIRD SysTerra.

3. Bilan et conclusion

3.1. Bilan

Les résultats à ce jour de cette collaboration débutée en 2005 incluent notamment :

- une quarantaine de publications co-signées (dont certaines sont référencées ici),

- 2 thèses de doctorat co-dirigées (Carolina Felicíssimo et Eurico Vasconcelos),
- 2 thèses de master co-dirigées (Maíra Gatti et Cynthia Moisés),
- un certain nombre de programmes de coopération : CAPES-COFECUB, ARCUS, et ANR-FINEP.
- Plusieurs nouveaux projets ont également été récemment soumis à des programmes de coopération : STICAmSud, PICS CNRS et SysTerra-AIRD.

La collaboration a également eu un effet d'entraînement sur :

- plusieurs autres thèses de doctorat menées entre France et Brésil, bien que non formellement co-dirigées (par exemple, Alessandro Almeida sur la tolérance aux fautes de systèmes multi-agents à l'aide de plans [Almeida *et al*. 08]),
- plusieurs autres séjours sandwich (par exemple, Bruno Silvestre sur la programmation concurrente [Silvestre *et al*. 20xx]),
- des collaborations avec plusieurs autres Professeurs du département d'informatique de PUC-Rio (outre ceux déjà cités), tels que : Arndt von Staa, Noemi Rodriguez [Silvestre *et al*. 20xx], Renato Cerqueira, Simone Barbosa [Vasconcelos *et al*. 2009] et Marco Casanova.

3.2. Analyse

De notre expérience à ce jour, après environ 4 ans de collaboration, nous pouvons tirer quelques premières constatations :

- Les programmes de type CAPES-COFECUB sont un excellent soutien à des initiatives de collaboration. Cependant, à la fin d'un tel projet, la question se pose de comment continuer et amplifier la collaboration (si elle en a le potentiel).
- De plus, ce type de programme de collaboration se borne (ce qui est déjà très utile) à des missions d'échanges, mais ne permet pas de financer les recherches dans chaque équipe/pays. Il faut donc coupler financement local et bilatéral, ce qui n'est pas toujours aisé du fait des contraintes (durée, etc.) propres à ces 2 types de financements. Les initiatives très récentes de couplage entre certains programmes de l'ANR et un organisme brésilien tel que la FINEP (idem pour des programmes européens étendus) permettent de combiner les deux types de financements dans un même cadre, avec un niveau de financement conséquent, ceci offrant ainsi des opportunités nouvelles très intéressantes.
- Le projet SimParc (voir le paragraphe 2.5) a représenté une nouvelle forme de collaboration, à la fois bilatérale et interdisciplinaire. Elle a été rendue possible par le programme ARCUS, qui inclut plusieurs axes disciplinaires, ce qui n'est pas le cas de la plupart des programmes bilatéraux, soit mono disciplinaires (MathAmSud, STICAmSud…), soit pluridisciplinaires (CAPES-COFECUB, CNRS-CNPq…) mais dans lesquels les projets sont en pratique essentiellement mono-disciplinaires (ne serait ce que du fait du montant des financements). La conjonction de la juxtaposition de plusieurs disciplines à l'intérieur du programme ARCUS, le montant plus conséquent du financement et la durée (3 ans), ont permis la création d'un projet interdisciplinaire original, ce qui a d'ailleurs représenté une première pour le programme ARCUS.
- Les programmes de type co-tutelle et collège doctoral franco-brésiliens sont très intéressants, mais en pratique nous avons rencontré des difficultés à les utiliser, du fait d'un relatif manque de flexibilité : la convention doit être signée avant le début de la thèse, ce qui élimine ainsi le cas d'émergence de co-directions en cours de thèse, pourtant des cas réels en pratique. De plus, les contraintes de durée de séjour du sandwich du collège doctoral peuvent parfois entrer en conflit avec certaines

règles d'écoles doctorales (contraignant les durées de séjours à l'étranger en fonction du calendrier de qualification et de défense de proposition de thèse).

Remerciements

À nos collègues et étudiants qui, par leur participation, ont permis la genèse et le développement réussi de cette collaboration.

Références

Almeida, A., Aknine, S. et Briot, J.-P. (2008) "Dynamic resource allocation heuristics for providing fault tolerance in multiagent systems", *23rd Annual ACM Symposium on Applied Computing (SAC'08)*, Fortaleza, CE, Brésil, mars, p. 66–70.

Briot, J.-P., Sordoni, A., Vasconcelos, E., Melo, G., de Azevedo Irving, M. et Alvarez, I. (2009) "Design of a decision maker agent for a distributed role playing *game* – Experience of the SimParc project", *AAMAS Workshop on Agents for Games and Simulations (AGS'09)*, édité par F. Dignum, J. Bradshaw, B. Silverman et W. van Doesburg, Budapest, Hongrie, mai, p. 16–30.

Briot, J.-P., Vasconcelos, E., Adamatti, D., Sebba Patto, V., Irving, M., Barbosa, S., Furtado, V. et Lucena, C. (2008) "A computer-based support for participatory management of protected areas: The SimParc project", *XXVIII Congresso da Sociedade Brasileira de Computação (SBC'08) – Seminário Integrado de Software e Hardware "Grandes Desafios"*, Belem, PA, Brésil, juillet, p. 1–15.

Briot, J.-P., Guessoum, Z., Aknine, S., Almeida, A., Faci, N., Gatti, M., Lucena, C., Malenfant, J., Marin, O. et Sens, P. (2006) "Experience and prospects for various control strategies for self-replicating multi-agent systems", *International Workshop on Software Adaptive and Self-Managing Systems (SEAMS'2006)*, Shanghaï, Chine, mai, ACM, p. 37–43.

Briot, J.-P. (2005) "Foreword", *Software Engineering for Large-Scale Multi-Agent Systems III*, édité par R. Choren, A. Garcia, C. Lucena et A. Romanovsky, No 3390, Lecture Notes in Computer Science, Springer Verlag, p. v–vi.

Cabral, G., Pachet, F. et Briot, J.-P. (2005) "Automatic x traditional descriptor extraction: The case of chord recognition", *6th International Conference on Music Information Retrieval (ISMIR'2005)*, Londres, Royaume-Uni, septembre, p. 444–449.

Carvalho, G., Lucena, C., Paes, R., Choren et Briot, J.-P. (2007) "Applying the governance framework technique to promote maintainability in open multi-agent systems", *Agent-Oriented Software Engineering: 7th International Workshop, AOSE 2006. Revised Papers*, édité par L. Padgham et F. Zambonelli, No 4405, Lecture Notes in Computer Science, Springer Verlag, p. 64–83.

Corruble, V. et Ramalho, G. (2009) "Jeux vidéo et systèmes multi-agents", *Technologies des systèmes multi-agents et applications industrielles*, édité par A. El Fallah-Seghrouchni et J.-P. Briot, Collection IC2, Hermes/Lavoisier, p. 235–264.

Felicíssimo, C., Chopinaud, C., Briot, J.-P., El Fallah-Seghrouchni, A. et Lucena, C. (2008) "Contextualizing normative open multi-agent systems", *23rd Annual ACM Symposium on Applied Computing (SAC'08)*, Fortaleza, CE, Brésil, mars, p. 52–59.

Felicíssimo, C., Lucena, C., Briot, J.-P. et Choren, R. (2007) "Informing regulatory dynamics in open MASs", *Coordination, Organizations, Institutions and Norms in Multi-Agent Systems II*, édité par P. Noriega, J. Vázquez-Salceda, G. Boella, O. Boissier, V. Dignum, N. Formara et E. Matson, No 4386, Lecture Notes in Artificial Intelligence, Springer Verlag, p. 140–155.

Garcia, A., Lucena, C., Zambonelli, F., Omicini, A. et Castro, J. (éditeurs) (2003) "Software Engineering for Large-Scale Multi-Agent Systems", No 2603, Lecture Notes in Computer Science, Springer Verlag.

Gatti, M., Carvalho, G., Paes, R., Lucena, C. et Briot, J.-P. (2007) "On fault tolerance in law-governed multi-agent systems", *Software Engineering for Large-Scale Multi-Agent Systems V*, édité par R. Choren, A. Garcia, H. Giese, H. Leung, A. Romanovsky et C. Lucena, No 4408, Lecture Notes in Computer Science, Springer Verlag, p. 1–20.

Guessoum, Z., Briot, J.-P., Faci, N., et Marin, O. (2010) "Towards reliable multi-agent systems - An adaptive replication mechanism", *Multiagent and Grid Systems (MAGS)*, 6(1), mars, p. 134, à paraître.

Guessoum, Z., Briot, J.-P., Marin, O., Hamel, A. et Sens, P. (2003) "Dynamic and adaptive replication for large-scale reliable multi-agent systems", *Software Engineering for Large-Scale Multi-Agent Systems*, édité par A. Garcia, C. Lucena, F. Zambonelli, A. Omicini et J. Castro, No 2603, Lecture Notes in Computer Science, Springer Verlag, p. 182–198.

Paes, R., Carvalho, G., Gatti, M., Lucena, C., Briot, J.-P. et Choren, R. (2007) "Enhancing the environment with a law-governed service for monitoring and enforcing behavior in open multi-agent systems", *Environments for Multi-Agent Systems III*, édité par D. Weyns, H. Van Dyke Parunak et F. Michel, No 4389, Lecture Notes in Computer Science, Springer Verlag, p. 221–238.

Silvestre, B., Rossetto, S., Rodriguez, N. et Briot, J.-P. (20xx) "Flexibility and coordination in event-based, loosely-coupled, distributed systems", *Computer Languages, Systems & Structures*, Elsevier, accepté pour publication/à paraître.

Vasconcelos, V., Briot, J.-P., Irving, M., Barbosa, S. et Furtado, V. (2009) "A user interface to support dialogue and negotiation in participatory simulations", *Multi-Agent-Based Simulation IX,* édité par N. David et J. Sichman, No 5269, Lecture Notes in Artificial Intelligence, Springer Verlag, p. 127–140.

Viterbo, J., Mazuel, L., Charif, Y., Endler, M., Sabouret, N., Breitman, K., El Fallah Seghrouchni, A. et Briot, J.-P. (2009) "Ambient intelligence: Management of distributed and heterogeneous context knowledge", *Context-Aware Self Managing Systems*, édité par W. Dargie, CRC Studies in Informatics Series, Chapman & Hall, p. 79–128.

Viterbo, J. F., Endler, M. et Briot, J.-P. (2008) "Ubiquitous service regulation based on dynamic rules", *13th IEEE International Conference on Engineering of Complex Computer Systems (ICECCS'2008)*, Belfast, Irlande du Nord, mars–avril, p. 175–182.

Towards a distributed computing model that characterizes dynamics of mobile networks

Luciana Arantes[1], **Alfredo Goldman**[2], **Pierre Sens**[1]

[1]Projet Regal – INRIA/LIP6 (Université Pierre et Marie Curie, CNRS)
104, av. du President Kennedy, 75016, Paris, France

[2]Instituto de Matematica e Estatistica - Universidade de São Paulo (IME)-USP
Rua do Matão 1010, CEP. 05508-090, São Paulo, SP, Brazil.

{luciana.arantes,pierre.sens}@lip6.fr,gold@ime.usp.br

Abstract. *Our main goal in this proposal is to present the research possibilities opened by two previous cooperations among groups from France and Brazil in dynamic systems.*
Basically, we intend to add up the expertise of the French group on distributed systems with the skills of the Brazilian group on the practical use of theoretical tools in order to create a new distributed model for mobile networks (MANET). Despite the fact that both groups have conducted their researches independently, we believe that to join efforts to define such a model would be very promising and would open an interesting field of research.

1. Teams Presentation

Regal is a joint research group of **LIP6, Paris6** and **INRIA-Rocquencourt** (`http://regal.lip6.fr/`) that investigates systems or distributed algorithms for large-scale, hierarchical or self-organized environments such as peer-to-peer, Grid platforms, and mobile or sensor networks. It aims at providing solutions for tolerating failures and/or reducing the time for getting an information or data. One of the fields of our research is to propose new distributed, adaptive and/or fault tolerant algorithms that cope with the dynamics, scalability or communication latency heterogeneity of such environments. Regal research has both a theoretical and an experimental approaches since we prove the correctness of the algorithms and we usually develop prototypes on top of simulators, cluster-based emulated platforms or testbed platforms like Grid5000 [gri] in order to perform evaluation experiments.

The **Distributed Systems Research** group of **IME-USP** (`http://gsd.ime.usp.br/`) gathers several professors and students interested in aspects of parallel, distributed, and mobile computing. Currently, the main research areas of the group are related to Grid Computing and the middleware InteGrade (`http://www.integrade.org.br`), mobile computing, and networks with somehow predictable behavior. The ongoing research on the group aims at producing software and experimental results based on a strong theoretical basis.

2. Brazilian-French Cooperations

Regal (INRIA/LIP6): A CAPES/COFECUB (projet 497/05) cooperation project with the Federal University of Bahia (UFBA), the Federal University of Campina Grande

(UFCG), and several French laboratories (ADEPT Team, GRAND LARGE Team, REGAL Team and ENST Bretagne) took place during the period of 2005-2008. The focus of the project was distributed/Grid computing systems and algorithms. In this projet, we have particularly worked with Prof. Fabiola Greve of UFBA (Distributed Computing Group GAUDI) by proposing an algorithm that implements an unreliable failure detector for mobile wireless network (MANET) [Sens et al. 2008a][Sens et al. 2008b]. Informally, an unreliable failure detector can be seen as a per process oracle, which periodically provides a list of processes suspected of having crashed. We have considered a distributed system where nodes connected by a wireless network could move, did not have a full knowledge of the network, and could only communicate to their respective neighbors. However, some specific properties were added to our model in order to satisfy the properties of the unreliable failure detector themselves. Furthermore, links between nodes were bi-directional, and we assume that, in spite of changes in the topology, there is always a path between two correct nodes. That is, there are no network partitions in the system in spite of node failures and mobility.

Distributed Systems Research (IME-USP): There were several well succeeded previous cooperations between the distributed systems research group of IME-USP and French institutions. The first projects were conducted together with INP Grenoble, and were mainly based on scheduling for parallel and distributed computing. Cooperations were supported either by CNPq-CNRS or USP-COFECUB. The main publications from the first cooperation were related to scheduling on clusters while the second cooperation focused on several aspects related to Grid Computing. More recently a very successful cooperation between the same Brazilian group and the research team Mascotte from INRIA-Sophia Antipolis was supported by FAPESP-INRIA. In this project, Evolving Graphs were used on several experiments on realistic settings and we were able to improve the state of art with several publications [Monteiro et al. 2006, Ferreira et al. 2009, Monteiro 2008, Ferreira et al. 2007a, Ferreira et al. 2007b]. A Brazilian student which did his master science degree during this project received a prestigious INRIA scholarship to do his PhD in another team of INRIA-Sophia Antipolis.

3. Proposal of a new Brazilian-French Cooperation

Some of the results and open questions raised on both CAPES/COFECUB cooperation (Regal group with UFBA) and FAPESP-INRIA cooperation (IME-USP distributed systems research group with INRIA Mascotte) have motivated the current proposal. The former has shown us the need of defining a more suitable and comprehensive distributed computing model that characterizes the dynamics of mobile networks and on top of which well-known distributed algorithms (e.g. consensus, unreliable failure detectors, partition detectors, etc.) would be built. At the same time, we believe that such a model is strongly related to the theory of Evolving Graphs [Ferreira 2004], since the latter is based on the definition of paths (journeys) that change over time and which we would like to exploit in our proposed model.

Due to arbitrary failures, energy restriction, disconnections, arrivals, departure, or mobility of nodes, a MANET is in fact an extremely dynamic system where connections between nodes change over time. The temporal variations in the network topology thus implies that MANET can not be viewed as a static connected graph over which paths between nodes are established before the sending of a message but it should be seen as a

set of dynamic graphs, as proposed by the Evolving Graph theory. Furthermore, lack of connectivity between nodes (temporal or not) makes of MANET a *partitionable system*, i.e., a system in which nodes that do not crash neither leave the system might be unable to communicate between themselves. It is worth mentioning that even if some works in the literature propose computing models for dynamic systems [Mostefaoui et al. 2005], [Tucci P. and Baldoni 2007], [Baldoni et al. 2007], none of them have considered that paths are dynamically established over the time nor that the system is a partionable one.

The specification of a distributed computing model for MANETs should try to characterize as much as possible the dynamic and self-organizing behavior of MANETs. In principle, the model should cover the following points:

- A node in MANET has not a global knowledge of the system. It neither knows the number of participants of the system nor the identity of them. In addition, such a number can be bounded or not;
- Nodes can move, crash/recover and dynamically enter or leave the system;
- The network is not fully connected and a node can only send messages to nodes that are within its transmission range. Hence, it may be that a message sent by a node should be routed through a set of intermediate nodes until reaching the destination node;
- Connections between nodes should be considered unidirectional. For instance, it might happen that a node can receive a message from another node but has insufficient remaining energy to send it back a message;
- A path between two nodes is built over the time, i.e., a connection between two intermediate nodes of a path is not necessarily established beforehand but dynamically whenever a node sends a message to the following one in the path;
- Different types of links can exist in the network: links can be reliable or not (loss of messages), they can present bounds for message transmission or not (synchrony assumptions), etc;
- Lack of links between nodes partitions the network into components.

Based on such dynamic characteristics, the reader can realize that the specification of a distributed computing model for MANETs is an interesting field of research. However, it is far from trivial. On the other hand, the previous mentioned cooperation FAPESP-INRIA validated the use of an interesting theory (Evolving Graphs) that characterizes dynamic systems. More than just the possible use of the theory itself there are several theoretically proved algorithms that could also be useful. We thus think that the theory of Evolving Graphs [Ferreira 2004] could help us and should be strongly exploited in the definition of our model since they characterize connectivity over time and temporal dependance of paths in dynamic systems like DTNs (disruption tolerant networks) [Arantes et al. 2009] or even MANETs. Concisely, an evolving graph is a time-step indexed sequence of subgraphs, where the subgraph at a given time-step corresponds to the network connectivity at the time interval indicated by the time-step value. Instead of dealing with end-to-end paths between nodes, there is the concept of journeys (paths that change over time). The FAPESP-INRIA cooperation project has proposed several optimal algorithms in order to calculate shortest journeys (with less nodes), foremost journeys (with smallest arrival time), and fastest journeys (with smallest time in "transit"). However, all these algorithms consider that nodes have a global knowledge, i.e., a centralized approach.

Another important aspect of MANETs that our model should consider is how to characterize stable periods in dynamic MANETs. In other words, since a MANET network is a partitionable system, its partitions (some or all) should present some *eventual stability condition* (or a *stability* whose duration is long enough) in order to ensure that applications that run on top of it can progress and terminate. In this sense, the theory of evolving graph could also be useful since in this theory a partition can be viewed as a so defined time-connected strongly connected component. Notice that it might be interesting to consider that stable partitions are not necessarily isolated one from another and also that the network presents some "stable partitions" or "connectivity islands" while the rest of the network has a dynamic behavior. Some applications could then run on stable partitions.

The main goals of our proposal is therefore:

- the implementation of a distributed version of the Evolving Graphs algorithms, i.e., nodes do not have a global view of the network;
- the definition of a distributed computing model for MANETs which (1) exploits the concept of Evolving Graph journey concept, (2) nodes do not have a global view of the network, (3) different types of links can be characterized and (4) network partitions can be detected;
- the characterization of time-dependent connected components, whose connections may be represented by journeys with bounded transmission delay which will thus allow the definition of stable partition concept.
- The proposal, implementation, and proof of correctness of well-known distributed algorithms (e.g. consensus, unreliable failure detectors, group communication, etc.) on top of the new model.
- a framework where the above algorithms could be easily implemented and tested.

References

GRID 5000. http://www.grid5000.org/.

Arantes, L., Goldman, A., and dos Santos, M. V. (2009). Using evolving graphs to evaluate dtn routing protocols. In *ExtremeCom 2009*. To be published.

Baldoni, R., Bertier, M., Raynal, M., and Tucci P., S. (2007). Looking for a definition of dynamic distributed systems. In *PaCT*, pages 1–14.

Ferreira, A. (2004). Building a reference combinatorial model for manets. *IEEE Network*, 18(5):24–29.

Ferreira, A., Goldman, A., and Monteiro, J. (2007a). On the evaluation of shortest journeys in dynamic networks. In *Proc. of the 6th IEEE International Symposium on Network Computing and Applications*, pages 3–10, Cambridge, MA, USA.

Ferreira, A., Goldman, A., and Monteiro, J. (2007b). Using evolving graphs foremost journey to evaluate ad-hoc routing protocols. In *Proc. of 25th Brazilian Symposium on Computer Networks (SBRC'07)*, pages 17–30, Belem, Brazil.

Ferreira, A., Goldman, A., and Monteiro, J. (2009). Performance evaluation of routing protocols for manets with known connectivity patterns using evolving graphs. *Wireless Networks*. In press.

Monteiro, J. (2008). The use of evolving graph combinatorial model in routing protocols for dynamic networks. In *Proc. of the XV Concurso Latinoamericano de Tesis de Maestría (CLEI'08)*, pages 41–57, Santa Fe, Argentina. Third prize in the CLEI'08 Master's Thesis Contexts.

Monteiro, J., Goldman, A., and Ferreira, A. (2006). Performance evaluation of dynamic networks using an evolving graph combinatorial model. In *Proc. of the 2nd IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob'06)*, pages 173–180.

Mostefaoui, A., Raynal, M., Travers, C., Patterson, S., Agrawal, D., and El Abbadi, A. (2005). From static distributed systems to dynamic systems. In *Proc. of the 24th IEEE Symposium on Reliable Distributed Systems*, pages 109–118.

Sens, P., Arantes, L., Bouillaguet, M., Simon, V., and Greve, F. (2008a). An unreliable failure detector for unknown and mobile networks. In *OPODIS*, pages 555–559.

Sens, P., Greve, F., Arantes, L., Bouillaguet, M., and Simon, V. (2008b). Um detector de falhas assincrono para redes moveis e auto-oganizaveis. In *SBRC: Simposio Brasileiro de Redes de Computadores*.

Tucci P., S. and Baldoni, R. (2007). Connectivity in eventually quiescent dynamic distributed systems. In *LADC*, pages 38–56.

