3^{rd} Workshop of

LICIA

http://licia-lab.org

October 21-22, 2013

Universidade Federal do Rio Grande do Sul
Instituto de Informática
Auditório 67

Porto Alegre

Scientific Program

---

1 Program at a glance

	Monday, October 21st, 2013
8h45	Reception
9h	Opening
9h45	GRDI Web Science, HOSCAR
10h	Coffee-break
11h	**Keynote A** Marta Mattoso (COPPE–UFRJ)
12h	Lunch-break (FAURGS)
14h	**Keynote B** Alexandre Termier (LIG)
15h	Coffee-break
15h30	**Emerging Trends** • Lucas M. Schnorr (UFRGS) • Franck Rousseau (LIG) • Christian Boitet (LIG)
17h	**LICIA Scientific Committee Meeting** Room **220** Building 65
19h30	Dinner (Location to be announced)

	Tuesday, October 22nd, 2013
9h	**Working Groups (in parallel)** • WG2 – FabLab • WG3 – Network • WG5 – Visualization • WG6 – Language Processing • WG8 – HPC
10h45	Coffee-break
11h	**Keynote C** Bruno Raffin (INRIA)
12h	Lunch-break (FAURGS)
13h30	**Working Groups (in parallel)** • WG1 – Performance, Trace Visualization & Analysis • WG7 – Distributed Systems • WG4 – WebScience
14h45	Coffee-break
15h	**Graduate Students Highlights** • Julio Toss (UFRGS-LIG) • Laercio Pilla (UFRGS-LIG) • Joao Lima (UFRGS-LIG) • Wagner Kolberg (UFRGS-LIG) • Damien Dosimont (LIG-UFRGS)
16h15	**Working Groups Results & Wrap-up**
17h	Closure
19h30	Dinner (Location to be announced)

2 Detailed program

2.1 Keynotes

- **Marta Mattoso** (COPPE–UFRJ)
  - **Location**: Keynote A, October 21st, 2013, 11h at Auditório 67
  - **Title**: Algebraic Dataflows for Big Data Analysis
  - **Abstract**: Analyzing big data requires the support of dataflows with many activities to extract and explore relevant information from the data. Recent approaches such as Pig Latin propose a high-level language to model such dataflows. However, the dataflow execution is typically delegated to a MapReduce implementation such as Hadoop, which does not follow an algebraic approach, thus it cannot take advantage of the optimization opportunities of PigLatin algebra. In this paper, we propose an approach for big data analysis based on algebraic workflows, which yields optimization and parallel execution of activities and supports user steering using provenance queries. We illustrate how a big data processing dataflow can be modeled using the algebra. Through an experimental evaluation using real datasets and the execution of the dataflow with Chiron, an engine that supports our algebra, we show that our approach yields performance gains of up to 19.6% using algebraic transformations in the dataflow and up to 39.1% of time saved on a user steering scenario.
  - **Short bio**: to be announced
- **Alexandre Termier** (LIG)
  - **Location**: Keynote B, October 21st, 2013, 14h at Auditório 67
  - **Title**: Data Mining at LIG/HADAS
  - **Abstract**: In this talk, we will give an overview of the data mining challenges tackled by the LIG/Hadas team. A major challenge for us is to broaden the range of "pattern mining" techniques (such as frequent itemset mining). In this regard, we proposed efficient generic pattern mining algorithms, and made them parallel in order to exploit the power of multicore processor for this computationnaly-intensive task. We will also present our ongoing works on large user-centric data, discussing the specificities of this data and the algorithmic challenges they pose.

- **Short bio**: Alexandre Termier got his PhD from Paris-South University, followed by postdocs in Japan at Osaka University and Tokyo's Institute of Statistical Mathematics. He then became associate professor at the University of Grenoble, LIG laboratory. His research interest include parallel pattern mining, trace analysis and interactive data exploration.