Combinatória e Problemas em Redes de Telecomunicações

Frédéric Havet[1], **Cláudia Linhares Sales**[2*]

[1]Projet Mascotte, I3S(CNRS, UNS) et INRIA
2004 route des Lucioles BP93, 06902 Sophia Antipolis Cedex, France

[2]ParGO, Departamento de Computação, Universidade Federal do Ceará
Campus do Pici – Bloco 910 – CEP 60455-760 – Fortaleza – CE

fhavet@sophia.inria.fr, linhares@lia.ufc.br

Abstract. *In this paper, we summarize some problems arising in telecommunication networks which have been studied in the scope of the cooperation between our teams ParGO (UFC) and Mascotte (INRIA). We also present their modeling by graph coloring problems and some partial results we have obtained.*

Resumo. *Nesse artigo, vamos mostrar alguns dos problemas em redes de telecomunicações que vendo sendo abordados dentro da cooperação entre as equipes ParGO (UFC) e Mascotte (INRIA). Em particular, vamos mostrar a modelagem desse problemas por problemas de coloração em grafos e alguns resultados parciais que obtivemos.*

1. Introdução

Todos os problemas em redes de telecomunicações citados aqui foram modelados como problemas de coloração cuja versão clássica segue. Dado um grafo $G = (V, E)$, uma k-coloração própria de G é uma função $c \longmapsto \{1, \ldots, k\}$ tal que se u e v, vértices de G, são adjacentes, então $c(u) \neq c(v)$. O número cromático de G é o menor número de cores k tal que G admite uma k-coloração própria. O problema consiste em dado um grafo $G = (V, E)$, determinar o seu número cromático. Este problema é um dos mais estudados em teoria dos grafos pela sua relevância tanto do ponto de vista teórico, visto que até mesmo encontrar uma aproximação para o número cromático é um problema computacionalmente difícil [Yannakakis and Lundi 1994], como pela suas inúmeras aplicações. Nas seções que seguem abordaremos três problemas modelados por variações do problema de coloração acima.

2. Alocação de Frequências

Considere um conjunto de antenas V para o qual deve-se associar um **conjunto** de freqüências. As antenas não são iguais nas suas necessidades de freqüências e portanto associa-se através de uma função de peso $p : V(G) \rightarrow \mathcal{N}$, uma demanda a cada antena. A proximidade das antenas no espaço (usualmente no plano) provoca interferências ou ruídos nas comunicações. Essas interferências mútuas são modeladas por um conjunto de arestas E, ou seja, se há interferência entre duas antenas, há uma arestas entre os vértices correspondentes. Deseja-se, a priori, atribuir conjuntos de freqüências distintos a antenas

*Este autor é parcialmente financiado pelo CNPq

que se interferem mutuamente. Logo, a rede pode ser modelada por um grafo ponderado (G, p) para o qual deseja-se encontrar uma atribuição de conjuntos de freqüências a cada vértice de forma que o número de freqüências utilizadas seja mínimo e vértices adjacentes não compartilhem freqüências. Portanto, deseja-se, em outras palavras, determinar o menor inteiro k tal que G admite uma coloração onde cada vértice v de G recebe a quantidade de cores solicitada, denotada pelo seu peso $p(v)$, e que vértices adjacentes não compartilhem cores. O problema de determinar esse menor inteiro k é conhecido como o problema de coloração ponderada. Entretanto, do ponto de vista prático, algumas redes podem lidar com alguma, até um certo limite k, interferência entre as antenas. Ou seja, tolera-se que uma freqüência atribuída a uma antena seja compartilhada por até k de seus vizinhos. Trata-se da coloração k-imprópria de vértices, bastante estudada nos grafos não ponderados. Mais formalmente, o problema é definido como:

Definição 1 (Coloração ponderada k-imprópria) *Dado um grafo ponderado (G, p), uma r-coloração ponderada de (G, p) é uma função $C : V \mapsto \mathcal{P}\{1, \ldots, r\}$ tal que $|C(v)| \geq p(v)$. Uma coloração ponderada C de (G, p) é k-imprópria se para qualquer cor i, o conjunto de vértices colorido com a cor i induz um subgrafo de G de grau máximo k. O número cromático ponderado k-impróprio de (G, p), denotado por $\chi_k(G, p)$, é o menor l tal que (G, p) admite uma l-coloração ponderada k-imprópria.*

Observe que, para $k = 0$, o problema acima é equivalente ao problema de coloração ponderada mencionado anteriormente. Além disso, quando $k = 0$ e $p(v) = 1$, para todo $v \in V(G)$, o problema acima é equivalente ao problema clássico de coloração. Em 2000, McDiarmid and Reed [McDiarmid and Reed 2000] provaram que é $\mathcal{NP}$-completo decidir se o número cromático ponderado de uma malha hexagonal é 3 ou 4. Com respeito à coloração imprópria, Havet, Kang and Sereni [Havet et al. 2005] generalizaram esse resultado provando que para $0 \leq k \leq 5$, o problema de determinar se uma malha hexagonal ponderada é 3-colorível de forma k-imprópria é também $\mathcal{NP}$-dificil. Em 2007, J-C. Bermond et al determinaram o número cromático ponderado k-impróprio das malhas hexagonais [Bermond et al. 2007] e das grades quando todos os vértices têm pesos iguais [Bermond et al. 2009]. Quando os vértices da malha têm pesos distintos, apenas algoritmos aproximativos foram encontrados. Um algoritmo aproximativo para encontrar uma coloração k-imprópria com fator de aproximação α_k é dito α_k-aproximativo, significando que a coloração retornada pelo algoritmo é k-imprópria e utiliza no máximo $\alpha_k \times \chi_k(G, p) + c$ colors, onde c é uma constante.

Teorema 1 *Para $1 \leq k \leq 5$, existe um algoritmo α_k-aproximativo para encontrar uma coloração k-imprópria para uma malha hexagonal ponderada, onde $\alpha_1 = \frac{20}{11}$, $\alpha_2 = \frac{12}{7}$, $\alpha_3 = \frac{18}{13}$, $\alpha_4 = \frac{80}{63}$, and $\alpha_5 = \frac{41}{36}$.*

Teorema 2 *Para $1 \leq k \leq 3$, existe um algoritmo α_k-aproximativo para encontrar uma coloração k-imprópria para uma grade ponderada, onde $\alpha_1 = \frac{3}{1}$, $\alpha_2 = \frac{27}{20}$, and $\alpha_3 = \frac{19}{16}$.*

3. Alocação de Buffers em Canais de Comunicação

Considere uma rede de comunicação definindo um grafo $G = (V, E)$, onde os vértices são roteadores e as arestas representam canais de comunicação. Para controlar as interferências, suponha que cada roteador pode estar envolvido em no máximo uma comunicação a cada unidade de tempo. Logo, um conjunto de comunicações simultâneas

viáveis representa um *emparelhamento* do grafo. O problema de coloração proporcional se apresenta quando consideramos que as demandas de comunicação obedecem a um padrão com respeito à quantidade *bits* a serem enviados. Cada demanda representa na verdade um caminho em G, entre a origem e o destino, e um padrão de comunicação (com respeito à quantidade de bits). O envio dos dados ao longo do caminho provoca a ativação de um canal em uma certa proporção do tempo. Essa *proporção* é modelada por uma função de peso $w : E(G) \to [0, 1]$. O problema torna-se agora, encontrar, se possível, um escalonamento periódico da ativação dos canais satisfazendo às proporções. Para diminuir a fila de mensagens (*buffers*) nos roteadores e para aumentar o total de comunicações ao longo do tempo, deseja-se encontrar o menor período no qual as comunicações podem ser realizadas satisfazendo às proporções dadas. Esse problema levou à definição do seguinte parâmetro [Huc et al. 2008]:

Definição 2 (Coloração Proporcional) *Dado um grafo ponderado (G, w), uma* coloração proporcional *de (G, w) é uma função $C : E \to \mathcal{P}(\{1, \ldots, c\})$ tal que para toda $e \in E$, temos*

1. $|C(e)| \geq cw(e)$*; e*
2. *para todo $e, f \in E^2$, $e \cap f \neq \emptyset \Rightarrow C(e) \cap C(f) = \emptyset$.*

Chamamos de índice cromático proporcional de G, $\chi'_\pi(G, w)$, o número mínimo de cores para o qual uma coloração proporcional de (G, w) existe. Se tal coloração não existe, definimos $\chi'_\pi(G, w) = \infty$.

Observe que há instâncias do problema que não têm solução. Nesses casos, pela definição acima, $\chi'_\pi(G, w) = \infty$. Determinar $\chi'_\pi(G, w)$ é $\mathcal{NP}$-difícil, entretanto, dada uma instância qualquer do problema, pode-se determinar se ela admite ou não uma solução em tempo polinomial [Huc et al. 2008]. F. Huc et al mostraram que limites inferiores e superiores podem ser obtidos em tempo polinomial. Além disso, no mesmo trabalho, mostraram que o problema é fácil para grafos bipartite ponderados, quando as instâncias admitem solução.

4. Protocolos em Redes Distribuídas de Duplo Acesso

Considere uma rede para transmissão de mensagens de tempo real formada por uma seqüência de estações $s_1, \ldots, s_n$, ligadas por canais que transmitem dados em apenas uma direção, ou seja, a estação s_i transmite dados para uma estação s_j, $j > i$, por um canal, mas existe outro canal para transmitir dados de s_j para s_i. Existe ainda nessa rede um gerenciador de pacotes que cria pacotes de tamanho fixo para o envio das mensagens que possui um controle da alocação da banda e dos pacotes, através de identificadores de circuitos virtuais. Quando uma estação s_i deseja enviar dados para uma estação s_j, $j > i$, ela faz uma requisição ao gerenciador de pacotes. O gerenciador a envia os pacotes em quantidade suficiente para que sejam enviados todos os dados, com o seu devido identificador de circuito virtual. A estação aloca os dados nos pacotes e os envia para a estação seguinte s_{i+i} e assim por diante, até a estação s_j, onde os dados são desempacotados e os pacotes continuam o trajeto até o fim da rede, passando por todas as estações de s_j a s_n. Esse tipo de protocolo de rede é conhecido como protocolo de controle de acesso de mídia em redes distribuídas de duplo canal.

Para melhor reaproveitar os pacotes após seu desempacotamento, uma outra versão para o problema de Coloração Ponderada pode ser utilizado neste protocolo.

Suponha que um conjunto de dados precisam ser transmitidos entre diversos pares de estações. Construa um grafo G ponderado, simples e não-direcionado de tal forma que para cada dado a ser transmitido é associado um vértice. Se dois dados não podem compartilhar o mesmo conjunto de pacotes, ou seja, se a estação de destino do primeiro a ser enviado possui um índice maior que o da a estação de origem do segundo, então é adicionada uma aresta entre os vértices que representam esses dados. O peso de cada vértice deve ser a quantidade de pacotes que o dado associado precisa para ser transmitido. Uma coloração ponderada ótima de G representa uma atribuição de identificadores de circuitos virtuais aos dados que otimiza a reutilização dos pacotes desse protocolo, pois os vértices que receberam a mesma cor devem ser enviados pelo mesmo conjunto de pacotes e a quantidade de pacotes de cada circuito virtual corresponde exatamente ao maior peso de um vértice nesse circuito. Definimos a seguir formalmente o problema de coloração ponderada, definido por Guan e Zhu [Guan and Zhu 1997]. Em uma coloração própria, os vértices coloridos com uma mesma cor i formam um conjunto independente que é chamado usualmente de *classe de cor*. Sendo assim, uma k-coloração possui k classes de cores. Dada uma k-coloração $c = \{c_1, c_2, \ldots, c_k\}$ de G, o peso de cada classe de cor de c_i, $w(c_i)$ é dado pelo maior peso de um vértice em c_i.

Definição 3 (Número Cromático Ponderado) *O* número cromático ponderado *de* $G = (V, E)$, $\chi_p(G)$ *é o mínimo* $\sum_{c_i \in c} w(c_i)$, *para toda coloração própria de* G.

Observe que quando todos os vértices do grafo têm peso unitário, esse problema corresponde ao problema clássico de coloração. Dada a dificuldade do problema de coloração ponderada, tenta-se encontrar resultados para subclasses de grafos. Para grafos com largura em árvore limitada, B. Reed e C. Linhares Sales determinaram a ordem de grandeza do número cromático ponderado [Linhares Sales and Reed 2006]:

Teorema 3 *Seja* $G = (V, E)$ *qualquer grafo ponderado com largura em árvore limitada* ω. *Então,* $\chi_p(G)$ *é* $O(\omega \lg n)$.

Para uma subclasse dos grafos P_4-esparsos, J.C. Araújo, C. Linhares Sales e I. Sau obtiveram [Araújo et al. 2009] o seguinte resultado:

Teorema 4 *Seja* G *um grafo* P_4*-esparso cuja a árvore de decomposição modular não contém nós paralelos. Então o número cromático ponderado de* G *pode ser determinado em tempo polinomial.*

References

Araújo, J., Linhares Sales, C., and Sau, I. (2009). Weighted coloring in p_4-sparse graphs. manuscript in preparation.

Bermond, J.-C., Havet, F., Huc, F., and Linhares Sales, C. (2007). Allocation de fréquences et coloration impropre des graphes hexagonaux ponderés. *In: AlgoTel 2007 (9ème rencontres francophones sur les aspects algorithmiques de télécommunications)*, 9:53–56.

Bermond, J.-C., Havet, F., Huc, F., and Linhares Sales, C. (2009). Improper coloring of weighted grid and hexagonal graphs. manuscript in preparation.

Guan, D. and Zhu, X. (1997). A coloring problem for weighted graphs. *Information Processing Letters*, 61:77–81.

Havet, F., Kang, R., and Sereni, J.-S. (2005). Improper colourings of unist disk graphs. *Eletronic Notes in Discrete Mathematics*, 22:123–128.

Huc, F., Linhares Sales, C., and Rivano, H. (2008). The proportional colouring problem: Optimizing buffers in radio mesh networks. *Eletronic Notes in Discrete Mathematics*, 30:141–146.

Linhares Sales, C. and Reed, R. (2006). Weighted coloring on graphs with bounded tree width. *In: 19th International Symposium on Mathematical Programming*, pages 146–146.

McDiarmid, C. and Reed, B. (2000). Channel assignment and weighted coloring. *Networks*, 36:114–117.

Yannakakis, M. and Lundi, C. (1994). On the hardness of approximating minimization problems. *Journal of ACM*, 41(5):960–981.

Using Heuristics To Guide Anisotropic Diffusion Filtering

Leandro Coser [1,2], Antonio Carlos Sobieranski [1], Eros Comunello [1], Aldo von Wangenheim [1,2]

[1]UFSC-INE-LAPIX Universidade Federal de Santa Catarina, Departamento de Informática e Estatística, Laboratório de Processamento de Imagens e Computação Gráfica – Florianópolis – SC – Brasil.

[2]UFSC-EGC Universidade Federal de Santa Catarina, Departamento de Engenharia e Gestão do Conhecimento – Florianópolis – SC – Brasil.

{leandro, asobieranski, eros}@cyclops.ufsc.br, awangenh@inf.ufsc.br

Abstract. *Anisotropic diffusion filtering is a well-established technique for image enhancement that works by means of diffusion functions. They are able to smooth images without destroying edge information. However, when many filtering iterations are applied or higher contrast parameter are used, edges gradually fade away and are ultimately smoothed by the process. We propose the adoption of a color gradient map and an adaptive contrast parameter to guide the smoothing in order to preserve the edges even after many iterations. Preliminary experiments show good results when compared with the traditional anisotropic diffusion filter.*

1. Introduction

The anisotropic diffusion filter discussed by Weickert (1998,2001) is a powerful image processing technique for noise removal and image enhancement. The filter works by performing smoothing on the image but at the same time preserving the boundaries between different regions. The process is controlled by a diffusion matrix that measures the color variation on the neighborhood of a hot spot and by a contrast parameter (λ) that defines where diffusion should be performed.

When the filter is applied over many iterations, the diffusion matrix slowly becomes adapted to the new local colors. When the original image contains tenuous edges, the smoothing process will gradually erase all edge information as iterations are sequentially applied. Other related problem is when higher λ values are used, because the matrix-kernel or diffusion tensor acts like a simple convolution operation over tenuous edges.

In this work, we propose improvements to the anisotropic diffusion filter that introducing a color gradient, which behaves as static boundary evidence, and local adapting λ contrast parameter. This heuristics enhances the response of the filter when applied several times to an image containing tenuous edges and/or higher λ values.

2. Orienting the Anisotropic Diffusion Filter

The anisotropic diffusion filter used in our approach has been extensively discussed by Weickert (2001). We can be regarded as a convolution technique with an adaptive matrix-valued kernel that performs a special smoothing on images. The anisotropic diffusion inhibits the smoothing on edge pixels and stimulates it on internal regions. The

basic diffusion equation (BROX, 2005) for an image $I(x, y)$ with M channels and a signal initialized with $u(x, y, 0) = I(x, y)$ is

$$\partial_t u_i = \operatorname{div}\left(D\left(\sum_{k=1}^{M} \nabla u_k \nabla u_k^{\mathrm{T}}\right)\nabla u_i\right) \quad , (1)$$

where D is a matrix-valued function or diffusion tensor, and $i = 1, \ldots, M$ are the individual channels. Each component of tensor D can be computed by the follow diffusivity equation given by:

$$g(x) = e^{-\left[\frac{x^2}{\lambda}\right]} \quad , (2)$$

where x^2 denote variation in the region over the hot spot, and λ determines how strong the diffusion must be into a region.

As the diffusion process is carried on through several iterations, the edges of the original image gradually fade away, because tensor D takes into account only the results from the previous iteration. After a certain amount of iterations, even edges that were initially well defined may become blurred. In order to keep well-defined edges present in the original image, we propose the introduction of a static factor G_i in equation (1) and an adaptive contrast parameter λ in Equation (2). These properties depend on the original image only, avoiding the effects introduced by the repeated application of the diffusion filter.

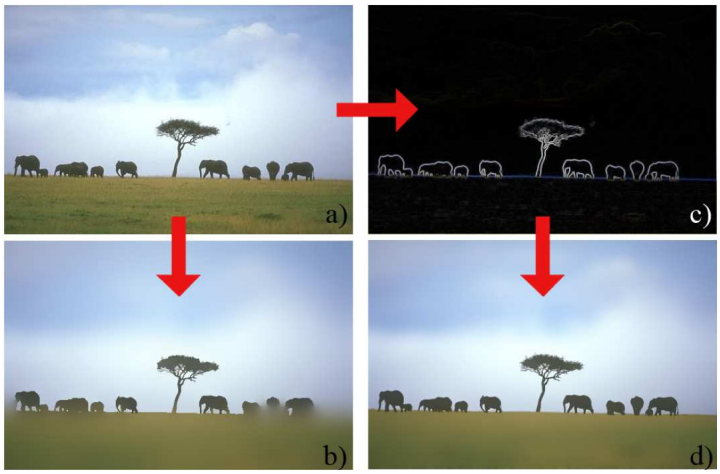


Figure 1. Overview of traditional and gradient map-oriented filtering. a) the original image; b) the traditional filtering; c) the gradient map; and d) the filtering oriented by the gradient map and adaptive λ.

An overview of this process is shown in Figure 1, where both versions of the filter are applied over 300 iterations. In a) is shown the original image used by the both approaches. In b) is demonstrated the result of traditional filtering algorithm and in d) the results of the proposed approach employing the color gradient map (in c) and the adaptive contrast parameter.

2.1. Color Gradient Map

The color gradient map G is a static factor obtained from the original image and remains constant throughout all iterations. In order to accommodate G_i, Equation (2) can be rewritten as

$$\partial_t u_i = \operatorname{div}\left(D\left(G_i \sum_{k=1}^{M} \nabla u_k \nabla u_k^{\mathrm{T}}\right)\nabla u_i\right) \quad . (3)$$

The gradient map G is calculated by a simple convolution operation using the following masks:

$$I_x = \frac{1}{4}\begin{pmatrix} -b & 0 & b \\ -a & 0 & a \\ -b & 0 & b \end{pmatrix} \text{ and } I_y = \frac{1}{4}\begin{pmatrix} -b & -a & -b \\ 0 & 0 & 0 \\ b & a & b \end{pmatrix} \quad . (4)$$

where $a = 2(\sqrt{2}-1)$ e $b = (2-\sqrt{2})$. The modulus of the vector (I_x, I_y) is then used as an estimation of the gradient for each channel. In figure 1-c is shown an example the color gradient map produced for each color channel on the image. As we can observe in equation 3, the color gradient map locally inhibits the diffusion on high gradient responses, and smooths intra-regions of the objects, as demonstrated in figure 1-d.

2.2. Adaptive Contrast Parameter

The adaptive contrast parameter λ is a heuristic used to control the diffusion at a region on the image. This is done by adapting the λ parameter according to the global and local neighborhood variation. The λ parameter is perceptually readjusted by increasing or decreasing its responses on diffusivity function (equation 2).

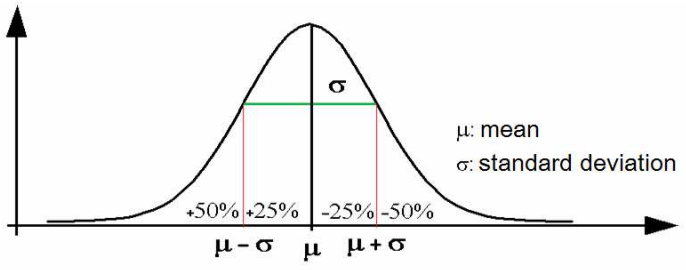


Figure 2. Lambda inhibition or stimulation regarding of local variation and global mean.

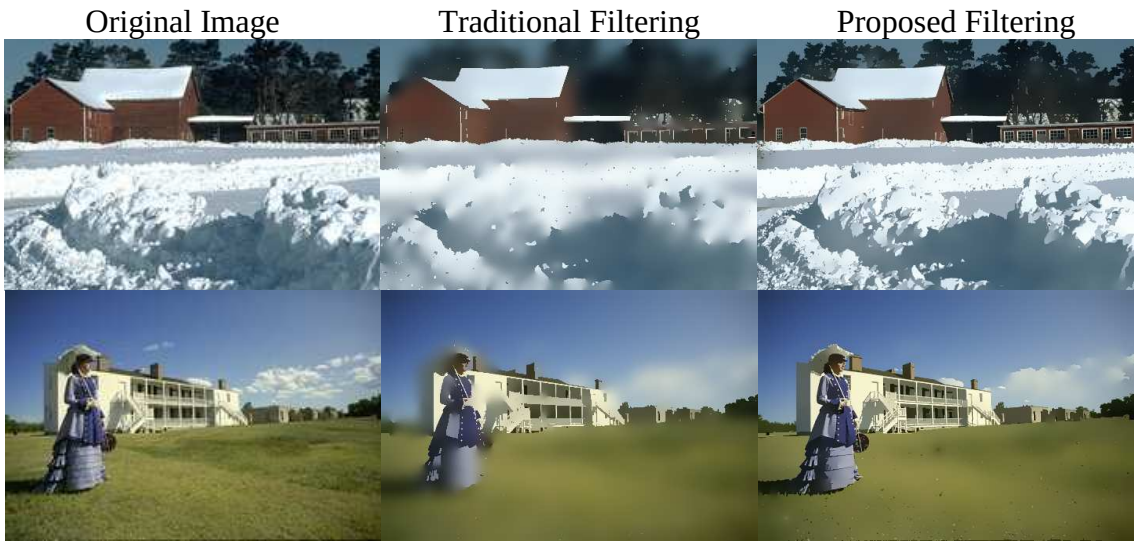


Figure 3. From left; original images, traditional filtering and proposed filtering.

In the figure 2 it is illustrated four ranges used to locally readjust the λ parameter according to the global arithmetic mean (μ) of RGB and the global standard deviation (σ) of RGB components. The inhibition is made decreasing the λ parameter by 25% if color components are above the average and up to one σ and by 50% if above the average and greater than σ. The stimulation is made by increasing the λ parameter by 25% if color components are bellow the average and up to one σ and by 50% if color components are bellow the average and greater than σ.

Finally, the λ is locally adapted allowing to perform strong smoothing for regions with low variation, and weak smooth for higher variations.

3. Conclusions and Discussions

The anisotropic diffusion filter is a powerful tool to improve the quality of images for its sophisticated border-preserving smoothing. We have shown how this filter can be further improved by using a color gradient map that remains the edge pixels unchanged over iterations and also by adaptive contrast parameter which controls the smoothing over gradient variation. These two heuristics preserve edges that are well defined in the original image.

In the figure 3 it is presented the filtering results of the both color gradient map and adaptive λ processed by the same parameters (iterations=120, λ=40). As it can be observed, pixels in the regions with high gradient levels are still preserved, while those in the regions with low gradient are smoothed. The internal regions are greatly smoothed, and the distinction between them remains clear. Other 40 images can be observed in http://150.162.202.1/m_adf/adf.html .

The traditional anisotropic diffusion filtering version has been explored in a parallel environment as described in (SOBIERANSKI, 2008). This new approach also could be easily extended to the environment.

In harmony with INRIA researches labs like WILLOW(Models of visual object recognition and scene understanding) and TEMICS(Digital image processing, modeling and communication) the improvements of anisotropic diffusion filter can be further explored with other applications context such as pre-processing in image segmentation for object recognition and video processing.

4. References

Brox, T. From Pixels to Regions: Partial Differential Equations in Image Analysis. PhD thesis, Saarland University, 2005.

Weickert, J. Anisotropic Diffusion in Image Processing. Teubner-Verlag, 1998.

Weickert, J. Applications of nonlinear diffusion in image processing and computer vision. Acta Math. Univ. Comenianae, 70:33–50, 2001.

Sobieranski, A.C, Coser, L, Dantas M.A.R., Wangenheim A, Comunello E., "An Anisotropic Diffusion Filtering Implementation to Execute in Parallel Distributed Systems," cseworkshops, pp.182-187, 2008 11th IEEE International Conference on Computational Science and Engineering - Workshops, 2008.

Supporting MPI Malleable Applications upon the OAR Resource Manager

Márcia Cristina Cera [1], **Yiannis Georgiou** [2], **Olivier Richard** [2], **Nicolas Maillard**[1], **and Philippe O. A. Navaux** [1]

[1] Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil
{marcia.cera, nicolas, navaux}@inf.ufrgs.br

[2]Equipe MESCAL – Laboratoire d'Informatique de Grenoble
ZIRST 51, avenue Jean Kuntzmann - 38330 - Montbonnot Saint Martin - France
{Yiannis.Georgiou, Olivier.Richard}@imag.fr

Abstract. *Malleable applications are able to adapt themselves, at execution time, to changes in the amount of resources available. Developing applications with a malleable behavior require some flexibility from programming environment. For instance, MPI-2 provides dynamic processes creation, which can be employed to add some flexibility to MPI applications. Further, an environment that can provide dynamic resources is also required to enable malleability. The OAR resource manager is an open source system composed by high level components, which can be easily extended to integrate new features. Upon this context, this paper exposes how to develop malleable applications using the MPI-2 features, in which the dynamic resources are known through interactions with OAR. Our first results shown that OAR will be able to support malleable jobs and that malleability can provide indeed a better resource utilization with an improvement of almost 35%. Besides resource utilization improvement, we present a discussion about some other advantages brought by some flexibility in job allocation.*

1. Introduction

Malleable applications perform growing and shrinking operations to adapt themselves to changes in resources availability. The development of MPI applications with certain flexibility become possible with the MPI-2 norm, for instance using dynamic process creation [Gropp et al. 1999]. With dynamic processes, when a MPI malleable application performs growing it grants some workload to new resources through processes spawning. In shrinking, processes running on resources announced as unavailable must be stopped and some fault tolerance mechanism is required to avoid crashes. We adopted a simple one: tasks running in resources being shrunk are identified and will be restart in the future. It is not optimal, but ensures application results correctness.

To develop MPI applications with dynamic processes, we designed a scheduler able to determine the physical location of spawning processes [Cera et al. 2006]. Such work was recently upgraded to handle malleable operations take into account resources with dynamic availability. Furthermore, some support in the resource manager is required to provide malleability. OAR [Capit et al. 2005] is a modular and extensible resource manager which we employ to provide malleability. An OAR module was developed to

manage and identify the resources availability, sending such information to the MPI application, which will adapts itself to it. This paper goal is presents the efforts to add some flexibility in job allocation aiming at resources utilization improvements. This work has been developed as part of an international cooperation project between the LIG laboratory and the GPPD/UFRGS (*Grupo de Processamento Paralelo e Distribuído*) group. Many experiences and expertises are changed between both, since the French group focuses in resource management and the Brazilian one in adaptable parallel applications. This paper presents our study case composed by a MPI malleable application, which are scheduled and launched by the OAR resource manager – Section 2. Experimental results are exposed in Section 3. Also, Section 4 discusses some advantages of malleable jobs and, at end, Section 5 concludes this paper.

2. Running MPI Malleable Applications upon OAR

Due to the modularity and flexibility of OAR resource manager, it was possible to construct a prototype, aside the core of OAR, enabling the execution of malleable jobs. The prototype is based on two notions: *(i) Best Effort* job, which is low priority job able to harness the idle resources of a cluster but it can be directly killed when the resource is required; and *(ii)* resource discovery command which provides the current and near-future resources availability. To provide malleable jobs in OAR, we consider that they are composed by a rigid and a *Best Effort* part. The rigid part stays always intact so that it can guarantee the job completeness. In the same time, the *Best Effort* part is responsible for the job flexibility. Hence, using the resource discovery command which informs the application for the variations on resources availability: *Best Effort* jobs can be killed meaning application shrinking or further submitted allowing the application growing.

MPI-2 dynamic process creation [Gropp et al. 1999] can be employed to enable malleability, but some care in application development are demanded. Issues like *when* and *where* spawn processes must be on-line answered. Here, the answer is spawn processes in new resources when they become available. Our previous work provide a scheduler able to determine the physical location of dynamic processes [Cera et al. 2006]. Such scheduler was upgraded to determine the location taken into account the resource availability information provided by OAR. In this way, the scheduler launches malleable operations, which will adapt the application according the resources available. In other words, scheduler performs like a bridge between the MPI application and the OAR.

Our first initiative consider a simple scenario where one malleable application performs upon all idle resources. Extensions of this scenario are awaited in future works. Considering our initial context, the rigid part of the MPI malleable application is the minimum unit required to execute it, *i.e.* one node in our case. The *Best Effort* part is as large as have resources available, which is determined by the resource discovery command. Malleable application starts with all resources answered by the discovery command. When a *Best Effort* job is further submitted characterizing a growing, the scheduler is notified and update its control structures. In the application side, after identify that there are new resources available, it spawns processes charging them. Scheduler will ensure that the spawning processes will be placed in the new resources. In the opposite case, when some nodes are demanded to satisfy arriving jobs, they will required from the *Best Effort* part. The scheduler is notified, performing the shrinking operation and the fault tolerance

procedures. To ensure that this operation will be safe performed, a grace time delay[1] is applied on OAR before the killing of the *Best Effort* jobs [Georgiou et al. 2007].

3. Malleability Improving the Resource Utilization

Aiming to verify the impact of malleable job execution together with rigid ones, we choose a static workload of 5 hours slice of DAS2 workload[2] with 40% cluster utilization. This workload is injected in OAR charging the resources and representing the normal workload of the cluster. At same time one malleable job per time, is submitted and will run upon the free resources, *i.e.* those are not used by the normal workload.

MPI malleable application employed in tests is an implementation of Mandelbrot set, which has adapted to performs growing and shrinking operations. Although Mandelbrot problem do not require a malleable behavior to be solved, we decide use a know MPI application in our tests because our target is the support of malleable operations. The minimum of resources required to start the malleable application is one node and all malleability operations will be performed using whole nodes. In this way, the normal workload jobs are simple 'sleep' jobs just occupying the resources for a specific time, since they do not affect the malleable execution. Results of executing malleable jobs (one per time) are compared to a non-dynamic approach. In other words, the malleable jobs submission is substituted by moldable-besteffort jobs submission. As moldable-besteffort job we define a moldable job, that can be executed upon all the free cluster resources but does not provide flexibility during runtime hence it will be immediately killed (like a besteffort job) when a rigid job asks for resources.

In terms of resources utilization, since the workload makes use of 40% of the cluster, this leaves a 60% of free resources. We observed that moldable-besteffort jobs use a 32% of the idle resources arriving at 72% of total cluster utilization. On the other hand, the malleable jobs use 57% of the idle resources arriving at 97% of overall cluster utilization. Hence, the improvement of the dynamic approach when comparing with the non-dynamic approach is almost **35%** of resources utilization. Furthermore, we observed the number of jobs executed in 5 hours of experimentation. In the dynamic context we obtain 8 successfully Terminated malleable jobs compared to 4 Terminated and 5 in Error State for the non-dynamic context. Finally the impact of the response time for the normal workload was also measured and the result was 8 sec of average response time in case of non-dynamic context, compared to 44 sec for the dynamic one. The response time for malleable MPI approach is explained by the grace time delay to ensure safe shrinking operations, which was 40 sec. Table 1 resumes the results exposed.

Table 1. Comparison between malleable and moldable-besteffort jobs.

	idle resources used	jobs Terminated	jobs Error	response time
Malleable	57%	8	0	44 sec
Moldable-besteffort	32%	4	5	8 sec

[1] Amount of the time that the OAR system waits before destinate the resources to another job, ensuring that they are free.

[2] http://www.cs.huji.ac.il/labs/parallel/workload/l_das2/index.html

4. Discussion about Malleability Advantages

Previously, we exposed the gain in resource utilization brought by the execution of malleable jobs (one per time) and rigid ones. Kalé et al. [Kalé et al. 2000] exposes some other advantages of malleable jobs. In this section, we discuss these advantages and make a relation with OAR and MPI malleable application context.

- Resize low priority jobs to respond to higher priority ones, insted of preempt them. In Table 1, malleable jobs (low priority) are resized to supply the arriving rigid ones and always finish their execution successfully. On the other side, moldable-besteffort jobs are sometimes preempted and get Error Status;
- Use idle-cycle jobs. MPI malleable jobs performs in the gaps of cluster resources enabling the utilization of idle-cycles. As shows in Table 1, the idle-cycle utilization of malleable MPI jobs is bigger than the moldable-besteffort ones;
- Improve job efficiency by its shrinking. Applications with a poor scalability usually are inefficient with many resources. In these cases, destinate some resources to other applications can improve their performance. This is not the case of Mandelbrot set that has a trivial parallelization. But this feature can taken into account in OAR requiring adaptations in the scheduling policies;
- Small-duration jobs can be promptly responded by shrinking malleable jobs. OAR can shrinking malleable jobs to attend small jobs reducing their waiting time, and after growing again when the small jobs finishes.

5. Conclusion

This paper shown a malleable application that take advantage of MPI-2 dynamic features. Such application, helped by a dynamic processes scheduler, execute upon the unused resources of a cluster improving its utilization. The malleable applications are launched by OAR resource manager, which is responsible to manage the resources dynamicity. OAR and the scheduler interact exchanging informations about the resources availability. In this paper, we have shown that malleable jobs upon OAR can improve the resource utilization of almost 35%. This work is part of a cooperation project between LIG, research about OAR, and GPPD/UFRGS, research about MPI malleability.

References

Capit, N. et al. (2005). A batch scheduler with high level components. In *5th International Symposium on Cluster Computing and the Grid*, pages 776–783. IEEE Comp. Society.

Cera, M. et al. (2006). Improving the dynamic creation of processes in mpi-2. In *13th European PVMMPI Users Group Meeting*, volume 4192 of *LNCS*, pages 247–255.

Georgiou, Y. et al. (2007). Evaluations of the lightweight grid cigri upon the grid5000 platform. In *3th IEEE Int. Conf. on e-Science and Grid Computing*, pages 279–286.

Gropp, W., Lusk, E., and Thakur, R. (1999). *Using MPI-2 Advanced Features of the Message-Passing Interface*. The MIT Press, Cambridge, Massachusetts, USA.

Kalé, L. V., Kumar, S., and DeSouza, J. (2000). An adaptive job scheduler for time-shared parallel machines. Technical Report 00-02, Parallel Programming Laboratory, Department of Computer Science, University of Illinois at Urbana-Champaign.

VI

Massive Processing of Multimedia Data, Virtual Reality

COLIBRI 2009 • 22-23 de julho de 2009 • ISBN 978-85-7669-245-4

SARAVÁ: data sharing for online communities in P2P

Marta Mattoso[1], Esther Pacitti[2], Patrick Valduriez[2], Reza Akbarinia[2], Vanessa Braganholo[3], Alexandre A. B. Lima[4]

[1]COPPE, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
[2]INRIA and LINA, Nantes, France
[3]Computer Science Department, Federal University of Rio de Janeiro, Brazil
[4]School of Science and Technology, Unigranrio University, Rio de Janeiro, Brazil

marta@cos.ufrj.br, Esther.Pacitti@univ-nantes.fr,
Patrick.Valduriez@inria.fr, Reza.Akbarinia@inria.fr,
braganholo@dcc.ufrj.br, abento@unigranrio.com.br

Abstract. *This paper describes SARAVÁ, a research project that aims at investigating new challenges in P2P data sharing for online communities. The major advantage of P2P is a completely decentralized approach to data sharing which does not require centralized administration. Users may be in high numbers and interested in different kinds of collaboration and sharing their knowledge, ideas, experiences, etc. Data sources can be in high numbers, fairly autonomous, i.e. locally owned and controlled, and highly heterogeneous with different semantics and structures. Our project deals with new, decentralized data management techniques that scale up while addressing the autonomy, dynamic behavior and heterogeneity of both users and data sources. In this context, we focus on two major problems: query processing with uncertain data and management of scientific workflows.*

1. Introduction

Online communities such as social networks (e.g. sites like MySpace and Facebook) and professional communities (e.g. scientific communities, online technical support groups) are becoming a major killer application of the web. Community members typically have common interests or purposes and are willing to share their data. Through data sharing, online communities foster mass collaboration whereby large numbers of members can contribute in parallel to a large project, e.g. developing a world-wide encyclopedia such as Wikipedia.

To scale up to very large amounts of data while providing performance and availability for ever growing numbers of users, they can rely on a large data center (e.g. using a cluster computer). However, this solution is expensive (both in hardware/software and electricity), requires skilled staff to administer, tune and repair the system and is hard as it requires centralized integration of heterogeneous data. Furthermore, community members may not trust the administrators of the centralized servers and prefer to maintain their data in their computer under their own control. Thus, radically new data sharing techniques where the user plays a central role are needed. A promising solution is to organize community members in a peer-to-peer (P2P) architecture where each member can share data with the others through a P2P overlay network. The major

advantage of P2P is a completely decentralized approach to data sharing which does not require centralized administration.

This paper describes SARAVÁ, a research project that aims at investigating new challenges in P2P data sharing for online communities. In Portuguese, SARAVÁ means "good luck", that is certainly an interesting name for a project that involves collaboration among Brazilian and French teams that have been working on successful projects over the past years (Ecobase, DAAD and GriData, to cite just a few).

2. Data Sharing in P2P: research challenges

The general problem we address in our project is P2P data sharing for online communities, by offering a high-level network ring [Abiteboul and Polyzotis 2007] across distributed data source owners. Users may be in high numbers and interested in different kinds of collaboration and sharing their knowledge, ideas, experiences, etc. Data sources can be in high numbers, fairly autonomous, i.e. locally owned and controlled, and highly heterogeneous with different semantics and structures. What we need then is new, decentralized data management techniques that scale up while addressing the autonomy, dynamic behavior and heterogeneity of both users and data sources.

To illustrate the data management requirements in this context, consider two representative, rather complex examples of online community applications: collaborative scientific research (e.g. bioinformatics) and social networking systems. As many other online community applications, these two applications have common requirements (e.g. high level data access, update support, data privacy,) and differences. For instance, collaborative scientific research may be quite demanding in terms of quantity of data exchanged while social networks may involve very high numbers of participants with dynamic behavior.

The large quantity of scientific data should be processed as a data warehouse. Traditionally, decision support queries (OLAP) are based on fully integrated data sources. Recent researches investigate techniques for OLAP query processing in computing grids [Furtado 2006, Paes et al. 2008], where data integration is not so strong. The dynamicity of P2P environments makes the data sources very loosely coupled and impose new challenges for OLAP. We intend to adapt the techniques proposed for clusters [Lima et al. 2009] and grids [Kotowski et al. 2008] for such an environment.

Also, collaborative scientific research typically requires input data to be processed by workflows of heterogeneous programs while social networks require high-level query capabilities as well as key-word search. A P2P architecture for data management, e.g. APPA [Akbarinia et al. 2006], provides important advantages like decentralized control and administration, scale up to high numbers of peers and support of the dynamic behavior of peers (who may join or leave the system at will) [Valduriez and Pacitti 2005]. These advantages are important for online communities.

In addition to the traditional requirements of distributed database or data integration systems [Özsu and Valduriez 1999] such as query expressiveness, semantic data integration, data replication and caching, we must deal with some new, hard problems:

- *Query processing with uncertain data*. Some data should not be assumed to be 100% certain, precise or correct, in particular, when coming from peers with different levels of confidence. Query processing techniques designed for P2P systems, e.g. top-k queries [Akbarinia, Pacitti and Valduriez 2007] should be revisited to deal with data uncertainty at all levels. Similarly, the recent extensions of DBMS with support for data uncertainty [Agrawal et al. 2006] should be revisited for P2P.
- *Scientific workflow management*. In a P2P setting, the traditional parallel techniques to manage scientific workflows no longer apply because many workflow components are legacy programs. Therefore, it is not possible to modify their code as they have to be treated like black box components [Meyer et al. 2005, Serra da Cruz et al. 2008]. Furthermore, exploiting the massive scale of computing power available in desktop peers for workflow management is an open problem [Pacitti, Valduriez and Mattoso 2007].

3. Main Project Tasks

We plan to pursue parallel work on the following three main tasks:

1. **P2P data management architecture**. Architectural work will be ongoing throughout the project as it is very hard to come up with the right architecture upfront, without having solved the research issues and come up with efficient techniques. Thus, in 2009, we will start by adopting the Data Ring vision [[AP07] which is well suited for community data sharing and general enough so we can adapt it to our requirements. In a Data Ring, we can model an online community as a set of *participants* and *relationships* between them. The participants are autonomous data sources (under their own control) in the network and can be heterogeneous (e.g. relational, XML, files) with very different levels of processing capabilities (e.g. ranging from a DBMS to a file system). The relationships are between any two or more participants and indicate how their data sources are related, e.g. one source is a copy of the other, or some sources share the same semantic domain or schema. Based on a selected collaborative scientific application from bio-informatics in Brazil and a social network application, we will analyze the requirements for the data model and query language for uncertain data, and for query processing and workflow management. We will reflect these requirements in application scenarios which should be useful for the two next tasks. For selecting the data model and query language with uncertainty, we will capitalize on probabilistic data models and study the support of imprecision on attribute values using fuzzy logic. In a P2P system with heterogeneous data sources, we need a model that can describe various annotated and structured data.

2. **P2P query processing with uncertain data**. To address P2P query processing with data uncertainty, we must deal with the problems of uncertain query routing (to the nodes holding relevant data), and ranking results with uncertain data. For uncertain query routing, P2P query routing algorithms cannot be used without significant revision. The reason is that a query must be routed only to the nodes which involve relevant data with certainties higher than a given certainty, i.e. that expressed in the query. Another major problem is the ranking of results since for each query, there

may be huge quantities of answers, most of them uninteresting for the user. This requires support of top-k queries with new techniques to deal with uncertainty. We will start designing query optimization and query routing techniques and reflect them to task 1, in terms of a query processing service. We will also validate these techniques by implementation and experimentation over the Grid5000 and Planetlab platforms, using synthetic and real data.

3. **P2P workflow management**. One typical scenario in genomic applications is having a set of input data to be processed by a workflow. Bioinformatics programs often generate and process large datasets. Thus, one question that can be asked is how to process this genome workflow in parallel on a high number of heterogeneous peers. There are many alternatives to execute a genome workflow in parallel because programs and data can be distributed among the peers in many different ways. Choosing the best strategy for parallel execution in a P2P system is difficult because this choice must consider: (i) the dependencies among the components; (ii) the unbalanced execution time of the peers; (iii) the different size of datasets; and (iv) the computational resources available at peers. The choice of the best execution alternative is much more difficult in a P2P system than in a parallel machine or a cluster due to the heterogeneity and the dynamic nature of the peers. Decisions to replicate code and datasets have to be taken either on demand or by pre-staging in order to provide better performance. We have started designing workflow management techniques to exploit parallel execution to reflect them to task 1, in terms of a workflow management service. Preliminary results using VisTrails workflow management system and a cluster machine show significant performance improvements. We will also validate these techniques by implementation and experimentation over the Grid5000 and Planetlab platforms, using synthetic and real data.

4. Concluding Remarks

The general impact of this collaboration on the two partners should be high quality joint papers, co-advising and exchange of Ph.D. students, training of Ms. students, dissemination of results through workshops jointly organized as well as special issues in top journals such as JOGC. Furthermore, the experience of both teams on using regularly the Grid'5000 platform [Furtado et al. 2008, Kotowski et al. 2008, Paes et al. 2008, Lima et al. 2009] fosters our joint work.

We also designed the SARAVÁ project in order to exploit our complementarities and develop much synergy. For the Atlas project-team in particular, SARAVÁ provides the opportunity to make progress on workflow management which is the area of expertise of the Brazilian group and to gain access to real users in bioinformatics and oil industry with which they have good collaboration. For the French side in particular, SARAVÁ provides the opportunity to learn more on P2P in the context of the APPA project and thus exploit new research opportunities for P2P workflow management.

References

Abiteboul, S. and Polyzotis, N. The Data Ring: Community Content Sharing. Conference on Innovative Data Systems Research (CIDR), 154-163, 2007.

Agrawal, P., Benjelloun, O., Das Sarma, A., Hayworth, C., Nabar, S., Sugihara, T. and Widom, J. Trio: A System for Data, Uncertainty, and Lineage. Int. Conf. on Very Large Databases (VLDB), 1151-1154, 2006.

Akbarinia, R., Martins, V., Pacitti, E. and Valduriez, P. Design and Implementation of APPA. Global Data Management (Eds. R. Baldoni, G. Cortese, F. Davide), IOS Press, 2006.

Akbarinia, R., Pacitti, E. and Valduriez, P. Best Position Algorithms for Top-k Queries. Int. Conf. on Very Large Databases (VLDB), 495-506, 2007.

Furtado, C., Lima, A. ,Pacitti, E. ,Valduriez, P. ,Mattoso, M. Adaptive Hybrid Partitioning for OLAP Query Processing in a Database Cluster. Int. Journal of High Performance Computing and Networking, Special issue on Best Papers from SBAC2005, 5(4), p. 251-262, 2008.