- **Bruno Raffin** (INRIA)
  - **Location**: Keynote C, October 22nd, 2013, 11h at Auditório 67
  - **Title**: Computer Graphics and Parallel Computing
  - **Abstract**: GPUs, first designed to efficiently render 3D scenes at a high frame rates, are today commonly present in supercomputers for accelerating all kind of computations. GPU's high flops numbers result from a massive internal usage of parallelism. Computer graphics have been relying on parallelism to face the needs for computation power for a long time, and conversely, computer graphics can help produce images that can convey very relevant information about the state of a parallel numerical simulations. In this talk we will show how different and complementary these two domains can be, through historical milestones and examples of recent investigations. Through this journey on the border between these domains, we will try to spot some possible future trends and opportunities for new fruitful collaborations.
  - **Short bio**: Bruno Raffin (PhD, HDR) is associate researcher at INRIA Grenoble since 2001. He has a PhD from the Université d'Orléans on parallel programming language design (1997). After a 2 year postdoc at Iowa State University he refocused his research activity on 3D high performance interactive computing. He led the development of the FlowVR middleware dedicated to parallel interactive applications, initiated and steered the multi-camera Grimage platform (`http://grimage.inrialpes.fr`) used to develop real-time full-body 3D interactions and 3D telepresence. He also works on parallel algorithms and cache-efficient parallel 3D data structures (cache oblivious mesh layouts), targeting multi-CPU and multi-GPU machines with applications to real-time physics simulations and scientific visualization. Bruno Raffin accounts for more than 50 international publications, 11 advised PhD students. He is a long time committee member of IEEE VR, the major international virtual reality conference,and steering committee member of the Eurographics Symposium on Parallel Graphics and Visualization.

2.2 Emerging Trends

- Lucas M. Schnorr (UFRGS)
  - **Location**: Emerging Trends section, October 21st, 2013, 15h30 at Auditório 67
  - **Title**: Large-Scale performance visualization analysis
  - **Abstract**: Considering the development of exascale systems, parallel and distributed applications are composed by an ever larger amount of processes to explore the extreme concurrency of such systems. This brings a series of complications to the performance analysis of these applications: the large number of trace files; the large volume of events to be analyzed, the need of a fast turn-around time considering the analysis phase, among others. This talk will present some of the challenges considering data aggregation and general guidelines to scale the analysis large-scale systems and applications. We detail how this emerging topic is appearing within LICIA and how we can built on to establish new collaborations.
  - **Short bio**: Lucas Mello Schnorr (Ph.D.) is an Associate Professor (pt-br: Professor Adjunto) at the Informatics Institut of the Federal University of Rio Grande do Sul (UFRGS). He is also part of the Graduate Program in Computer Science (PPGC), advising students in parallel and distributed computing, and computer graphics. His research interests are on data visualization techniques that can be used to analyze and understand the behavior and performance of large-scale distributed and parallel applications. Lucas used to be a research scientist in CNRS (2009 – 2013), France, working in the MESCAL team with Arnaud Legrand and Jean-Marc Vincent. He holds a Ph.D. (2009) degree in Computer Science from UFRGS, Brazil and INPG, France through a coadivising agreement. He also holds a Bachelor degree (2003) in Computer Science from UFSM, Brazil.

- Franck Rousseau (LIG)
  - **Location**: Emerging Trends section, October 21st, 2013, 16h at Auditório 67
  - **Title**: Mobile & wireless networks for the future Internet of Things
  - **Abstract**: The Drakkar group has been working on wireless networks for the last 15 years. We will present some of our activities on wireless LANs, mobility and multi-hop networking, and focus on our on-going work on WSN —wireless sensor networks. WSN and particularly low power wireless communications are one of the key enabling technologies for the Internet of Things. But this technology alone cannot solve the problems we face when trying to interconnect billions of tiny devices to the Internet, moreover when they have to operate for years on scarce energy resources, either drawn from small batteries or harvested from their environment. We will discuss some of the challenges that need to be met in order for the IoT to become a reality and try to identify possible collaboration opportunities with our brazilian colleagues.

  - **Short bio**: Franck Rousseau is Associate Professor at the Grenoble Institute of Technology - Ensimag. He received his PhD from INPG in 1999. Previously, he was a member of the technical staff at the OSF Research Institute in Grenoble and engineer at Bull. Dr. Rousseau has published papers in the fields of multimedia and networking. His current research interests are in wireless networking — mobility, ad hoc, mesh, medium access. He is actively involved in several national and European projects, as well as project with industrial partners.