Furtado, P. Node Partitioned Data Warehouses: Experimental Evidence and Improvements. Journal of Database Management 17(2), 42-60, 2006.

Lima, A. A. B., Furtado, C., Valduriez, P., Mattoso, M. Parallel OLAP Query Processing in Database Clusters with Data Replication. Distributed and Parallel Databases 25(1-2), 97-123, 2009.

Kotowski, N., Lima, A. A. B., Pacitti, E., Valduriez, P., Mattoso, M. Parallel query processing for OLAP in Grids. Concurrency and Computation. Practice & Experience 20(17), 2039 - 2048, 2008.

Meyer, L., Scheftner, D., Vöckler, J-S., Mattoso, M., Wilde, M. and Foster, I. An Opportunistic Algorithm for Scheduling Workflows on Grids. Int. Conf. on High Performance Computing for Computational Science (VecPar -2004), LNCS 3402, Springer, 1-12, 2005.

Özsu, T. and Valduriez, P. Principles of Distributed Database Systems. 2nd Edition, Prentice Hall, 1999 (3rd Edition forthcoming).

Pacitti, E., Valduriez, P. and Mattoso, M. Grid Data Management: open problems and new issues. Journal of Grid Computing, 5(3), 273-281, 2007.

Paes,M., Lima, A., Valduriez, P. and Mattoso, M. High-performance Query Processing of a Real-world OLAP Database with ParGRES. Int. Conf. on High Performance Computing for Computational Science (VecPar), Best student paper award for M. Paes, LNCS 5336, Springer, 188-200, 2008.

Serra da Cruz, S., Nogueira da Silva, F., Gadelha Jr., L., Reis Cavalcanti, M.C., Machado Campos, M.L., Mattoso, M. A Lightweight Middleware Monitor for Distributed Scientific Workflows. IEEE Int. Symp. on Cluster Computing and the Grid (CCGrid,), 693-698, 2008.

Valduriez, P., Pacitti, E. Data Management in Large-scale P2P Systems. Int. Conf. on High Performance Computing for Computational Science (VecPar -2004), Invited paper, LNCS 3402, Springer, 109-122, 2005.

Using Semantics in Peer Data Management Systems

Carlos Eduardo Pires[1], Damires Souza[1], Zoubida Kedad[2], Mokrane Bouzeghoub[2], Ana Carolina Salgado[1]

[1]Universidade Federal de Pernambuco (UFPE), Centro de Informática, Av. Prof. Luiz Freire, S/N 50.740-540 Recife, PE, Brazil

[2]Université de Versailles et Saint-Quentin-en-Yvelines (UVSQ), 45 Avenue des Etats-Unis, 78035 Versailles, France

{cesp,dysf,acs}@cin.ufpe.br, {zoubida.kedad, Mokrane.Bouzeghoub}@prism.uvsq.fr

Abstract. *Data management in Peer Data Management Systems (PDMS) is a challenging problem considering the excessive number of peers, their autonomous nature, and the heterogeneity of their schemas. To help matters, semantic knowledge in the form of ontologies has proven to be a helpful support for the techniques used for managing data in such systems. Ontologies can be used, for instance, to represent the semantic content of data sources as well as to unify the semantic relationships between their schemas. In this sense, the goal of this paper is to highlight the use of semantics in order to enhance data management issues in a PDMS. We present the current status of scientific cooperation between the database groups of Centro de Informática from Universidade Federal de Pernambuco (CIn/UFPE) and PRiSM laboratory from Université de Versailles Saint-Quentin-en-Yvelines (PRiSM/UVSQ). In addition, we point out further work to be done.*

1. Introduction

Peer Data Management Systems (PDMS) [Halevy *et al*., 2003; Valduriez and Pacitti, 2004; Mandreoli *et al*., 2007; Lodi *et al*., 2008; Kantere *et al*., 2009] came into the focus of research as a natural extension to distributed databases in the Peer-to-Peer (P2P) setting [Herschel and Heese, 2005]. They consist of a set of peers, each one with an associated schema (called exported schema) that represents the data to be shared with other peers. In such systems, schema matching techniques are used to establish schema mappings (i.e., correspondences between schema elements) which form the basis for query answering and peer clustering. Schema mappings are defined between pairs of semantic neighbor peers, i.e., peers that are semantically related as previously identified by a clustering process. Queries submitted at a peer are answered with data residing at that peer and with data that is reached through mappings over the semantic neighbors.

Data management in PDMS is a challenging problem considering the excessive number of peers, their autonomous nature, and the heterogeneity of their schemas. To help matters, semantic knowledge in the form of ontologies has proven to be a helpful support for the techniques used for managing data in such systems. For instance, ontologies can be used to represent the semantic content of data sources as well as to unify the semantic relationships between their schemas. Thus, the goal of this research project is to exploit the benefits provided by semantics through ontologies to enhance

data management issues in PDMS. To this end, we present semantic-based approaches to support peer clustering, schema summarization, schema matching, and query reformulation. In the following, we present a description of such approaches as well as some obtained experimental results.

2. Work in Progress

The main categories of problems which have been particularly addressed by this research are described in the following.

Ontology-based PDMS

The establishment of schema mappings and consequently query answering in PDMS can be improved if semantically similar peers are put together in the overlay network. In this sense, we have proposed a semantic-based PDMS [Pires *et al*., 2007] whose mixed network is mainly designed to assist the organization of peers according to their exported schema (represented by an ontology). Peers are grouped according to their knowledge domain (e.g., *Education* and *Health*), forming semantic communities. Inside a community, peers are organized in a finer grouping level, named semantic clusters, where peers share similar ontologies (schemas). A semantic cluster has a cluster ontology which represents the ontologies (schemas) of the peers within the cluster. Each cluster maintains a link to its semantic neighbors in the overlay network, i.e., to other semantically similar clusters. Regarding implementation, a PDMS simulator has been developed through which we were able to reproduce the main conditions characterizing the proposed system's environment.

Ontology Matching

We have proposed a semantic-based ontology matching process, named *SemMatch* [Pires *et al*., 2009a], that considers, besides the traditional terminological and structural matching techniques, a semantic-based one. The process produces a set of semantic correspondences and a global similarity measure between two peer ontologies. The former is used to enhance query reformulation while the latter is used, for instance, to determine semantic neighbor peers in the overlay network. A tool implementing the semantic-based ontology matching process has been developed.

Ontology Summarization

We have proposed an automatic process to build summaries of cluster ontologies [Pires *et al*., 2009b]. Such summaries are used as a semantic index to assist the identification of similar peers when a new peer joins the system. The summarization process is divided into several steps and is based on the notions of centrality and frequency. Centrality is used to capture the importance of a given concept within an ontology. The use of frequency is motivated by the fact that a cluster ontology is obtained by merging several different local ontologies. The summaries are used as a semantic index to indicate an initial cluster for new peers during their connection to the system. We have developed a preliminary implementation of an ontology summarization tool.

Ontology-based Peer Clustering

Peer connection in the proposed PDMS is mainly an incremental clustering process. When a new peer arrives, it searches for a corresponding semantic community in a DHT

network. Then, within a semantic community, the new peer searches for a semantically similar cluster in an unstructured network. The search for a cluster starts when the new peer sends its exported schema (i.e., an ontology) to a promising initial cluster (provided by the semantic index) and proceeds by following the semantic neighbors of the initial cluster until a certain limit (TTL) is reached. At each visited cluster, *SemMatch* is executed taking as arguments the current cluster ontology and the exported schema of the new peer. Each cluster returns its global similarity measure to the new peer. The set of global measures are used by the new peer to determine if it will join an existing cluster or create a new one. The proposed process has been implemented in the mentioned simulator and submitted to experimental evaluation. Validation has been performed using clustering indices and by executing query answering simulations.

Query Reformulation

In our PDMS, a query posed at a peer is routed to other peers in order to find answers to the query. An important step in this process is reformulating a query issued at a peer into a new query expressed in terms of a target peer, considering the correspondences between them. In this light, we have worked on a semantic-based query reformulation approach, named *SemRef* [Souza *et al*., 2009], which brings together both query enrichment and query reformulation techniques in order to provide users with a set of expanded answers. Exact and enriched query reformulations are produced as a means to obtain this set of answers. To this end, we make use of semantics which is mainly acquired from a set of semantic correspondences that extend the ones commonly found. Also, we take into account the context of the user, of the query and of the environment as a way to enhance the overall process and to deal with information that can only be acquired on the fly.

3. Further Work

There are a number of ongoing research issues concerned with the use of semantics in PDMS. Among them, we will focus in two relevant issues: (i) the maintenance of semantic communities; and (ii) query routing. Concerning the former, an issue to be studied in deep detail regards the evolution of cluster ontologies. In order to reflect the content available in a semantic cluster, cluster ontologies should be created and maintained dynamically, in an automatic way, according to peers' intermittence. A cluster ontology should be able to evolve not only when a requesting peer joins the cluster but also when a participating peer leaves it.

Regarding the latter, query reformulation strategies and query routing mechanisms [Montanelli and Castano, 2008] have a great influence on each other. In our approach, we consider that every peer P_i maintains a neighborhood $N(P_i)$ selected from the set of existing peers in the setting. In this sense, a submitted query must be reformulated in such a way that it is possible to ensure effective query routing, preserving the query semantics at the best possible level of approximation. Furthermore, we intend to use semantics to enhance the selection of relevant semantic neighbors and their ranking.

4. Cooperation Status

The scientific cooperation between the database groups of Centro de Informática from Universidade Federal de Pernambuco (CIn/UFPE) and PRiSM laboratory from

Université de Versailles Saint-Quentin en Yvelines (PRiSM/UVSQ) began in the nineties when two students from CIn/UFPE were accepted as PhD students in PRiSM/UVSQ supervised by Prof. Mokrane Bouzeghoub. This cooperation was intensified in 2002 when a PhD student from CIn/UFPE did a 'sandwich' stage in PRiSM/UVSQ. Since then it has been established a regular cooperation which has included research visits, cooperative projects, a sabbatical year of the CIn/UFPE database group leader, and another PhD 'sandwich' stage in PRiSM/UVSQ. Currently, we have a STIC/AMSUD project (2008-2009) which motivated the organization of a workshop in Recife last July with participation of researchers from the groups involved on the project. One of the main research areas of this cooperation project is the use of semantics to enhance data management in dynamic distributed environments.

References

Halevy, A. Y., Ives, Z. G., Mork, P., and Tatarinov, I. (2003) "Piazza: Data Management Infrastructure for Semantic Web Applications", In: World Wide Web Conference, pp. 556-567, Budapest, Hungary.

Valduriez, P., and Pacitti, E. (2004) "Data Management in Large-Scale P2P Systems", In: International Conference on High Performance Computing for Computational Science, pp. 104-118, Valencia, Spain.

Mandreoli, F., Martoglia, R., Penzo, W., Sassatelli, S., and Villani, G. (2007) "SUNRISE: Exploring PDMS Networks with Semantic Routing Indexes", In: 4th European Semantic Web Conference, Innsbruck, Austria.

Lodi, S., Penzo, W., Mandreoli, F., Martoglia, R., and Sassatelli, S. (2008) "Semantic Peer, Here are the Neighbors You Want!", In: 11th Extending Database Technology, pp. 26-37, Nantes, France.

Kantere, V., Tsoumakos, D., Sellis, T., and Roussopoulos, N. (2009) "GrouPeer: Dynamic clustering of P2P databases", In: Information Systems Journal, Volume 34, Issue 1, pp. 62-86.

Herschel, S., Heese, R. (2005) "Humboldt Discoverer: a Semantic P2P Index for PDMS", In: Int. Workshop Data Integration and the Semantic Web, Porto, Portugal.

Pires, C. E. S., Lóscio, B. F., and Salgado, A. C. (2007) "Semantic-based Connectivity in a Peer Data Management System", In: 6th Workshop of Thesis and Dissertation on Data Base, in conjunction with the 22th SBBD, pp. 65-72, João Pessoa, Brazil.

Pires, C. E. S., Souza, D., Pachêco, T., and Salgado, A. C. (2009a) "A Semantic-based Ontology Matching Process for PDMS", to appear in the 2nd International Conference on Data Management in Grid and P2P Systems (Globe'09), Linz, Austria.

Pires, C. E. S., Alencar, V. B., Kedad, Z., and Salgado, A. C. (2009b) "Building Ontology Summaries for PDMS", Submitted to the 28th Conference on Conceptual Modeling, Gramado, Brazil.

Souza D., Arruda T., Salgado A. C., Tedesco P., and Kedad, Z. (2009) "Using Semantics to Enhance Query Reformulation in Dynamic Environments", to appear in the 13th East European Conference on Advances in Databases and Information Systems (ADBIS'09), Riga, Latvia.

Montanelli S., Castano S. (2008) "Semantically Routing Queries in Peer-based Systems: the H-Link Approach", In: Knowledge Eng. Review 23(1): 51-72.

MinTI: Mineração de Texto e Imagem para Recuperação Inteligente de Documentos do CPDOC-FGV

Leonardo Silva Kury[1], Aristófanes Corrêa[1], Anselmo Paiva[1,] Asla M. Sá[2], Moacyr Silva[2], Paulo Cezar Carvalho[2].

[1]Universidade Federal do Maranhão (UFMA)
Núcleo de Computação Aplicada (NCA) - São Luís – MA – Brasil
[2]Fundação Getúlio Vargas (FGV)
Centro de Matemática Aplicada (CMA) - Rio de Janeiro – RJ – Brasil

leokury@gmail.com, ari@dee.ufma.br, paiva@deinf.ufma.br, {asla.sa, moacyr.silva, pcezar}@fgv.br

Abstract. *CPDOC-FGV (Centro de Pesquisa e Documentação de História Contemporânea do Brasil) presently hosts a large set of texts and images, totalizing 1,8 million documents, mostly digitalized. Traditional types of search are inefficient and moreover, the data base grows faster than the human capacity to categorize and analyse the information. In this context, CMA-FGV (Centro de Matemática Aplicada), in partnership with NCA-UFMA (Núcleo de Computação Aplicada), agreed to develop a software tool called MinTI, supported by recent advances in text and image mining. Our goal is to create a tool for intelligent information retrieval from CPDOC-FGV database. By automatizing the process of information extraction we hope to improve the potentialities of the database.*

Resumo. *O Centro de Pesquisa e Documentação de História Contemporânea do Brasil (CPDOC-FGV) tem atualmente conjuntos documentais que totalizam cerca de 1,8 milhões de documentos. As formas tradicionais de busca em acervo se tornam ineficientes à medida que a base de dados cresce numa velocidade que ultrapassa a capacidade humana de catalogação e análise da informação. Neste contexto, o Centro de Matemática Aplicada (CMA-FGV), em parceria com o Núcleo de Computação Aplicada (NCA-UFMA), estão pesquisando e desenvolvendo a ferramenta computacional MinTI, baseada em tecnologia recente desenvolvida nas áreas de mineração de textos e imagens, para recuperação inteligente de informações da base de dados do CPDOC-FGV, visando acelerar e automatizar o processo de extração de informações relevantes para melhorar o aproveitamento do potencial da base de dados.*

1. Introdução

O Centro de Pesquisa e Documentação de História Contemporânea do Brasil (CPDOC-FGV) tem como um de seus objetivos abrigar conjuntos documentais relevantes para a história recente do país. Os conjuntos documentais do CPDOC-FGV hoje totalizam cerca de 1,8 milhões de documentos. Com o passar dos anos, o CPDOC-FGV tem

armazenado cada vez mais informações (texto, áudio, imagem e vídeo) em suas bases de dados. A quantidade de informação armazenada aumenta diariamente e ultrapassa a habilidade técnica e a capacidade humana de interpretação dessa informação. Apesar do enorme valor desses dados, a sua organização atual não consegue aproveitar totalmente o material que está armazenado em sua base. A maior parte dessa informação encontra-se na forma textual descrita em linguagem natural e é atualmente recuperada utilizando o sistema de computação *Accessus*. Este sistema ainda não é dotado de recursos inteligentes de agrupamento e classificação de informações.

Nesse contexto, propusemos como colaboração do Centro de Matemática Aplicada (CMA-FGV), em parceria com o Núcleo de Computação Aplicada (NCA-UFMA), o desenvolvimento de uma ferramenta computacional baseada em mineração de textos e imagens para recuperação inteligente de informações da base de dados do CPDOC-FGV, visando a extração de informações relevantes para acelerar e automatizar o aproveitamento do potencial da base de dados.

Neste artigo, abordaremos brevemente o problema de Mineração de Textos na Seção 2. A Seção 3 mostra o protótipo inicial da ferramenta de mineração de texto e imagem, denominado *MinTI*, que está sendo desenvolvida em parceria pelo CMA-FGV e pelo NCA-UFMA. Na Seção 4 descreveremos o problema de mineração de imagens, no contexto do CPDOC-FGV, como trabalho futuro.

2. Mineração de Texto

A Mineração de Texto é o processo de obtenção de informações relevantes a partir de textos descritos em linguagem natural. Inspirado originalmente na mineração de dados, que consiste em extrair informação de banco de dados estruturados, difere desta pela natureza dos dados processados que são não-estruturados ou semi-estruturados (Hotho and Nürnberger, 2005). A Mineração de Textos combina técnicas de Banco de Dados, Inteligência Artificial, Aprendizado de Máquina e Processamento de Linguagem Natural entre outros. Alguns exemplos de aplicações da área de mineração de textos são: sumarização, classificação, agrupamento, tradução automática e extração de informações de textos.

O processo de mineração de textos é bastante complexo e pode ser subdividido nas seguintes etapas (Aranha, 2007):

- **Coleta** é a etapa inicial do processo em que é feita a aquisição dos dados. É de fundamental importância para o desenvolvimento que se obtenha dados de qualidade. Ao fim dessa etapa, espera-se que tenha formado uma base de dados textual, conhecido na literatura como *corpus*.
- No **pré-processamento**, os dados são preparados para as etapas seguintes com a finalidade de serem formatados e representados para o processamento das etapas seguintes. Dependendo dos dados, pode ser uma etapa demorada e consumir boa parte do cronograma do processo.
- A **indexação** é responsável por organizar os termos adquiridos de forma a facilitar o acesso e recuperação dos dados. Uma boa estrutura de indexação garante agilidade e rapidez ao processo de mineração.

- Na etapa de **mineração** são executados algoritmos, cálculos estatísticos e inferências com o objetivo de extrair automaticamente informações relevantes da base de dados.
- Finalmente, a **análise** dos dados é realizada para interpretação e visualização dos resultados obtidos.

No presente contexto, efetuaremos as etapas de coleta, pré-processamento e indexação sobre a base de dados do CPDOC-FGV. O foco de pesquisa deste trabalho está no estudo comparativo e aplicação de algoritmos de aprendizado de máquina e de outras áreas relacionadas durante a etapa de mineração.

3. *MinTI*

O protótipo inicial por nós desenvolvido, chamado *MinTI*, implementa nesta fase somente o módulo de mineração de texto. A fase atual o protótipo aceita dois tipos de entrada de texto: texto inserido diretamente na caixa de texto do programa, ou texto obtido a partir de um arquivo especificado pelo usuário (ver Figura 1).

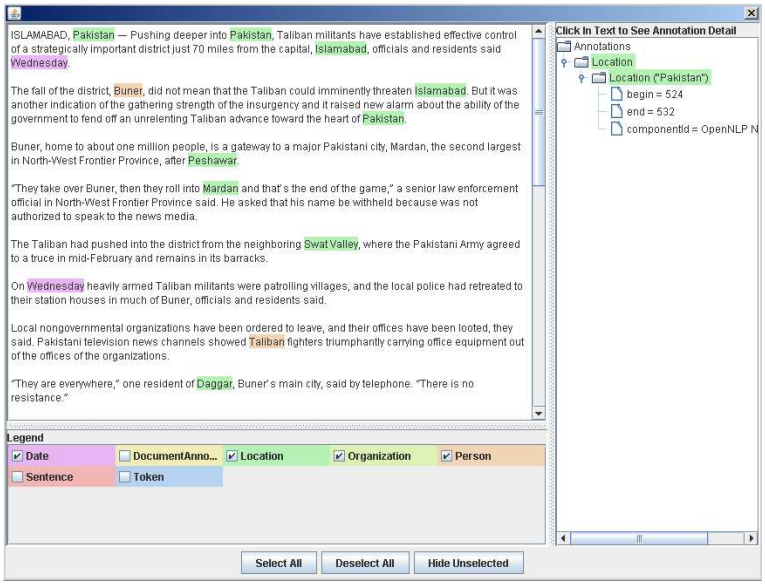


Figura 1 - Tela do *MinTI* exibindo o resultado da tarefa de NER para um artigo do *The New York Times*

O *MinTI* é baseado na arquitetura do padrão UIMA (*Unstructured Information Management Architecture*) da OASIS (*Organization for the Advancement of Structured Information Standards*). A implementação desta arquitetura utiliza o Apache UIMA - *framework* de código aberto desenvolvido em linguagem JAVA - que suporta o idioma português e o inglês. Para o idioma português o protótipo executa a tarefa de etiquetação de classes gramaticais (*Part-of-Speech Tagging*, ou simplesmente POS Tagging). Esta tarefa resume-se a atribuir uma classe gramatical da linguagem para cada *token* do texto. A tarefa executada no idioma inglês é o reconhecimento de entidades nomeadas (*Named Entity Recognition*, ou NER). Essa é uma subtarefa da extração de informações que visa localizar e classificar elementos atômicos do texto em categorias pré-definidas como nome de pessoas, organizações, locais, datas, etc. Essa diferenciação entre as tarefas dos

idiomas deve-se a falta de *corpus* anotado em português para o treinamento dos algoritmos.

A próxima fase de desenvolvimento do *MinTI* implementará a etapa de mineração propriamente dita. Em seguida passaremos a abordar a mineração de imagens.

4. Trabalhos Futuros: Mineração de Imagem

Um problema comum enfrentado pelos grandes arquivos de dados é a concentração de conhecimento nas pessoas que manipulam e/ou alimentam a base de dados. Uma vez que essa pessoa não está mais disponível perde-se a *expertise* de busca às informações arquivadas. No arquivamento de imagens de personalidades históricas este fato se torna especialmente crítico uma vez que uma determinada pessoa se "familiariza" aos personagens de um dado período e passa a reconhecê-los em fotografias ainda não catalogadas, tarefa que requer a habilidade particular deste indivíduo que outra pessoa não familiarizada não seria capaz de executar.

Para o CPDOC-FGV um procedimento de reconhecimento automático de personalidades em bancos de fotografias e vídeos seria de grande utilidade. O reconhecimento de face a partir de imagens fotográficas e imagens de vídeo está emergindo como uma atividade na área de pesquisa com numerosas aplicações comerciais. Estas aplicações requerem algoritmos robustos para reconhecimento de faces humanas sob diferentes condições de iluminação, expressões faciais e orientações. A mineração de imagens agrupa novos conceitos e tecnologias que englobam as áreas de processamento de imagens e vídeos e conceitos de aprendizado de máquinas. A mineração de imagens ocorre sobre os dados extraídos da imagem, essas características são utilizadas por modelos de mineração.

Referências

A Hotho, A Nürnberger (2005) A brief survey of text mining, LDV Forum-GLDV Journal for Computational Linguistics and Language Technology

CD Manning, P Raghavan, H Schütze (2008) Introduction to Information Retrieval, Cambridge University Press

MW Berry (2004) Survey of Text Mining: Clustering, Classification, and Retrieval, Springer

MW Berry, M Castellanos (2007) Survey of Text Mining II: Clustering, Classification, and Retrieval, Springer

CPDOC. Centro de Pesquisa e Documentação de História Contemporânea do Brasil, Fundação Getúlio Vargas. Sítio: http://www.cpdoc.fgv.br

ARANHA, C.N. (2007) Uma abordagem de pré-processamento automático para mineração de textos em português: sob o enfoque da inteligência computacional, Tese de Doutorado, Departamento de Engenharia Elétrica, PUC-Rio

An Approach for Ontology Modularization

Camila Bezerra [1], Fred Freitas[1], Jérôme Euzenat[2], Antoine Zimmermann [3]

[1]Centro de Informática - Universidade Federal de Pernambuco
Av. Prof. Luis Freire, s/n, Cidade Universitária - 50740-540
Recife-PE - Brasil

[2]INRIA Grenoble - Rhône-Alpes Inovallée,
655 avenue de l'Europe, Montbonnot 38 334
Saint Ismier, Cedex, France

[3]Digital Enterprise Research Institute, National University of
Ireland, Galway IDA Business Park, Lower Dangan, Galway, Ireland

{cbs,fred.freitas}@cin.ufpe.br

{Jerome.Euzenat, antoine.zimmermann}@inrialpes.fr

antoine.zimmermann@deri.org

Abstract. *Ontology modularization could help overcome the problem of defining a fragment of an existing ontology to be reused, in order to enable ontology developers to include only those concepts and relations that are relevant for the application they are modeling an ontology for. This paper presents a concrete tool that incorporates an approach to ontology modularization that inherits some of the main principles from object-oriented software engineering, which are encapsulation and information hiding. What motivated us to track that direction is the fact that most ontology approaches to the problem focus on linking ontologies rather than building modules that can encapsulate foreign parts of ontologies(or other modules) that can be managed more easily.*

1. Introduction

Although the field of ontology engineering would take advantage of reuse, a culture of building blocks for ontologies – such as the component culture that took over in the realm of object-oriented software engineering – has not shown up yet. This constitutes an absence that should not be neglected, once the realization of the Semantic Web depends on the ability to reuse ontologies [Stuckenschmidt and Klein 2004].

Currently, there are two major ways of reusing an ontology. Ontology editors such as Protégé[1] allow the reuse of another ontology by including it in the model that is being designed. The Web ontology language [McGuinness and van Harmelen 2004] offers the possibility to import an OWL ontology by means of the owl:imports statement. In both cases, the whole ontology has to be included. Ontology modularization could help overcome the problem of defining a fragment of an existing ontology to be reused, in order to enable ontology developers to include only those concepts and relations that are relevant for the application they are modeling an ontology for.

[1]http://protege.stanford.edu/

This paper presents a concrete tool that incorporates an approach to ontology modularization that inherits principles of object-oriented software engineering. What motivated us to track that direction is the fact that most ontology approaches to the problem focus on linking ontologies (or modules) rather than building modules that can encapsulate foreign parts of ontologies (or other modules) that can be managed more easily.

2. Ontology Modularization

The notion of modularization comes from Software Engineering where it refers to develop software in structured way that supports the combination of self-contained components that are easier to build, reuse and maintain.

From an ontology engineering perspective, modularization should be considered as a way to structure ontologies, meaning that the construction of a large ontology should be based on the combination of self-contained, independent and reusable knowledge components[d'Aquin et al. 2007]. A possible definition for an ontology module is the following: "An ontology module is a reusable component of a larger or more complex ontology, which is self-contained but bears a definite relationship to other ontology modules" [P.Doran 2006]. Alan Rector adds to that definition by specifying three requirements that should hold for ontology modules [Blomqvist 2004]: loose coupling, which means the modules might have very little in common, and therefore as little interaction as possible should be required between the modules; self-containment, which means that every module should be able to exist and function without any other module; integrity, which means that there should be ways to check for, and adapt to, changes in order to ensure correctness. In the next section, we describe our approach to modularization of ontologies and our modularization tool, ModOnto.

3. The Proposed Ontology Modularization Approach

The goal of our ongoing research is to define and implement an ontology modularization approach that provides the definition and composition of modules(or ontologies). The approach was inspired by the object-oriented components market, for it is indeed the most successful reuse approach in the computer science mainstream. The properties typically expected from a module system are: separate development, separate execution, separate compilation and reusability.

Bearing this in mind, we defined a module language which comprehends some aspects related to an organized way of implementing module composition with the above principles. On the syntactic side, a module language must be able to encapsulate ontology fragments, to refer to other modules and to define the interface between these modules. The module tags that compose the language syntax to define modules are displayed in Figure 1 below.

From this syntax, every feature represents a block of information of the module.

- The Uses Feature is the list of external modules or ontologies being used to build the new one.
- Imports are the list of entities (classes, properties and instances) from other ontologies or modules that are needed to create the module.
- Contains represents the new entities and axioms created using or not imported entities.

Features	Description
Uses	List of used Ontologies /Modules
Imports	List of imported entities
Contains	New entities and axioms
Alignments	List of alignments between the module and the used Ontologies/Modules
Exports	List of entities can be reused for others modules

Figure 1. Module Language tags and descriptions.

- The alignments feature is needed when working with different modules and mappings are needed involving exported entities.
- The exports are entities from the content or from the list of the imported entities that will be available when one reuses the module being defined, i.e. , the classes, properties, axioms and individuals ready for reuse.

On the semantic side, it is necessary to define unambiguously what is the meaning of "encapsulating" is, i.e., what is hidden and what is exposed in a module. This definition leads to a way to determine which assertions are logical consequences of the defined modules. For that purpose we have provided our ontology modularization approach with very general definitions for the semantics, and then show that different actual semantics can be applied to our module language, each having advantages and drawbacks. The disadvantage of most distributed semantics is that knowledge does not fully propagate from imported modules to importing modules. It can be an advantage too, with regard to robustness towards heterogeneity. Besides this flexibility, our semantics' definitions possess another interesting feature in the context of ontology modules: It does not depend on a particular ontology semantics, but only needs to be able to determine what are the logical consequences of an ontology. This is very useful when using modules in which the ontology languages can be different and also enables connected ontologies or modules to be replaced by similar ones.

Beyond the design of the language with its syntax and semantics, a software tool is needed for modularization to be accepted at a larger scale, thus avoiding burden over ontology developers. A Tool, called ModOnto (see figure2), has been developed as a standalone tool, providing an intuitive graphical interface for modularization that facilitates the selection of ontology entities for building modules.

Module API: based on the implementation of the module specifications as described in the syntax. Provide the basic data structure and accessors as well as a parser and serializer on top of the structure.

Module extractor: consist on the part of the tool that provide a graphical interface to extract elements from the imported ontology. A user of this tool should be able to make a 'select' over an ontology in a number ways, such as selecting/ruling out entities (classes, roles, individuals). It is possible to import items from more than one ontology.

Module checker: the checker should verify if all of the definitions needed by one module are present on its imported modules or in their own imported modules.

Module library: building on the Module API, the Module library is a repository for "off-the-shel" modules that will help developers to choose its proper modules.

Module linker: is possible to import entities from other other modules. Could be to interact with the module library to pick up imported modules to be included,

Reasoner interface for modules: modular ontologies are ontologies. So they should be usable just as yet another ontology: being able to query if an assertion is entailed by a particular module.

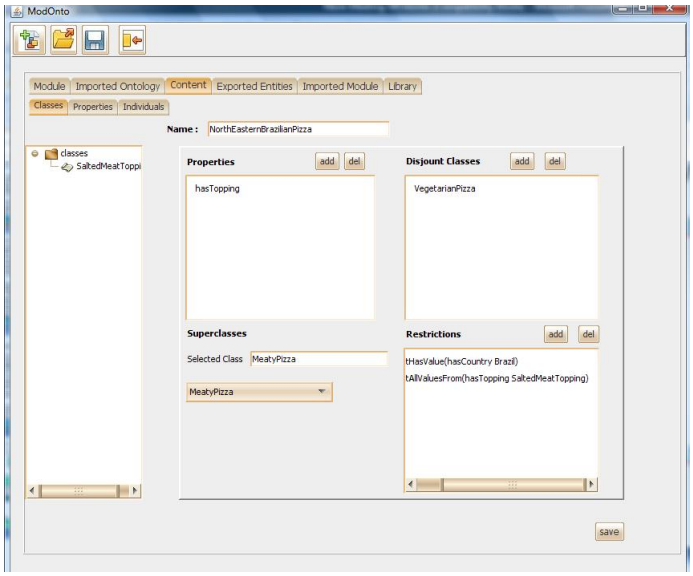


Figure 2. Create a new class in ModOnto

The Java programming language, the OWL API from Manchester, which contains an implementation for the W3C Web Ontology Language OWL, were used for the tool, as well as the alignment API[Euzenat 2004] and a reasoner based on IDDL[Zimmermann and Duc 2008] developed in INRIA Rhone-Alpes.

4. Related Work

During the last three years, there was an increasing interest and active research about ontology modularization. Some approaches propose translation of so called modular languages (DDL,e-connection, etc.) into standard DL and describe new reasoning services to ensure modularity, without introducing new formal languages. This can be satisfactory from a logical point of view but, is does not take into account engineering aspects like encapsulation, separation of ontologies from mappings, etc. For example, P-DL([Bao et al. 2006]) was designed in order to improve modularization of web ontologies. In P-DL, an ontology is divided into several packages (i.e., a module in their sense), which are interpreted separately, but the interpretations have to coincide on imported terms.

There are two different approaches have been considered for the modularization of existing ontologies: ontology partitioning, which divides an ontology into a set of modules, and module extraction, which reduces an ontology to a module focusing on a given set of elements. For example, Doran([P.Doran 2006]) proposed an approach which has been implemented as a standalone application, called ModTool which has an intuitive graphical interface. ModTool recursively applies a set of rules to generate an ontology module and define an abstract graph model and an algorithm for module extraction.

Regarding tool support for ontology modularization, for the goals of creating and encapsulating ontology modules (among other operations), there is no framework available yet, what could assure the originality of our work. However, at least one implementation for an ontology linkage approach deserves attention. The Manchester's OWL-API([Bechhofer et al. 2003]), the SWOOP Web Ontology browser and editor and the Pellet OWL reasoner have been extended to comply with E-connections ontologies. Some good features have been added to SWOOP with this extension, like incorrect editing prevention and foreign ontology visualization with graph layouts.

5. Conclusions and Future Work

Along the OntoCompo project, we have designed an ontology modularization approach inspired by one of the most successful techniques of computer science, object orientation. We developed a module language with syntaxes and a formally well-defined semantics and we present a concrete implementation of the language, the ModOnto tool that assists users in the construction of new modules starting from ontologies or other modules. This project anticipates an ontology module market that can possibly come out as a consequence of the Semantic Web, in the same flavor as the Java components market in the Web, when module repositories become available.

As we just produced the first prototype of the modularization tool, it naturally will be extended and improved in many ways. The suite must be integrated with popular ontology frameworks like Protégé [2], KAON2[3] and the ones from NeOn project[4] too.

6. Acknowledgements

The authors wish to thank CNPQ (National Council of Technological and Scientific Development) and INRIA (The French National Institute for Research in Computer Science and Control) for its partial sponsorship through the research project Composition and Modules for Ontology Engineering.

References

Bao, J., Caragea, D., and Honavar, V. (2006). Towards collaborative environments for ontology construction and sharing. In *CTS '06: Proceedings of the International Symposium on Collaborative Technologies and Systems*, pages 99–108, Washington, DC, USA. IEEE Computer Society.

Bechhofer, S., Volz, R., and Lord, P. (2003). Cooking the semantic web with the owl api. pages 659–675. Springer.

Blomqvist (2004). State of the art: Patterns in ontology engineering. Technical report, Jönköping University.

d'Aquin, M., Schlicht, A., Stuckenschmidt, H., and Sabou, M. (2007). Ontology modularization for knowledge selection: Experiments and evaluations. In *DEXA*, pages 874–883.

[2]http://protege.stanford.edu/

[3]http://kaon2.semanticweb.org/

[4]http://www.neon-project.org/

Euzenat, J. (2004). An api for ontology alignment. In *Proc. 3rd conference on international semantic web conference (ISWC), Hiroshima (JP)*, pages 698–712.

McGuinness, D. L. and van Harmelen, F. (2004). Owl. http://www.w3.org/TR/owl-guide/m.

P.Doran (2006). Ontology reuse via ontology modularisation. In *Proceedings of KnowledgeWeb PhD, Budva, Montenegro*.

Stuckenschmidt, H. and Klein, M. C. A. (2004). Structure-based partitioning of large concept hierarchies. In *International Semantic Web Conference*.

Zimmermann, A. and Duc, C. L. (2008). Reasoning on a Network of Aligned Ontologies. In Calvanese, D. and Lausen, G., editors, *Web Reasoning and Rule Systems, Second International Conference, RR 2008, Karlsruhe, Germany, October/November 2008, Proceedings*, volume 5341 of *Lecture Notes in Computer Science*, pages 43–57. Springer.

Enhancing Affective Communication in Embodied Conversational Agents

Michelle D. Leonhardt[1,2], **Rosa M. Vicari**[1], **Sylvie Pesty**[2]

[1]Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

[2]Laboratoire D'Informatique de Grenoble (LIG)
110 Av. de la Chimie – 38041 – Grenoble cedex 9
Domaine Universitaire de Saint Martin d'Heres – France

{mdleonhardt,rosa}@inf.ufrgs.br, sylvie.pesty@imag.fr

Abstract. *In this article, we present current work being developed in collaboration between UFRGS (Brazil) and LIG laboratory (France) inserted in the scope of PRAIA Project - international project of cooperation between UFRGS and LIG. We first motivate the use of Embodied Conversational Agents (ECAs) by showing and contextualizing research, in order to establish a foundation for understanding how these agents may be useful in human-computer interaction. We then give more details about the agent, its purpose and scenario, together with important characteristics of its development. Finally we conclude with more information about the cooperation and also demonstrating the importance of the research in this field.*

1. Introduction

One of the objectives of AI (Artificial Intelligence) is to provide formal tools for the development of computerized systems which express human beings' intellectual behavior when performing a certain activity [Russel and Norvig 1995]. Since early times, human beings have tried to develop natural communication among interlocutors. As technology has developed and consequently the access to personal computers and to the Internet, such concern has also focused on machines. Interface between machines and humans is getting more complex and better planned, moving towards more human-like ways of communication.

Embodied Conversational Agents (ECAs) can be defined as computer-generated characters that are able to demonstrate many of the same properties as humans in face-to-face conversation, including the ability to produce and respond to verbal and nonverbal communication [Cassell et al. 2000]. In order to emulate the experience of human face-to-face conversation, several verbal and nonverbal modalities of communication can be used: speech, intonation, gaze, facial display, gesture, among others. Because of the importance of nonverbal issues, ECAs must also be conversational in their behaviors, and human-like in the way they use their bodies in conversation, aiming to empower, facilitate, and enrich interaction between humans and machines [Bickmore 1999].

The goal of researchers in the field of ECAs is to create agents that can be more natural, believable and easy to use, including, in this expectation, several other specific

goals like: enrich conversational capabilities, deliberate in order to provide adequate feedback in different modalities of communication, consider social intelligence issues while developing conversational agents (trust, persuasion, personality...), use emotion as a form of improving deliberation, expression, and conversation [Cassell et al. 2000]. Due to the broad scope of research and the multidisciplinary of the field, many other investigations can arise in many different areas, leading researchers with the problem of choosing among different technologies and approaches, together with developing reasoning mechanisms to achieve the main goal.

Following the motivation presented above, the objective of this work is to present an ECA capable of communicating affectively with users. Actually we are strongly interested in an ECA capable of providing affective feedback considering an expressive communication language (briefly discussed in next section) and emotion as a tool for reasoning and behaving.

2. The agent

The role of our agent is to guide its own construction, as well as participate on it. We are developing a scenario where two participants must collaborate in order to define not only physical characteristics of the ECA, but also emotional ones (based on personality behaviors).

Initially, the role of the agent is to guide the user while deciding the appearance and personality of an agent for use as part of other systems (there are many possibilities in different applications: representing the student or the teacher inside educational environments, representing characters in games, interface web agents...). As far as the conversation goes and participants start to decide on aspects of the agent personality, the agent will go deep in conversation, reacting according to its characteristics (by means of language expression and other multimodalities of communication - i.e - if participants decide the agent will have soft personality, the conversation will be conducted accordingly). The idea behind the strong focus on emotion in our agent can be justified by the fact that emotions represent one important modality when communicating a message and latest scientific findings indicate that the use of emotions in ECAs may contribute to various domains of application. Also, the use of emotions in such agents can contribute to their credibility.

Emotional display can be a very complex phenomenon involving a wide range of verbal and nonverbal behavior, making the integration between different modalities of communication and textual exchange of messages (together with reasoning mechanisms that can take into consideration the beliefs and goals of the agent in specific environments) an interesting focus of research. By observing and participating in its own construction, our agent will consider the conditions of success and satisfaction present in each message exchanged. In this context, attitudes can affect their way of speaking and acting.

In order to achieve our goals, our agent is being developed considering a set of conversation acts formally defined in order to allow expressive communication between communities of mixed agents [Berger and Pesty 2005a, Berger and Pesty 2005b]. Mixed communities are those where agents and humans interact together. The idea behind the language is to allow expressive communication between not only software agents but also human agents. This language takes into consideration aspects such expectations, condi-

tions of success, among other characteristics that are present in human communication. The conditions of success and satisfaction are explicitly defined as well as the elements from the conversational background. The thirty two formalized conversation acts are:

- assertive $\rightarrow$ confirm, deny, think,say, remember, inform and contradict;
- commissives $\rightarrow$ commit oneself, promise, guarantee, accept, refuse, renounce and give; directives $\rightarrow$ request, ask a question, suggest, advise, require, command and forbid;
- declaratives $\rightarrow$ declare, approve, withdraw, cancel;
- expressives $\rightarrow$ thank, apologize, congratulate, compliment, complain, protest, greet.

The architecture of our agent will consist of three different modules: the understanding module, the processing module, and the generation module. The understanding module is responsible for understanding and treating the textual input received. It consists of two different phases: the input analysis and the act association. The input analysis is the process of verifying the textual input as a preparation for an act association. The act association is responsible for narrowing the analysis, choosing, through the use of knowledge base information, the different acts that can fit the input. The processing module is responsible for choosing and deciding the actions of the agent. The deliberation of the agent is based on the BDI approach [Rao and Georgeff 1991]. In this module, the action planner is responsible for generating the actions of the agent, considering the knowledge of the scenario, the point in conversation and agent emotional characteristics. Finally, The generation module is responsible for taking the actions generated by the response planner and adapt them to be expressed by the virtual agent.

3. Conclusion

We presented in this paper current research towards the development of an affective ECA. The importance of research in ECAs is evident. In recent years, many papers on the topic and related aspects can be found in leading AI scientific conferences like AAMAS and IJCAI (highlighting the workshops and special tracks focusing on the subject: special track on Virtual agents and Workshop on AI for Human Computing, respectively from AAMAS 2008 and IJCAI 2007). Another important conference on the topic is the International Conference on Intelligent Virtual Agents (IVA). IVA is the major annual meeting of the intelligent virtual agents community, attracting interdisciplinary minded researchers and practitioners from embodied cognitive modeling, artificial intelligence, computer graphics, animation, virtual worlds, games, natural language processing, and human-computer interaction.

This work is inserted in the Artificial Intelligence group of the Institute of Informatics - UFRGS (Brazil), and the research this group carries out on AI and its applications: intelligent tutoring systems, education, agents, among others. Additionally, it is being developed in collaboration with MAGMA team (Modélisation d'agents autonomes en univers multi-agents) at LIG laboratory (France). MAGMA team develops theoretical studies, computer tools and practical applications for the user in the field of MAS (multi-agent systems). This work is also inserted in the scope of PRAIA Project (Pedagogical Rational and Affective Intelligent Agents) - international project of cooperation between UFRGS and LIG, supported by Capes-Cofecub [Jaques et al. 2009]. The main goal of

the project is to develop methodologies, models, tools and solutions for handling student affect in the interaction between tutor and student.

Inside PRAIA project, a platform was defined in order to test and validate the research developed inside the scope of the project. It consists in a collaborative game, called "Collaborative Sudoku", and is a multi-user version of the popular logic-based number placement puzzle, which requires basically simple spatial reasoning. In the game, a team collaborates through a web-based interface. Supported by a game server, the partners interact, negotiating and coordinating actions in order to construct a shared solution to each proposed reasoning problem. The main goal of each team is to complete the task faster than an adversary team, matched by the server at random.

Although we are (initially) not focusing on this platform, the agent can be easily integrated inside it in the future, with the adaptation of the dialogue plan and knowledge base. The reason why this platform is not being used as the scenario for the development of the agent is that the conversation possibilities would be limited for test (helping users to solve the game by giving hints or stimulating users to perform tasks by sending supportive messages) and the agent would have a passive role in the environment.

International cooperation provides exchange of expert knowledge and collaboration of efforts in order to consolidate and deeply explore common interests of research. Some joint publications related to this work are [Leonhardt et al. 2008a, Leonhardt et al. 2008b].

References

Berger, A. and Pesty, S. (2005a). Towards a conversational language for artificial agents in mixed community. In *Proceedings of Fourth International Central and Eastern European Conference on Multi-Agent Systems CEEMAS'05*, pages 31–40. Springer Verlag, Budapest.

Berger, A. and Pesty, S. (2005b). Vers un langage de conversation entre agents pour l'interaction dans les communautés mixtes. In *Actes du Colloque Jeunes Chercheurs en Sciences Cognitives*, Bordeaux, France.

Bickmore, T. W. (1999). Social intelligence in conversational computer agents. Proseminar conceptual analysis of thesis area, Massachussets Institute of Technology, add:cambridge.

Cassell, J., Sullivan, J., Prevost, S., and Churchill, E. (2000). *Embodied Conversational Agents*. MIT Press, Cambridge, Massachusets.

Jaques, P., Pontarolo, E., Vicari, R., and Pesty., S. (2009). Project praia: Pedagogical rational and affective intelligent agents. Submmited to COLIBRI Coloquium 2009.

Leonhardt, M. D., Pesty, S., and Vicari, R. M. (2008a). Towards expressive communication in embodied conversational agents. In *Proceedings of Speech and Face-to-Face communication Workshop*, pages 49–51, Grenoble, France.

Leonhardt, M. D., Pontarolo, E., Jaques, P., Pesty, S., and Vicari, R. M. (2008b). Towards an affective embodied conversational agent for collaborative educational environments. In *Proceedings of Affective Aspects of Cooperative Interactions Workshop*

- *8th International Conference on the Design of Cooperative Systems*, pages 7–15, Carry-le-Rouet, France.