- Christian Boitet (LIG)
  - **Location**: Emerging Trends section, October 21st, 2013, 16h30 at Auditório 67
  - **Title**: Emerging trends towards a multilingual semantic web
  - **Abstract**: The main challenges in multilingual NLP now are: the number of languages to handle; the difficulty of creating the very large resources needed to build even low-quality general purpose MT systems; the necessity to use linguistically and semantically motivated representations of utterances, texts, discourses and dialogues in order to get better quality in all applications; the impossibility to find (and pay) experts to handcraft HQ NLP modules. We will present the ongoing research done in the framework of the COFECUB Caméléon project and in the framework of LICIA on the extraction and representation of "polylexical expressions" (such as complex technical terms, compound predicates, and idioms). We will also mention the future directions proposed by the (submitted) AIM-west project. The Watson/QA experiment on 14/2/2011 has shown that, by indexing enough unstructured knowledge expressed in NL texts (in that case, 1M books, or 50 G words) and analysed by a state of the art expert parser (in that case, LMT slot grammar parser), and by coupling it with a formal representation of the domain of the application (ontology) and with inference mechanisms both general and tailored to the tasks at hand, one can do significantly better than human experts. Watson/QA has in fact shown that CLIR can now be replaced by the more interesting (for users) QA paradigm. A new challenge, then, is to make that architecture multilingual, arriving at a new "MLQA" paradigm. We will briefly present a project (BigTextIF) just submitted to build a recommendation system based on the processing of tweets in a multilingual context.
  - **Short bio:** Christian Boitet is codirector of GETALP (Study Group on the Translation and Processing of Languages and Speech). On graduating from École Polytechnique (Paris) in 1970, he joined the Centre National de la Recherche Scientifique (CNRS) and Prof. Bernard Vauquois' pioneering MT team (GETA) at Université Joseph Fourier in Grenoble. He has advised 49 PhD and State Doctoral Thesis students to their defense, and advises or co-advises 4 PhD students at the moment. He has been principal investigator for several industrial research contracts (about 150 contract reports). He has also participated in or led GETA's involvement in several cooperative European and international research efforts. He has been an invited researcher in several laboratories, notably TAUM (UdM, Montréal), SFB-100 (Saarbrucken), UTMK (USM, Penang), KDD (Tokyo), NII (Tokyo), and ATR Interpreting Telephony Research Laboratories, where from 1988 he became interested in speech translation and multimodal interactive disambiguation. He spent in total about 2 years in Japanese laboratories, with a sabbatical in 1992-93. He has organized workshops on Dialog-Based MT and multi-modal interactive disambiguation, co-organized the COLING-92 international conference, and chaired COLING-ACL'98 (Montréal) and COLING-2012 (Bombay).

2.3 Working Groups

Working groups will be carried-out in parallel according to people's interests.

WG1 – Performance, Trace Visualization & Analysis
Lucas M. Schnorr (UFRGS)
Jean-Marc Vincent (LIG)
Oct. 22nd, 2013, 13h30 – 14h45, Room **215** Building 65,

WG2 – FabLab
Didier Donsez (LIG)
José Palazzo Moreira de Oliveira (UFRGS)
Oct. 22nd, 2013, 9h – 10h45, Room **220** Building 65

WG3 – Network
Luciano Paschoal Gaspary (UFRGS)
Franck Rousseau (LIG)
Oct. 22nd, 2013, 9h – 10h45, Room **218** Building 65

WG4 – WebScience
Alexandre Termier (LIG)
Dante Barone (UFRGS)
Oct. 22nd, 2013, 13h30 – 14h45, Room **214** Building 65

WG5 – Visualization
Bruno Raffin (INRIA, LIG)
João Comba (UFRGS)
Oct. 22nd, 2013, 9h – 10h45, Room **105** Building 67

WG6 – Language Processing
Aline Villavicencio (UFRGS)
Christian Boitet (LIG)
Oct. 22nd, 2013, 9h – 10h45, Room **215** Building 65

WG7 – Distributed Systems
Claudio Geyer (UFRGS)
Vivien Quéma (LIG)
Oct. 22nd, 2013, 13h30 – 14h45, Room **218** Building 65

WG8 – HPC
Philippe Olivier Alexandre Navaux (UFRGS)
Oct. 22nd, 2013, 9h – 10h45, Room **104** Building 67