Rao, A. S. and Georgeff, M. P. (1991). Modeling rational agents within a bdi-architecture. In *Proceedings of International Conference on Principles of Knowledge Representation and Reasoning*, San Francisco, CA, USA. Morgan Kaufmann Publishers.

Russel, S. and Norvig, P. (1995). *Artificial Intelligence: a Modern Approach*. Prentice-Hall, Englewood Cliffs.

PROJECT PRAIA
PEDAGOGICAL RATIONAL AND AFFECTIVE INTELLIGENT AGENTS

Patrícia Jaques[3], Edilson Pontarolo[4], Magda Bercht[1], Rosa Vicari[1], Sylvie Pesty[2]

[1]PGIE/Universidade Federal do Rio Grande do Sul (UFGRS)
[2]Laboratoire D'Informatique de Grenoble (LIG)
[3]PIPCA/Universidade do Vale do Rio dos Sinos (UNISINOS)
[4]Universidade Tecnológica Federal do Paraná (UTFPR)

pjaques@unisinos.br, epontarolo@utfpr.edu.br, bercht@inf.ufrgs.br
rosa@inf.ufrgs.br, sylvie.pesty@imag.fr

Abstract. *This paper describes the ongoing research of the project Capes/Cofecub PRAIA. The project aims at studying how to handle users' emotions in intelligent learning environments. A collaborative reasoning game is proposed as platform to test and validate the developed research. In this game, we use a psychologically grounded model to infer peer-related emotions that will be used by an Embodied Conversational Agent to improve collaboration between students.*

1. Introduction

This article aims at describing some partial results from the project PRAIA: a cooperation project between researchers from UFRGS (in Brazil) and LIG and LIMSI (in France). PRAIA (http://gia.inf.ufrgs.br/praia/) is an interdisciplinary project involving three main areas of research: Education, Computer Science and Cognitive Science. Education and Computer Science because we are interested in developing computational solutions for a more effective learning, and Cognitive Science because we aim at handling the emotions in intelligent learning environments.

The field of the Computer Science that studies the potential use of computing resources in learning is called Computer in Education. Some researchers of Artificial Intelligence (AI) interested in Education study the possibility of using techniques of AI in order to turn the educational software more customized to the user. Pedagogical theories and sophisticated techniques of user modeling have been investigated.

On the other hand, little attention has been paid to the role of the affect in cognition. However, some recent works of psychologists and neurologists have pointed out the important role of the emotions in some cognitive activities such as, for example, decision taking [Damásio 1994] and learning [Goleman 1995].

Thus, researchers of Artificial Intelligence have considered the emotions in intelligent systems modeling, appearing thus a new field of research in AI: "Affective Computing" (AC). Picard [Picard 1997] defines AC as "computing that relates to, arises from or deliberately influences emotions". This area is divided in two major branches of research interest. The first one studies mechanisms to recognize human emotions or to express emotions by machine in human-computer interaction. The second branch investigates the simulation of emotion in machines (emotion synthesis) in order to discover more about human emotions and to construct more realistic robots.

COLIBRI 2009 ● 22-23 de julho de 2009 ● ISBN 978-85-7669-245-4

In the first one, the researchers are mainly interested in recognizing the users' affect and to respond emotionally to them with the goal of adapting the system to the users, in this case, to their emotions. The systems that are part of this branch have the capacity of *recognizing* and *expressing* emotions. The proposed project is mainly situated in this first branch of research in Affective Computing.

Aiming at contributing to the works in affective computing applied to education, the project PRAIA has as main goal to develop models, methodologies, techniques and tools that handle the users' emotions in intelligent learning environments. A platform was defined in order to test and validate the research developed in this project. It consists in a collaborative game, called "Collaborative Sudoku", and is a multi-user version of the popular logic-based number placement puzzle, which requires basically simple spatial reasoning. We are employing this game to study empirically what happens in terms of peer-related emotions when this kind of simple reasoning task is addressed collaboratively. In the game, a team collaborates through a web-based interface. Supported by a game server, the partners interact, negotiating and coordinating actions in order to construct a shared solution to each proposed reasoning problem. They seek to complete the task faster than an adversary team, matched by the server at random.

We also have developed an affective model to recognize the specific emotions students feel towards their peers during collaboration. Our modeling follows an approach based on cognitive psychology, i.e. the students' emotions are inferred based on their interactions with the system's interface and the produced outcomes, evaluated according to cognitive criteria. More specifically, the model is based on the OCC psychological model of emotions [Ortony *et al.* 1988]. The model is focused on inferring four OCC "attribution" emotions: pride, shame, admiration and reproach.

In order to develop different affective models applied to education, we are also working in affective states like mood [Longhi *et al.* 2008], adopting the researches developed by Scherer [Scherer, 2001].

The information about users' emotions is used by an Embodied Conversational Agent (ECA), which will be integrated to this platform. ECAs are intelligent agents with a humanlike representation that are able to engage in a conversation with humans [Cassel and Sullivan 2000]. In a learning environment, recognizing the students' emotions can increase the believability of an ECA by making possible to maintain a more credible dialog with the students [Lester and Stone 1997]. As a consequence, the ECA will have a more effective base to support collaboration.

The first part of our work was dedicated to the development of an emotions inference model applied to the context of collaborative learning environments, which is described in more detail in the next sections. This affective model is the result of Edilson Pontarolo thesis supervised by Rosa Vicari and Patricia Jaques, and with a sandwich with Syvie Pesty. The ECA is being developed in the context of the PhD work of Michelle Leonhardt. The latter will be explored in another paper submitted to this same colloquium [Leonhardt *et al.* 2009].

2. Cognitive-based Inference of Emotions in Computing Systems

In order for a computing system to recognize users' emotions, two main steps are necessary. First, the system should be able to catch some users' data from which it

could infer their emotions. Nowadays, emotional states can be recognised by the following source of data: (*i*) voice (prosody); (*ii*) facial expressions; (*iii*) physiological signs (blood volume pulse, muscle tension, skin conductivity, breathing), and (*iv*) behavioural data, i. e. user's actions and choices in the system interface (for example, chosen options and typing speed). Second, the system should have a user model that is able to infer her/his emotional states from past emotions and also from the caught data. The capacity of the computational system to model the user's affective states is known as Affective User Modeling (AUM) [Elliott *et al.* 1999]. The affective user model must be dynamic enough to consider the changes in emotional states, since emotion is an ephemeral process.

In a well-defined context (for example, a learning environment), the users' **behavioral data** may be a path to predict, recognize and interpret their affective states. In this case, the system is conceptually grounded on a psychological model of emotions that follows a cognitive approach in order to infer the users' emotions from their actions and choices in the computing system. This approach was denominated **Cognitive-Based Affective User Modeling (CB-AUM)** [Martinho *et al.* 2000]. One of the most employed cognitive-based psychological models for inference of emotions in computing environments is Ortony, Clore and Collins Cognitive Structure of Emotions [Ortony *et al.* 1988], generally called the OCC model. The OCC model aims at explaining the appraisal, the cognitive process that elicits an emotion, of 22 types of emotions.

In the next section, we describe the proposed model, which is grounded on the OCC model for the inference of emotions.

3. The Peer-related Affective Student Model

We have developed an affective model to recognize the specific emotions students feel towards their peers during collaboration. Our modeling follows an approach based on cognitive psychology, i.e. the students' emotions are inferred based on their interactions with the system's interface and the produced outcomes, evaluated according to cognitive criteria. More specifically, the model is based on the OCC psychological model of emotions [Ortony *et al.* 1988]. The model is focused on inferring four OCC "attribution" emotions: pride, shame, admiration and reproach.

These peer-related emotions are inferred by representing the student's cognitive appraisal of her/his own collaborative interactions (giving rise to pride or shame), as well as the appraisal of her/his partner's collaborative interactions (giving rise to admiration or reproach). The praiseworthiness of collaborative interactions is evaluated according to student's goals-related standards, i.e. her/his assumed prototypical models of behavior. In both situations, the model assumes that the student applies the same set of standards, based in the fact that people tend to believe that other people have the same knowledge as theirs.

Some examples of collaborative interactions are "make a justified (or an unjustified) proposition", "agree (or disagree) with a proposition", and "send a positive/encouraging (or a negative/discouraging) message". Some examples of behavioral standards are "the student prefers a quite negotiated solution to a more individualist one", and "the student believes that it is important to keep the partner motivated to solve the problem".

The chosen modeling technique relies on Bayesian Networks [Jensen 2001]. Some conceptually derived causal dependencies among students cognitive and affective variables were mapped and related by means of conditional probabilities. This knowledge representation allowed us to deal with the uncertain domain of cognitive-based emotions recognition and is suitable to be refined through controlled experiments, by learning the conditional probability tables between the random variables. The initial assumption of multivalued independence among the random variables simplifies a lot the task of learning the probability density distributions and keeps an acceptable level of generalization. We have started to evaluate the model using a collaborative game test-bed prototype we have developed called "Collaborative Sudoku".

4. Conclusions

The proposed model is already developed and additional information about its functioning and validation can be found in [Pontarolo *et al.* 2008]. As a future improvement, the group aims at aggregating a face-based inference of emotions tool that is already being developed [Oliveira and Jaques 2008]. The ECA is an ongoing work of the PhD student Michelle Leonhardt and its description can be found in more details in [Leonhardt *et al.* 2009].

5. References

Cassel, J. and Sullivan, J. (2000). Embodied conversational agents. MIT Press.

Damásio, A. (1994). Descartes' Error: Emotion, reason, and the human brain. New York, G. P. Putnam.

Elliott, C., Rickel, J. and Lester, J. C. (1999). Lifelike Pedagogical Agents and Affective Computing: An Exploratory Synthesis. In Artificial Intelligence Today, p. 195-211.

Goleman, D. (1995). Emotional Intelligence. New York, Bantam Books.

Jensen, F.V. (2001). Bayesian networks and decision graphs. Berlin, Springer.

Leonhardt, M., Vicari, R. M., Pesty, S. (2009). Enhancing Affective Communication in Embodied Conversational Agents. Submitted to COLIBRI Colloquium 2009.

Lester, J. C. and Stone, B. (1997). Increasing Believability in Animated Pedagogical Agents. International Conference on Autonomous Agents. Marina del Rey, p. 16-21.

Longhi, M. T. ; Bercht, M. ; Behar, P. A. . AnimA: Protótipo para Reconhecimento dos Estados de Ânimo no Processo de Aprendizagem. In: II Workshop - Escola de Sistemas de Agentes para Ambientes Colaborativos, 2008, Santa Cruz do Sul.

Martinho, C., Machado, I. and Paiva, A. (2000). A Cognitive Approach to Affective User Modeling. In Affective Interactions.Springer, p. 64-75.

Oliveira, E. and Jaques, P. A. (2008). Inferindo as emoções do usuário pela face através de um sistema psicológico de codificação facial. In Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais, SBC/ACM, p. 156 - 165.

Ortony, A., Clore, G. and Collins, A. (1988). The Cognitive Structure of Emotions. Cambridge.

Picard, R. W. (1997). Affective computing. Cambridge, Mass., MIT Press.

Pontarolo, E., Jaques, P. A. and Vicari, R. M. (2008). An Affective Model Suitable to Infer the Student's Emotions in a Collaborative Learning Game. In TeLearn 2008.

Scherer, K. R., & Tran, V. Effects of emotion on the process of Organization Learning. In I. Nonaka (Ed.), Handbook of organizational learning and knowledge (pp. 369-392). Oxford, UK: Oxford University Press. 2001c.

User-Centered Evaluation of Information Visualization Techniques: Issues and Perspectives

Carla M.D.S. Freitas[1], **Marcelo S. Pimenta**[1] **and Dominique Scapin**[2]

[1] Universidade Federal do Rio Grande do Sul (UFRGS) – Instituto de Informática
Porto Alegre, RS, Brasil, 91.501-970

[2] INRIA Rocquencourt

{mpimenta, carla}@inf.ufrgs.br, Dominique.Scapin@inria.fr

Abstract. *This paper discusses some important issues related with usability evaluation of information visualization techniques, particularly using a user-centered point of view. This way, we address one of the goals established in the EDGE ("Evaluation methods, Design Guidelines and Environments for Virtual Reality and Information Visualization Techniques") project, namely developing new methods for evaluating information visualization techniques, taking advantage of the experience of both teams in this matter.*

1. Introduction

Following the same path of desktop graphical user interfaces (GUI) and Web-based interfaces, a large use of information visualization techniques will depend on their usability. Whereas the first information visualization (IV) techniques were presented without thorough evaluation studies, researchers have become aware of the importance of such usability studies [Plaisant 2004]. However, despite an evident progress in recent years to establish some 'good practices' for design and usability evaluation of such techniques, the definition of several aspects related to an user-centered perspective for IV techniques remains an open issue.

Historically, a user-centered perspective has its origin related to User Centered Design (UCD), which involves basically simplifying the structure of tasks, making things visible, getting the mapping right, exploiting the powers of constraint, and designing for error. Thus, a user-centered perspective tries to optimize the user interface around how people can, want, or need to work, rather than forcing the users to change how they work to accommodate the system or function. In short, evaluation with a user-centered perspective is an evaluation based on the needs of the user and, for that, we need to know them, their goals and tasks.

In the last years, we have been particularly interested in evaluation of IV techniques. Although there is a great variety of models and techniques for information visualization, there is not yet a consensus about their usability evaluation: what is the meaning of usability for IV techniques? Which characteristics we have to evaluate and how to evaluate them? In addition, which aspects are generic to all types of interactive systems and which are specific to IV techniques? Which of these issues are particularly relevant to IV techniques' evaluation?

Our goal in the EDGE (Evaluation methods, Design Guidelines and Environments for Virtual Reality and Information Visualization Techniques) project is to investigate if and how sound ergonomic user-centered knowledge can be transferred to IV techniques' context. This paper intends to start a discussion and presents some issues concerning IV techniques evaluation we are working on based on past experience [Scapin and Bastien 1997][Freitas et al. 2002].

2. Current Studies in Evaluation of Information Visualization Techniques

The different aspects related to the evaluation of IV techniques have become research issues among the (Human-Computer Interaction) HCI and Visualization community. We are particularly interested in finding answers to the following question: "How do we know if information visualization tools are useful and usable for real users performing real visualization tasks?", i.e., how do we know if a IV technique is able to provide insight to users? In fact, for effective and well-accepted adoption of information visualization tools, they have to be effective, efficient and satisfying for the intended users [Yi et al. 2008].

As new applications are more often related to larger and more complex datasets, the challenges of information visualization involves not only the selection of typical datasets and tasks but defining evaluation methodologies [Shneiderman and Plaisant 2006] and finding the appropriate case studies and users [Seo and Shneiderman 2006] [Valiati et al. 2008]. There has been already some research aiming at consolidating visualization tasks [Amar et al. 2004], collecting benchmark datasets [Plaisant et al. 2008], and experimenting human-computer interaction evaluation methods [Valiati et al. 2008].

The work on identifying and understanding "visualization" users tasks started in 1990, but remained set aside for a long time. Recently reports have been published devoted to understanding and representing the tasks users perform to accomplish their goals. Indeed, the identification and understanding of the nature of the users' tasks in the process of acquiring knowledge from visual representations of data is a recent branch in information visualization research, and some taxonomies have been proposed, for example, [Amar et al. 2004]. We have been working on that subject [Valiati et al. 2007], and have achieved a stable set of tasks that are useful both in the evaluation of new developments and to compare different techniques for some specific application.

Many of the evaluation reports we find in the literature are descriptions of experiments with users targeting the comparison of different IV techniques or assessing a specific technique regarding different tasks [Bertini et al. 2006] [Bertini et al. 2008]. Most of those works focused on controlled experiments in-laboratory, with specific tasks selected from the typical ones performed by the potential users of the visualization technique. Some works addressed different aspects of the evaluation process [Whiting et al. 2008]. The experimental procedure may be inadequate sometimes, mainly during a research exploratory stage when goals and tasks may not be already defined. Thus, for evaluating visualizations, longitudinal or (more broadly) qualitative research studies involving actual users' participation [Shneiderman and Plaisant 2006] [Isenberg et al. 2008] have been strongly recommended.

3. Some Important Issues for a User-Centered Viewpoint of Usability Evaluation of Information Visualization Techniques

A review of existing work on IV techniques evaluation allowed us to identify a significant number of problems.

First of all, from those methods used in traditional interaction evaluation, few are actually used for IV techniques evaluation: *the diversity of methods is quite limited*. In practice, most of the evaluation in IV is oriented to user testing methods. These methods try to check if usability goals are met and also to identify usability problems by conducting experiments, in which users try to solve realistic tasks with the technique. The dependent variables usually measured in this process are task time and task accuracy. Data collected is analyzed statistically, in order to capture the central usability measures effectiveness, efficiency and user satisfaction. However, user testing is not always the best choice because it is a very time consuming process, with high costs.

Two other problems arise from that one. *Evaluation happens too late when employing only user testing*, because such testing should be mainly applied in later stages of development: to testing, a running system is mandatory, and in general, *evaluation process does not follow a general usability evaluation methodology*. Sometimes, instead of empirical methods like user testing (which can only be used after some form of interaction design is available, and for which direct access to end-users is mandatory), it may be interesting to adopt some analytical methods like expert evaluation (inspection based solely on the evaluator's knowledge and experience), document-based evaluation (inspection based on some guiding documents, at various degrees of precision) or even model-based methods (inspection based on some theoretical models, usually cognitive models). These methods are particularly useful when it is not possible to collect data directly from users; but also they can be simply a useful first step to uncover some major usability flaws before investigating more complex problems further.

Our intention is to put emphasis on the proposal of guidelines (not strong recommendations or a strict methodology). From several case studies conducted by Valiati et al. (2008) following the MILCs approach [Shneiderman and Plaisant 2006], we devised that: (1) The context of usage for evaluation must be defined before the beginning of evaluation; (2) Evaluation needs to know who the users are, of what their goals are and to decide which users to support; (3) Evaluation needs to understand which tasks users need to perform and their characteristics (steps, constraints, and other tasks attributes like frequency, priority, etc) and to decide which tasks to support; (4) Evaluation should be performed earlier in the design-development cycle.

We are also focusing on the identification of appropriate usability methods for IV techniques evaluation. Methods for evaluating usability are described in terms of the characteristics, which should be present in the way the evaluation is conducted. The detailed prescription of a method to be used is in relation to a narrowly defined set of user goals, in a specified task domain, with limited metrics. However, much work is needed to extend the scope of current evaluation methods to cope with the many possible IV techniques' usability dimensions. For example, how do we apply usability inspection methods (like heuristics evaluation, for example) to IV techniques? Or how do we guide IV techniques development with these usability considerations?

Consequently, some interesting questions are open for discussion: (1) Can the sound ergonomic knowledge (style guides, architectures, and design and evaluation methods that have been proved adequate for GUIs and Web-based contexts) be transferred to this novel context of IV techniques? If so, how do we deal with the idiosyncrasies of IV techniques: in a generic or in a customized to IVT-specificities way? How do we ensure user involvement in usability evaluation? In fact, user involvement is a direct way to accelerate the process of improving usability evaluation of IV techniques. We think that doing user testing earlier allows usability knowledge be gained rapidly, rather than having simply the technology perfected without user concern.

References

Amar, R., Eagan, J., Stasko, J. (2005) "Low-Level Components of Analytic Activity in Information Visualization . Proc. Information Visualization 2005, pp. 111-147.

Bertini, E., Plaisant, C., Santucci, G. (2006) "BELIV: Beyond Time and Error: Novel Evaluation Methods for Information Visualization". Proceedings (Workshop of the Advanced Visual Interfaces Conference). Venice, Italy, May 23, ACM Digital Library.

Bertini, E. et al.. (2008) "BELIV: Beyond Time and Error: Novel Evaluation Methods for Information Visualization". Proceedings (Workshop of the Conference on Human Factors in Computing Systems). Florence, Italy, April 5, ACM Digital Library.

Freitas, C.M.D.S. et al. (2002) "Evaluating Usability of Information Visualization Techniques". In Proc. Symposium on Human Factors in Computer Systems (IHC 2002), Fortaleza, Brazil.

Isenberg, P. et al. (2008) "Grounded evaluation of information visualizations". In BELIV'08, April 5, Florence, Italy. ACM Digital Library.

Plaisant, C. (2004) "The Challenge of Information Visualization Evaluation". Proc. AVI'04 - ACM Conference on Advanced Visual Interfaces, ACM Press, pages 109-116.

Plaisant, C., Fekete, J.-D., Grinstein, G. (2008) "Promoting Insight Based Evaluation of Visualizations: From Contest to Benchmark Repository". IEEE Transactions on Visualization and Computer Graphics, 14(1): 120-134.

Scapin, D. L. and Bastien, J. M. C. (1997) "Ergonomic Criteria for Evaluating the Ergonomic Quality of Interactive Systems". Behaviour and Information Technology, 16 (4) 220-231.

Seo, J. and Shneiderman, B. (2006) "Knowledge discovery in high dimensional data: Case studies and a user survey for the rank-by-feature framework". IEEE Trans. on Visualization and Computer Graphics 12(3): 311-322.

Shneiderman, B., Plaisant, C. (2006) "Strategies for Evaluating Information Visualization Tools: Multi-dimensional In-depth Long-term Case Studies". In BELIV'06 , May 23, Venice, Italy, 38-43.

Valiati, E. et al. (2007) "Experimental Evaluation of Tasks Classification in Multidimensional Information Visualization". In Proc. CLIHC'07 – Workshop of INTERACT'07.

Valiati, E., Pimenta, M., Freitas, C.M.D.S. (2008) "Using Multi-dimensional In-depth Long-term Case Studies for Information Visualization Evaluation". In BELIV'08, April 5, Florence, Italy. ACM Digital Library.

Whiting, M.A., Haack, J., Varley, C. (2008) "Creating realistic, scenario-based synthetic data for test and evaluation of information analytics software". In BELIV'08, April 5, Florence, Italy. ACM Digital Library.

Yi, J.S. et al. (2008) "Understanding and Characterizing Insights: How do People Gain Insights Using Information Visualization?" In BELIV'08, April 5, Florence, Italy. ACM Digital Library.

Perspectives in the Development of ErgoCoIn

Walter de Abreu Cybis[1] , Dominique L. Scapin[2], Marcelo Morandini[3]

[2]Ecole Polytechnique de Montréal
2500, chemin de Polytechnique, C.P. 6079, succ. Centre-ville - Montréal - Canada.

[2]Institut National de Recherche en Informatique et Automatique
Domaine de Voluceau - Rocquencourt - B.P. 105, 78153 Le Chesnay - France

[3]School of Arts, Science and Humanities – University of Sao Paulo
Av Arlindo Bettio, 1000 – Ermelino Matarazzo – Sao Paulo - Brazil

walter.cybis@polymtl.ca, dominique.scapin@inria.fr,
m.morandini@usp.br

Abstract. *This paper presents current actions and perspectives concerning the development of the ErgoCoIn approach, which allows non expert inspectors to conduct ergonomic inspections of e-commerce web sites. An environment supporting inspections based on this approach was designed and a tool is being developed in order to allow the accomplishment of the approach validation plan. Perspectives concern the development of a wiki environment aimed to support this approach evolution as well as the integration of a specific task oriented log analysis tool.*

Resumo. *Este artigo apresenta as ações atuais e as perspectivas relacionadas ao desenvolvimento da abordagem ErgoCoIn, destinada a permitir que não especialistas possam realizar inspeções da ergonomia de sites de comércio na web. Um ambiente de apoio às inspeções baseadas nesta abordagem foi projetado e um protótipo está sendo desenvolvido visando a validação desta abordagem. As perspectivas ficam por conta do desenvolvimento de um ambiente compartilhado do tipo wiki para apoiar a evolução da mesma, assim como a integração de uma ferramenta especial que realize análises de dados de log segundo uma abordagem orientada a tarefas.*

1. Introduction

ErgoCoIn [1, 3] is an approach designed to provide support to inspectors in order to allow them to perform objective web sites ergonomic inspections[1]. With the goal of improving the quality of theirs diagnoses, this approach postulates that inspectors consider web sites context of use data, including: users, tasks and environments attributes. For objectiveness, only questions associated to information demand predefined by checklists are integrated to the contextual analysis interviews/questionnaires. With the same goal, only applicable questions are presented to inspectors while inspecting the web site. They will be performing "discount" context

[1] This environment was designed with the funds of the CNPq/ProTeM-CC/INRIA project called " TVU - CECI " (March 1999 to April 2001).

of use analysis and applying "low-cost" checklists. Besides objectiveness, the ErgoCoIn approach aims at supporting inspectors producing more homogeneous results when compared to those produced by ergonomic experts.

In this paper we introduce a prototype aimed at supporting the ErgoCoIn approach's validation plan. We discuss the recently identified requirements for the ErgoCoIn development, concerning specifically to the variety and the novelty of the approach's knowledge base. In order to achieve the fulfillment of these requirements, we propose the development of a collaborative effort aimed to insure the continuous enrichment of the ErgoCoIn knowledge base. Finally, we introduce the integration of a task oriented log analyzer to the original EgoCoIn approach.

2. The ErgoCoIn Environment and Validation Tool

Based on the ErgoCoIn logic architecture, we have modeled, designed and implemented a prototype aimed to support the phase of ergonomic inspections. 182 questions linked to 18 Ergonomic Criteria properly ranked are filled into the prototype data base. Once the tool is completed, we will start accomplishing cycles of validation studies focusing on the underlying approach. At each cycle, a number of inspectors will be invited to use the tool in order to perform inspections of a given e-commerce web site. Results from all inspectors, as well as the log of their actions will be gathered and analyzed from both homogeneity and objectiveness points of view. The goal behind revision proposals is to get inspections more objective and reports more coherent. Validation cycles will be repeated until expected objectiveness and homogeneity criteria have being reached.

The inspections cycles will allow us to have a better understanding of the way tasks concerning ergonomic inspections of web sites are accomplished, and specify a tool specially fitted to those tasks. Indeed, we intent to specify an ErgoCoIn user interface able to support inspectors spread all over the world performing ergonomic inspections of web sites from different domains, not only the ones concerning e-commerce. The idea is to offer the tool to those who want to make inspections, and want to contribute to the enrichment of the ErgoCoIn knowledge base and programming code.

3. The Wiki-ErgoCoIn

We propose to change the scope of the ErgoCoIn development in order to support a collaborative initiative. In fact, this kind of initiative is among the most interesting phenomena observed in the recent history of the web. Collaboration is authorized by special functions offered by web sites allowing users to create, share and organize the content by themselves. Best examples of socially constructed web sites are Facebook, Youtube, Flickr, Digg, del.ici.ous and Wikipedia.

Particularly, the Wikipedia is the most successful example of collaboration concerning scientific content on the web. This socially constructed encyclopedia features remarkable internet traffic numbers as it is the 9th most visited web site in the whole Web. From 2001 to now, 7.2 million of articles were posted in Wikipdia. Those were produced by 7.04 million of editors following some style and ethic rules [4]. Wilkison and Huberman [5] performed a study concerning 52.2 million edits in 1.5 million articles in the English language Wikipedia posted by 4.79 million contributors

between 2001 and 2006. They splitted out a group of 1,211 "*featured articles*", which accuracy, neutrality, completeness and style are assured by Wikipedia editors. Comparisons between featured and normal articles showed a strong correlation among the article quality, the number of edits and the number of distinct editors. In the same study, the authors were able to associate attractiveness of the articles (number of visits) to the edits novelty.

The goal of having ErgoCoIn as a collaborative web initiative is to increase the generality and attractiveness of its contents as well as the quality of the results this approach could lead to. Indeed, the Wiki-ErgoCoIn is being designed in order to allow ergonomic inspectors all over the world to share efforts and responsibilities concerning the ErgoCoIn knowledge base extension and generalization. In doing so, we can expect that the Wiki-ErgoCoIn will always feature newly proposed questions concerning ergonomics of web sites from different application domains, interface styles and components. Contributions should fulfill a basic requirement: follow free-content collaboration rules like those developed by Wikipedia. We believe that the results obtained by such a cooperative approach can be much more efficient and reliable than the ones that would be obtained solely by individual initiatives.

4. The UseMonitor component

Another kind of extension that is being considered concerns the integration of the results from the analysis of usage log data produced with this approach. Such data can be collected using specific software tools for this purpose. In fact, a usability oriented web analyzer called UseMonitor is being developed and associated to the ErgoCoIn approach. This tool can present warnings about the "a posteriori" perspective on usability problems, i.e., interaction perturbations occurring while users are interacting with the web site in order to accomplish their goals. Basically, the UseMonitor can indicate when the observed efficiency rate is particularly low. Detailed efficiency indication is about the rates and time spent of unproductive users' behaviors like solving error, asking help, hesitation, deviation, repetition and so on. Furthermore, the UseMonitor can indicate web pages related to this kind of perturbations. A logic architecture based on the integration of (i) a typology of usability problems, (ii) the ergonomic criteria/recommendations and (iii) a model of interface components is also being defined. This will allow the UseMonitor warning the inspectors about a detailed interface aspect causing an actual usability perturbation (a posteriori result), while ErgoCoIn will be helping inspectors identifying the user interface component responsible for such perturbation as well as indicating how to fix it (a priori result). The integration of ErgoCoIn and UseMonitor defines the ErgoManager environment [2]. As a tool for usability evaluation of such an environment will be automating both processes, the failure identification (by log analysis) and failure analysis (by guidelines processing. Details of this architecture are being defined and will be introduced in future publications.

5. Conclusions

ErgoCoIn is an inspection approach strongly based on knowledge about ergonomics of web user interfaces. This knowledge is intended to guide inspectors while undertaking contextual data gathering and analysis, checklists based inspections and report actions. In this paper we described details of this approach and the environment designed to

support it. We have also introduced the tool that is under development to validate its structure and contents. We will perform the validation activities following cycles of application-analysis-revisions until the approach reaches expected objectiveness and homogeneity goals.

However, the success of the ErgoCoIn initiative depends basically on the variety and the novelty of its knowledge. Nowadays, this approach is linked to the ergonomics of the current e-commerce web applications and interfaces technologies, styles and components. Indeed, all these aspects may evolve continuously using just e-commerce may be a very limited scope. Consequently, there is the need to undertake actions in order to face the challenge of continuously getting ErgoCoIn contents up to date and varied to support the production of inspection reports in different web sites domains. An open initiative is being proposed by which anybody knowledgeable will be authorized to contribute to the enrichment of the Wiki-ErgoCoIn knowledge base. Consultative and executive boards will be created to define strategies and policies concerning implementation of this ergonomics inspection wiki.

6. References

1. Cybis, W.A ; SCAPIN, D. & ANDRES, D. P. . Especificação de Método de Avaliação Ergonômica de Usabilidade para Sites/Web de Comércio Eletrônico. In: Workshop on Human Factors in Computer Systems, 2000, Gramado. Proccedings of 3rd Workshop on Human Factors in Computer Systems. Porto Alegre : Ed. Sociedade Brasileira de Computação, 2000. v. I. p. 54-63.
2. Cybis, Walter de Abreu . UseMonitor: suivre l'évolution de l'utilisabilité des sites web à partir de l'analyse des fichiers de journalisation. In: 18eme Conférence Francophone sur l'Interaction Humain-Machine, 2006, Montréal. Actes de la 18eme Conférence Francophone sur l'Interaction Humain-Machine. New York : ACM - The Association for Computing Machinery, 2006. v. 1. p. 295-296.
3. Cybis, Walter de Abreu . ErgoManager: a UIMS for monitoring and revising user interfaces for Web sites . Rocquencourt: Institut National de Recherche en Informatique et en Automatique, 2005 (Research report).
4. Wikipedia. www.wikipedia.org, assessed in february 2009.
5. Wilkinson, D & Huberman, B., Assessing the value of cooperation in Wikipedia, First Monday 12 (4), 2007. http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1763/1643.

Partitioning Clustering Algorithms for Interval-Valued Data based on City-Block Adaptive Distances

Francisco de A.T. de Carvalho[1], Yves Lechevallier[2]

[1]Centro de de Informática (CIn) – Universidade Federal de Pernambuco (UFPE)
Av. Prof. Luiz Freire, s/n – Cidade Universitária
CEP 50740-540 – Recife – PE – Brazil

[2]INRIA – Rocquencourt
Domaine de Voluceau – Rocquencourt
B.P. 105 – 78153 Le Chesnay Cedex, France

fatc@cin.ufpe.br, yves.lechevallier@inria.fr

Abstract. *The recording of interval-valued data has become a common practice nowadays. This paper presents some partitioning clustering algorithms for interval-valued data. The proposed methods furnish a partition of the input data and a corresponding prototype (a vector of intervals) for each cluster by optimizing an adequacy criterion which is based on suitable adaptive city-block distances between vectors of intervals. Experiment with a real interval-valued data set shows the usefulness of the proposed method.*

1. Introduction

Clustering is one of the most popular tasks in knowledge discovery and is applied in various fields, including data mining, pattern recognition, computer vision, etc. These methods seek to organize a set of items into clusters such that items within a given cluster have a high degree of similarity, whereas items belonging to different clusters have a high degree of dissimilarity. Partitioning clustering methods [Everitt 2001], [Gordon 1999], [Jain et al 1999] seek to obtain a single partition of the input data into a fixed number of clusters. Such methods often look for a partition that optimizes (usually locally) an adequacy criterion function.

The partitioning dynamic cluster algorithms [Diday and Simon 1976] are iterative two steps relocation clustering algorithms involving at each iteration the construction of the clusters and the identification of a suitable prototype (means, factorial axes, probability laws, etc.) of each cluster by locally optimizing an adequacy criterion between the clusters and their corresponding prototypes.

The adaptive dynamic clustering algorithm [Diday and Govaert 1977] also optimize a criterion based on a measure of fitting between the clusters and their prototypes, but there are distances to compare clusters and their prototypes that change at each iteration. These distances are not determined once and for all, and moreover, they can be different from one cluster to another. The advantage of these adaptive distances is that the clustering algorithm is able to recognize clusters of different shapes and sizes.

Often, objects to be clustered are represented as a vector of quantitative features. However, the recording of interval data has become a common practice in real world

applications and nowadays this kind of data is often used to describe objects. Symbolic Data Analysis (SDA) is an area related to multivariate analysis, data mining and pattern recognition, which has provided suitable data analysis methods for managing objects described as a vector of intervals [Bock and Diday 2000].

Concerning dynamical cluster algorithms for symbolic interval data, [Chavent and Lechevallier 2002] proposed an algorithm using an adequacy criterion based on Hausdorff distances. [Souza and De Carvalho 2004] presented a dynamic cluster algorithm for symbolic interval data based on City-Block distances. [De Carvalho et al 2006] proposed an algorithm using an adequacy criterion based on adaptive Hausdorff distances for each cluster. In this paper, we introduce partitioning dynamic clustering methods based on single adaptive city-block distances. These adaptive distances change at each iteration but are the same for all clusters.

2. Partitioning dynamic clustering algorithms for interval-valued data based on city-block adaptive distances

This section presents partitioning dynamic clustering methods for interval-valued data based on single adaptive city-block distances. These adaptive distances are defined by one weight vectors. The main idea of these methods is that there is a distance to compare clusters and their representatives (prototypes) that changes at each iteration but is the same for all clusters.

Let Ω be a set of n objects described by p interval-valued variables. An interval-valued variable [Bock and Diday 2000] is a mapping from Ω in $\Re$ such that for each $i \in \Omega$, $X(i) = [a, b] \in \Im$, in which $\Im$ is the set of closed intervals defined from $\Re$. Each object i $(i=1,\ldots,n)$ is represented as a vector of intervals $\boldsymbol{x}_i = ([a_i^1 , b_i^1], \ldots , [a_i^p , b_i^p])$.

This adaptive clustering method looks for a partition of Ω into K clusters $P_1,\ldots,P_K$ and their corresponding prototypes $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_K$ such that an adequacy criterion J measuring the fitting between the clusters and their prototypes is locally minimized. Assuming that each cluster P_k is also represented as a vector of intervals $\boldsymbol{y}_k = ([\alpha_k^1,\beta_k^1], \ldots , [\alpha_k^p,\beta_k^p])$, the criterion J is defined as:

$$J = \sum_{k=1}^{K} \sum_{i \in P_k} d(\mathbf{x}_i, \mathbf{y}_k) \text{ in which } d(\mathbf{x}_i, \mathbf{y}_k) = \sum_{j=1}^{p} \lambda^j \left[\max\{ | a_i^j - \alpha_k^j |, | b_i^j - \beta_k^j | \} \right]$$

is an adaptive city-block distance measuring the dissimilarity between an object $\boldsymbol{x}_i$ $(i = 1, \ldots , n)$ and a cluster prototype $\boldsymbol{y}_k$ $(k = 1, \ldots ,K)$ that is parameterized by the weight vector $\boldsymbol{\lambda} = (\lambda^1, \ldots , \lambda^p)$, which changes at each iteration but is the same for all clusters.

This algorithm sets an initial partition and alternates three steps until convergence, when the criterion J reaches a stationary value representing a local minimum.

2.1. Step 1: definition of the best prototypes

The partition of Ω into K clusters and the weight vector $\boldsymbol{\lambda}$ are fixed. The prototype $\boldsymbol{y}_k = ([\alpha_k^1,\beta_k^1], \ldots , [\alpha_k^p,\beta_k^p])$ of cluster P_k, which minim**izes** the clustering criterion J, has the boundaries of the interval $[\alpha_k^j,\beta_k^j]$ $(j = 1, \ldots , p)$ calculated **a**ccording to:

α_k^j is the median of the set $\{ a_i^j : i \in P_k \}$ and β_k^j is the median of the set $\{ b_i^j : i \in P_k \}$

2.2. Step 2: definition of the best adaptive distances

The partition of Ω into K clusters and the prototypes $\mathbf{y}_k$ $(k=1,\ldots,K)$ are fixed. The vector of weights $\boldsymbol{\lambda} = (\lambda^1, \ldots, \lambda^p)$, which minimizes the clustering criterion J under $\lambda^j > 0$ and $\prod_{j=1}^{p} \lambda^j = 1$, has its weights λ^j $(j=1,\ldots,p)$ calculated according to the following expression:

$$\lambda^j = \frac{\left\{\prod_{h=1}^{p}\left(\sum_{k=1}^{K}\left[\sum_{i\in P_k}\left(\left|a_i^h - \alpha_k^h\right| + \left|b_i^h - \beta_k^h\right|\right)\right]\right)\right\}^{\frac{1}{p}}}{\sum_{k=1}^{K}\left[\sum_{i\in P_k}\left(\left|a_i^j - \alpha_k^j\right| + \left|b_i^j - \beta_k^j\right|\right)\right]}$$

2.3. Step 3: definition of the best partition

T the prototypes $\mathbf{y}_k$ $(k=1,\ldots,K)$ and the vector of weights $\boldsymbol{\lambda} = (\lambda^1, \ldots, \lambda^p)$ are fixed. The clusters P_k $(k = 1, \ldots, K)$, which minimizes the clustering criterion J, are updated according to the following allocation rule:

$$P_k = \{i \in \Omega : d(\mathbf{x}_i, \mathbf{y}_k) \leq d(\mathbf{x}_i, \mathbf{y}_h), \forall h \neq k\}$$

3. City Temperature Interval-Valued Data Set

The city temperature interval-valued data set gives the average minimal and average maximal monthly temperatures of cities in degrees centigrade. This data set consists of a set of 503 cities described by 12 interval-valued variables (see Figure 1).

	January	February	...	November	December
Amsterdam	[−4, 4]	[−5, 3]	...	[1, 10]	[-1, 4]
Athens	[6, 12]	[6, 12]	...	[11, 18]	[8, 14]
...	...	...	...	...	...
Mauritius	[22, 28]	[22, 29]	...	[19, 27]	[21, 28]
...	...	...	...	...	...
Vienna	[−2, 1]	[−1, 3]	...	[2, 7]	[1, 3]
Zurich	[−11, 9]	[−8, 15]	...	[0, 19]	[−11, 8]

Figure 1. City Temparature Interval-Valued Data Set

With this city temperature interval-valued data set, the clustering algorithm was run until the convergence to a stationary value of the criterion J 100 times and the best result according to the adequacy criterion was selected. The main characteristics of the 6-cluster partition furnished by algorithm were:

- Cluster 1: the cities have very cold temperatures in winter similar to that of northern and eastern Europe;

- Cluster 2: the cities have temperatures similar to that of cities located in the southern hemisphere.

- Cluster 3: the cities have temperatures similar to that of western and central Europe.

- Cluster 4: the cities have a tropical climate and warm to hot temperatures

- Cluster 5: the cities have temperatures similar to that of southern Europe

- Cluster 6: the cities have a sub-tropical climate and warm to hot temperatures

Figure 2 shows the most discriminant months for the prototypes.

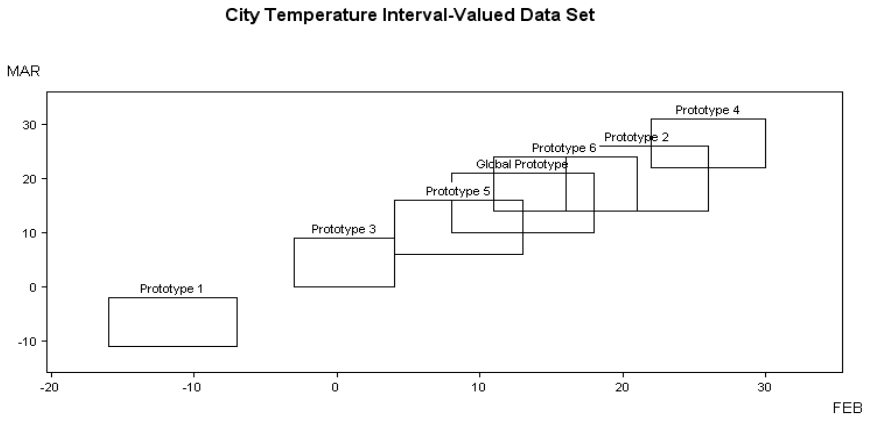


Figure 2. The most discriminant months

7. Final Remarks

In this paper, a partitioning clustering algorithm for interval-valued data was presented. This method furnishes a partition of the input data and a corresponding prototype for each cluster by optimizing an adequacy criterion which is based on adaptive city-block distances between vectors of intervals. An application with a city temperature interval-valued data set illustrates the usefulness of the proposed approach.

References

Bock, H.H. and Diday, E. (2000), Analysis of Symbolic Data, Exploratory methods for extracting statistical information from complex data., Springer, Berlin Heidelberg.

Chavent, M. and Lechevallier, Y. (2002) "Dynamical Clustering Algorithm of Interval Data: Optimization of an Adequacy Criterion Based on Hausdorff Distance", In: Classification, Clustering and Data Analysis, Edited by A. Sokolowski and H.-H. Bock, Springer, Heidelberg, 53–59.

De Carvalho, F.A.T, Souza, R.M.C.R., Chavent, M. and Lechevallier, Y. (2006), Adaptive Hausdorff distances and dynamic clustering of symbolic data, Pattern Recognition Letters, 27 (3), 167–179.

Diday, E. and Govaert, G. (1977), Classification Automatique avec Distances Adaptatives, R.A.I.R.O. Informatique Computer Science, 11 (4), 329–349.

Diday, E. and Simon, J.C. (1976) "Clustering analysis", In Digital Pattern Classification, Edited by K.S. Fu, Springer, Berlin et al, 47–94.

Everitt, B. (2001), Cluster Analysis, Halsted, New York.

Gordon, A. D. (1999), Classification, Chapman and Hall/CRC, Boca Raton, Florida.

Jain, A.K., Murty, M.N. and Flynn, P.J. (1999), Data Clustering: A Review, ACM Computing Surveys, 31 (3), 264-323.

Souza, R. M. C. R. and De Carvalho, F. A. T. (2004), Clustering of interval data based on city-block distances, Pattern Recognition Letters, 25 (3), 353–365

From data to knowledge through collaboration: bridging Wikis and knowledge systems

H. Astudillo[1], G. Canals[2], A. Diaz[3], A. Napoli[2], and M.G. Pimentel[4]

[1]Departamento de Informática, Universidad Técnica Federico Santa María (UTFSM)
Valparaiso, Chili (hernan@inf.utfsm.cl)

[2]LORIA (CNRS – INRIA – Nancy Université), Vandoeuvre-lès-Nancy, France
(Gerome.Canals@loria.fr,Amedeo.Napoli@loria.fr)

[3]LIFIA, Universidad de La Plata, Argentina
(alicia.diaz@lifia.info.unlp.edu.ar)

[4]Laboratório Intermidia Departamento de Ciências de Computação,
Universidade de São Paulo (São Carlos), Brazil (mgp@icmc.usp.br)

Abstract. *This paper presents some scientific aspects of the STIC AmSud project "Semantic-based support for Collaborative Design Activity". The main goal of this cooperation project is to define an operational environment for demonstrating how semantic technologies can be associated with collaborative design activities. In this way, semantic Wikis provide an original and operational infrastructure for efficiently combining both technologies.*

1. Introduction

This paper briefly presents the scientific aspects of a STIC AmSud project entitled as "Semantic-based support for Collaborative Design Activity" and involving researchers from Argentina, Brazil, Chili, and France. This project is aimed at integrating and sharing experiences and research efforts for defining an operational environment –including models, methods and tools– where semantic technologies are associated to collaborative computer-supported design activities. In particular, the project tries to demonstrate that semantic Web technologies are a suitable option to guide and improve a collaborative computer-assisted design process. Such an environment has to take advantage of domain ontologies for helping designers in assembling design components by guiding search for adequate documents, components, skills, and detection of conflicts as well. This research work involves competencies for: (i) knowledge engineering: knowledge representation, reasoning, knowledge discovery, ontologies and semantic Web technologies, (ii) support to collaborative design activities, with a focus on coordination and awareness, but also knowledge management for design process. The competencies of each team in the project are complementary w.r.t. the original topic of the project.

Researchers in the project met three times since the beginning of the project (April and October 2008, February 2009). For carrying on a proper research line in accordance with the original objective, it was decided to work on semantic Wikis with an application in the teaching domain, namely how to use semantic Wiki technology for designing a course on semantic Wikis themselves. For understanding this proposition, we introduce some background elements. A Wiki is roughly speaking a Web site (built around a set of pages) that can be edited by several people, possibly at the same time. Editing is

the collaborative task and problems of coherence and needs for synchronization appear while several people are working together. In this way, combining Wikis and semantic Web technologies is considered as a promising alternative for collaboratively creating and using information on the Web. The user-friendliness of Wikis as regarding multi-site content generation and the power of semantic technologies for organizing and retrieving knowledge may complement one another towards a new generation of Web-based content management systems. Accordingly, a semantic Wiki can be seen as a Wiki including an associated ontology, i.e. an operational representation model of domain knowledge, that can be used for annotating the content of Wiki pages and used for typing hyperlinks and testing consistency of contents (e.g. two users cannot state contradictory facts w.r.t. the underlying ontology). Moreover, an annotation process w.r.t. an ontology has a direct impact on knowledge access, semantic search and reuse, collaborative authoring, and social collaboration [Auer et al. 2006, Krötzsch et al. 2006].

In the present application context, i.e. designing a course on the topic of semantic Wikis, one has to be able to determine a set of resources –mainly textual documents but possibly extensible to multimedia support, e.g. video and sounds– to be analyzed and to set on a process able to extract useful units from the resources to be considered for designing the target course in being guided by a domain ontology. However, the transformation process from data to knowledge is hard to be fully automated: it may be oriented and guided by human collaboration at each step of the knowledge discovery process, i.e. when preparing data, tuning data mining algorithms, and interpreting extracted units. The process of turning the content of textual documents into knowledge units embedded in Wiki pages, and the different operations are illustrated on Figure 1. In this way, A parallel can be drawn with the way an analyst usually guides a knowledge discovery process [Lieber et al. 2008].

This is the purpose of this paper to introduce and explain a practical way of integrating semantic and collaborative technologies through a “semantic Wiki”. Then, it is shown how to combine semantic techniques such as knowledge representation and knowledge discovery within a Wiki environment for guiding the transformation process from data to knowledge units for a practical application.

2. Knowledge discovery and representation: from resources to an ontology

Ontologies are the backbone of semantic Web in allowing software and human agents to communicate and to share domain knowledge [Antoniou and van Harmelen 2004, Staab and Studer 2004]. From a formal point of view, an ontology is considered as an explicit specification of a domain conceptualization. For being operational, an ontology has to be encoded within a knowledge representation language such as a description logic [Baader et al. 2003] or OWL [Antoniou and van Harmelen 2004]. From a practical point of view, besides ontologies, there exist different types of “ontological resources” such as thesaurus, vocabularies, dictionaries, collections of documents, and databases. Every ontological resource provides a specific aspect of domain knowledge. For taking into account these ontological resources, a framework can be designed in which the content of resources can be integrated for being used as a “domain knowledge container” for knowledge sharing and reuse.

Following this way, Formal Concept Analysis (FCA) and its extension Rela-

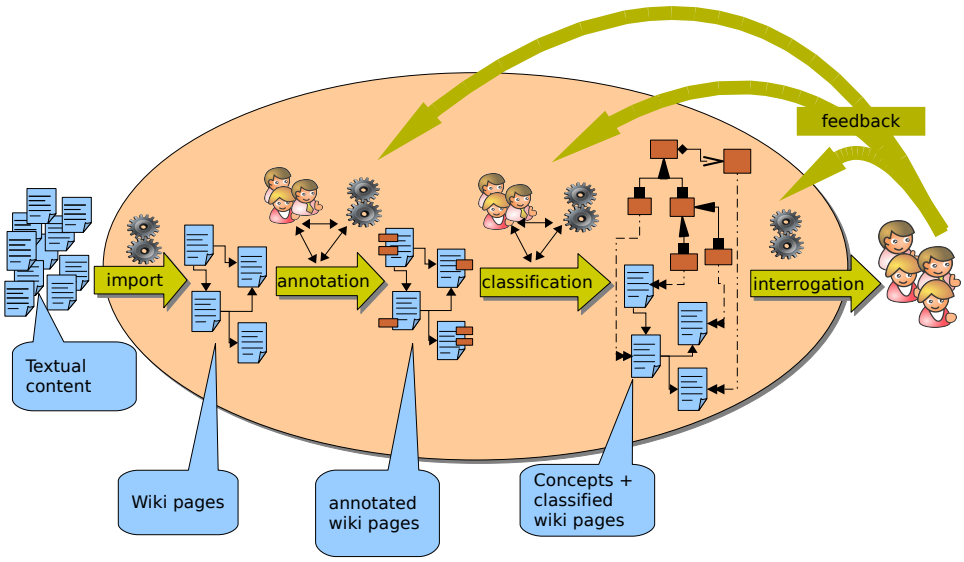


Figure 1. The transformation process from textual content to knowledge units.

tional Concept Analysis (RCA) can be used for designing and completing an ontology from a set of resources (especially textual documents as in [Bendaoud et al. 2008b, Bendaoud et al. 2008a]). FCA and RCA lead the transformation process between ontological resources and a concept lattice supporting an "ontological schema", i.e. a set of related concept definitions. One way of processing is to build a "source ontology" from resources and then to extend this source ontology by progressively adding new units extracted from additional resources. The transformation process can be based on FCA operations (e.g. apposition) and on RCA operations (e.g. extracting relations between concepts). Finally, elements in the resulting concept lattice can be represented within a knowledge representation language such as OWL for obtaining a concept hierarchy representing the ontological schema given by the concept lattice. In this way, FCA and RCA are "core processes" in the design of a target ontology from a set of heterogeneous resources. Firstly, FCA and RCA take into account all basic elements within an ontology, such as objects (or individuals), attributes, and relations, i.e. roles with a domain, a range, and possibly a cardinality and a quantification. Secondly, FCA provides operations for creating, managing, and updating concept lattices. When the resulting concept lattices are transformed into concept hierarchies, a classifier can be used for classification-based reasoning, content-based information retrieval, and answering queries.

This is an operational way of designing ontologies from a set of resources but this process is hard to be fully automated and requires a human intervention. This calls for collaborative architecture such as provided by a Wiki within which the transformation process from resources to knowledge units is explained in the next section.

3. Wikis and Semantic Wikis

A Wiki is a collection of Web pages allowing collaboration and user-contributed "knowledge production", by enabling users to contribute or modify content of pages

[Cunningham and Leuf 2001]. One of the best-known repository of user-contributed knowledge is Wikipedia, also one of the largest and fastest growing online sources of encyclopedic knowledge [Krötzsch et al. 2007]. The richness of its embedded structural information is mainly based on hyperlinks, with relations such as synonymy, polysemy, and additional tools such as infoboxes and templates. A Wiki offers simplicity and a social dimension, and emergence of structured knowledge repositories of collaborative nature.

Here, knowledge has to be understood as "knowledge for human agents" and not for software agents. This is why, for guiding a coherent development of a Wiki infrastructure, semantic technologies and especially ontologies can be used within Wikis, leading to semantic Wikis [Auer et al. 2006, Krötzsch et al. 2006, Buffa et al. 2008]. Semantic Wikis allow knowledge processing for humans and machines. The Wiki infrastructure can be used for supporting an ontology: the Wiki can be considered as a front-end of the ontology maintenance system, with Wiki pages as concepts, typed links as relations and attributes. By contrast, ontologies can be used within Wikis for supporting page selection, annotation (tag organization), searching, querying, faceted navigation, guided edition, and consistency checking. For example, the semantic mediawiki system addresses consistency of contents, knowledge access, and reusing information [Krötzsch et al. 2006]. Moreover, an annotation process w.r.t. an ontology has a direct impact on querying information: annotations can be categories, relations, and attributes, and can be represented as logical statements manipulated within reasoning schemes. In the same way, three main tasks linked to Wikis are improved by the presence of an ontology, namely collaborative authoring (editing), social collaboration (change tracking), and semantic search (browsing) [Auer et al. 2006].

4. Combining knowledge discovery, ontologies, annotation, and semantic Wikis

There are two main views relating Wikis and ontologies: (i) Wikis for improving ontology infrastructure, and (ii) ontologies for improving the development and the management of Wikis.

- In the first view, collaboration plays an important role for editing pages and for gathering and integrating resources of different types. Knowledge discovery (KD) processes can be applied to such a container of resources for extracting units. These units, after interpretation by an analyst, can be embedded within an ontology. In this view, a semantic Wiki can be used for selecting, collecting, and preparing data (documents) in a collaborative way for ontology design and extension.
- In the second view, an ontology plays the role of a domain model, providing a "domain terminology" with terms and associated meaning. In this view, improvements for the Wiki activity are mainly based on document searching –searching by content– and understanding, coherence checking, and guided editing. Both views involve collaborative aspects and "knowledge production" (i.e. extraction or creation): collaboration based on a Wiki guides knowledge organization and evolution in the first view while an ontology can be used for controlling the evolution and checking the consistency of the new elements brought through collaboration in the second view.

Actually, a semantic Wiki can be used as a support for knowledge discovery and knowledge management by combining two interrelated views of Wikis and ontologies. The following loop summarizes the operations that are under study within the research group and that will be tested on the design of a course :

- Based on an initial ontology, a set of documents of interest "ranked" by their content w.r.t. a given topic is selected.
- Then, documents are annotated and related (through hyperlinks) in a collaborative way and under the control of the ontology, i.e. using terms defined within the ontology.
- The resulting set of annotated documents can be analyzed using data mining algorithms for extracting elements of interest.
- These new elements can be interpreted and then proposed as knowledge units for extending and improving the initial ontology. The Wiki infrastructure can be used for making easier interactions.
- The loop is closed: starting from an ontology and going back to the ontology through a collaborative activity within a semantic Wiki using knowledge discovery and knowledge representation techniques.

This process can be seen as a slight adaptation of the transformation process given on Figure 1. An example on the design of a course will allow us to instantiate this design loop. The following operations can be carried out and the content of the course could be roughly the following:

- Presentation of a running example and its context: e.g. organization of the collections in a library or in a museum and visits.
- What is a Wiki and what are the principles of "computer supported cooperative work" (CSCW).
- What are semantic technologies, knowledge representation, reasoning, ontologies, annotation, and semantic Web.
- How Wikis meet semantic technologies, or alternatively how CSCW meet semantic Web: anatomy of a semantic Wiki (e.g. reasoning tools, storage, querying).
- Case study: collaborative course design.
- Advanced topics: e.g. knowledge discovery guided by domain knowledge, good practices in practical aspects, study and use of feedback and of recommendations.
- Conclusion, generalization, and reuse aspects.

5. Conclusion

In this paper we presented some scientific aspects of a STIC AmSud project entitled as "Semantic-based support for Collaborative Design Activity". One objective of this research project is to define an environment where collaborative and semantic technologies meet for improving design activities. We are currently building a framework for building new courses from a set of resources (such as pedagogical documents) using collaboration, knowledge discovery, and annotations processes, supported by the infrastructure of a semantic Wiki. The semantic Wiki infrastructure allows guided interactions between users, documents, and ontologies. The underlying knowledge system, combining knowledge discovery and knowledge representation processes, guides and improves information retrieval, resource classification and annotation, course design, and reasoning for problem solving and evaluation.

At the moment, people in the research project are working on the details of the design process both from theoretical and practical points of views. One important thing is to deeply understand the interactions of semantic and Wiki technologies, and to formally describe the role from each sides. Embedding collaborative aspects of Wiki and semantic techniques in association with knowledge discovery and representation (text mining, annotation) determines a new research direction that has still to be fully investigated.

The existence of this AmSud research project gives us the opportunity of working together with different and complementary backgrounds on these aspects with the hope of obtaining substantial results in a next future.

References

Antoniou, G. and van Harmelen, F. (2004). *A Semantic Web Primer*. The MIT Press, Cambridge, Massachusetts.

Auer, S., Dietzold, S., and Riechert, T. (2006). Ontowiki - a tool for social, semantic collaboration. In *The Semantic Web – ISWC 2006*, LNCS 4273, Springer, pages 736–749.

Baader, F., Calvanese, D., McGuinness, D., Nardi, D., and Patel-Schneider, P., editors (2003). *The Description Logic Handbook*. Cambridge University Press, Cambridge, UK.

Bendaoud, R., Napoli, A., and Toussaint, Y. (2008a). Formal concept analysis: A unified framework for building and refining ontologies. *Proceedings of the 16th EKAW*, LNCS 5268, Springer, pages 156–171.

Bendaoud, R., Toussaint, Y., and Napoli, A. (2008b). Pactole: A methodology and a system for semi-automatically enriching an ontology from a collection of texts. *Proceedings of ICCS 2008*, LNCS 5113, Springer, pages 203–216.

Buffa, M., Gandon, F. L., Ereteo, G., Sander, P., and Faron, C. (2008). Sweetwiki: A semantic wiki. *Journal of Web Semantics*, 6(1):84–97.

Cunningham, W. and Leuf, B., editors (2001). *The Wiki Way: Quick Collaboration on the Web*. Addison-Wesley.

Krötzsch, M., Vrandecic, D., and Völkel, M. (2006). Semantic mediawiki. In *The Semantic Web – ISWC 2006*, Lecture Notes in Computer Science 4273, pages 935–942. Springer.

Krötzsch, M., Vrandečić, D., Völkel, M., Haller, H., and Studer, R. (2007). Semantic wikipedia. *Journal of Web Semantics*, 5(4):251–261.

Lieber, J., Napoli, A., Szathmary, L., and Toussaint, Y. (2008). First elements on knowledge discovery guided by domain knowledge (KDDK). *Concept Lattices and Their Applications*, LNAI 4923, Springer, pages 22–41.

Staab, S. and Studer, R., editors (2004). *Handbook on Ontologies*. Springer, Berlin.

Towards Contextualized Web Searches

Adriana S. Vivacqua[1], Jano M. de Souza[2]

[1]RCT/PURO – Universidade Federal Fluminense (UFF)
[2]PESC/COPPE – Universidade Federal do Rio de Janeiro (UFRJ)
Rio de Janeiro, Brazil

avivacqua@vm.uff.br, jano@cos.ufrj.br

Abstract. *This paper presents a collaborative project currently under discussion with a French partner, building on a prior successful experience with another French institution. We address the explosion of internet resources and ensuing information overload, which makes it hard for individuals to find relevant information. We believe in a move towards a more personalized web experience, and start by focusing on the search activity: searches are hardly ever done in a vacuum (there is an underlying activity) and should be contextualized with other actions the user is taking. In this project, we plan to develop techniques for personalized search, using models developed by French and Brazilian parties and creating new ones.*

1. Introduction

The large amount of information currently available online makes it hard for a user to identify the importance, quality and relevance of data retrieved. Data without provenance, web pages with conflicting or incomplete information and questionable sources are some of the problems a user must face in his or her search for information on the web. These issues make the user's task harder, as he or she faces numerous sources and is forced to identify the one that contains the information he or she needs (Gauch et al., 2007).

Most current search engines make little distinction between users and provide the same results for all of them, based only on the search terms used. However, searches are hardly ever executed "in a vacuum": most searches serve an ulterior purpose, and are conducted as a part of a larger activity. As such, the most useful search results are those that serve that larger activity, and will not necessarily be the same for a different activity.

As a simple example, consider an arts student writing a paper on an analysis of the usage of the image of the Madonna in the Italian Renaissance period. An obvious initial search for the keyword "Madonna" in Google[1] yields no useful results for this student. All results relate to Madonna, the singer. The first related result appears only in the fifth page of search results. An improved search for "Madonna Italian" yields a relevant result in the first page (ranked fifth), but further searches would be needed to improve results. An even better query, "Madonna renaissance" yields several useful

[1] www.google.com

results in the first and subsequent pages. However, three attempts were necessary to arrive at the correct search terms for the query to generate useful results.

We contend that a contextualized approach will yield better results than the general approach currently adopted. Understanding the context in which the search is conducted and using it to improve the retrieval of web documents should result in more appropriate results. Our goal is to provide contextualized results. This also means that the definition of a good result dependent on the context in which the search was performed. Precision and recall will be used to evaluate results, as they are standard search measurements. Evaluation will also characterize the context in which the search was undertaken.

One previous approach (Lakshmanan, 2004) uses TFiDF and the vector space model, standard information retrieval techniques, to, given a set of documents that define the context of a search, expand the query to yield better results. We believe this approach can be improved on. The problem of producing personalized results has also been recognized by the web search giants Google and Yahoo!, who have recently developed ways to personalize search (in projects such as Yahoo! SearchPad[2], Y!Q[3] and Google Web History[4]).

2. The Contextual Search Approach

Our approach to contextual searching is based on a prior successful collaboration with a French university: in collaboration with Prof. Jean Paul Barthès of Université Technologie de Compiègne, we developed an agent based collaborative work tool that used individual contexts to disseminate knowledge, under a *co-tutelle* agreement. This partnership resulted in a doctoral thesis, after a one year exchange period at UTC.

Based on this previous experienced, we propose a layered approach that captures contextual knowledge of individual work and applies it to the individual's activities. In our architecture (illustrated in Figure 1), a Context Acquisition Layer helps determine the user's activity and instantiates the contextual model for the current activity. This model is then used by query filtering and expansion modules to perform better searches. Context acquisition would ideally be performed automatically, but may involve some user input. The feasibility of automatic context acquisition will be evaluated, but is not the main thrust of this project.

[2] http://help.yahoo.com/l/us/yahoo/search/searchpad/

[3] http://yq.search.yahoo.com/

[4] https://www.google.com/accounts/ServiceLogin?hl=en&continue=http://www.google.com/psearch&nui=1&service=hist

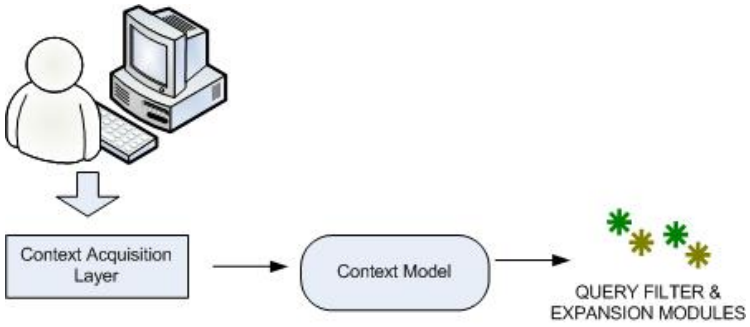


Figure 1. Envisioned System Architecture

Query manipulation modules are the main focus of this project. Our goal is to improve precision and recall of web searches, by taking into account contextual information. Two lines of work will be explored (possibly in parallel):

- Query Filtering: this involves submitting the query, as defined by the user, to the search engine(s), and filtering the results before presentation to the user. A system would be able to analyze more results than the user normally would and rank these according to the context of the search, going beyond search engine rankings. This approach involves manipulating only query results and is illustrated in Figure 2.

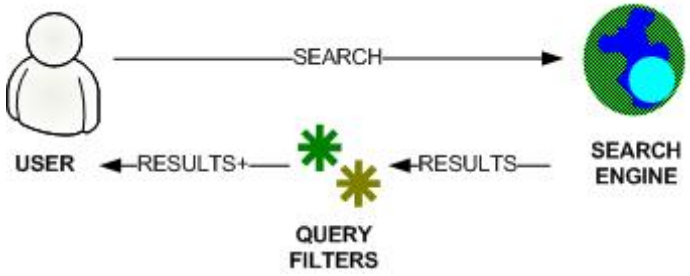


Figure 2. Query Filtering Approach

- Query Expansion: involves manipulating the query before it is submitted to the search engine, altering it through the addition of keywords or usage of advanced settings. The system would analyze the query after it is submitted by the user and alter it before submitting it to the engine. This approach is illustrated in Figure 3.

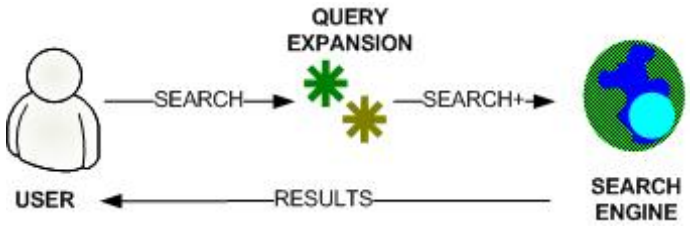


Figure 3. Query Expansion Approach

- Filter and Expansion Combination: a third approach, which combines the first two, can be tried after the first two have been developed. An expansion-filter combination might yield better results than any of the previous two alone. This is illustrated in Figure 4.

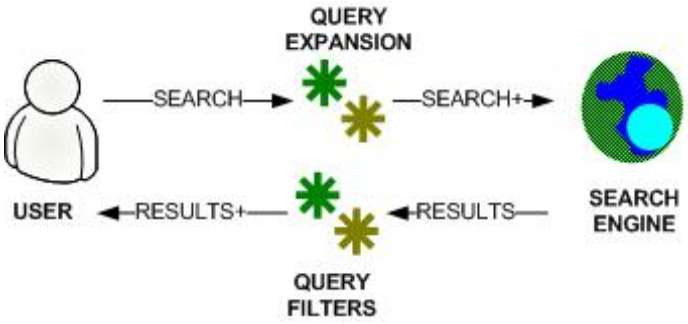


Figure 4. Combination Approach: Query Expansion and Filtering

These approaches will be measured individually for improvements in the results yielded. Our goal is to produce techniques to perform better web searches and later expand these to interaction with web content as a whole. The development of the Context Acquisition Layer is another topic for further work.

3. Discussion

We have presented an approach for the application of context in web searches. This project builds upon prior experiences with individual activity contexts developed in cooperation with a French institution. With this project, we expect not only to generate more contextually relevant search results but also to gain a better understanding of how individuals interact with information on the web and how to personalize these interactions, opening up more possibilities for research.

We are currently discussing the possibility to establish formal collaboration with a professor in France (LIP6 – Paris). This collaboration would involve using his context models as outcome for the Context Acquisition Layer and basis for reasoning on the query expansion and filtering modules. Tools already developed by his group could also be adapted as part of the Context Acquisition Layer. This would enable us to test his models and tools and explore new needs generated by this specific problem.

This project falls under the umbrella of INCT Ciência da Web (National Institute of Web Science), approved by CNPq in 2008. In addition to the professors listed above, we envision the participation of one or two graduate students in this project, which would possibly involve a post-doc or sandwich/*co-tutelle* agreement.

References

Gauch, S. Speretta, M., Chandramouli, A., Micarelli, A. (2007) User profiles for Personalized Information Access. In The Adaptive Web, Lecture Notes in Computer Science 4321. P. Brusilovsky, A. Kobsa and W. Nejdl (Eds.) pp. 54-89, 2007. Springer Verlag, Berlin Heidelberg

Lakshmanan, H. (2004) A client side tool for contextual Web search. MSc. Thesis. Massachusetts Institute of Technology, Department of Civil and Environmental Engineering, 2004.

List of the authors

COLIBRI 2009 • 22-23 de julho de 2009 • ISBN 978-85-7669-245-4

Authors