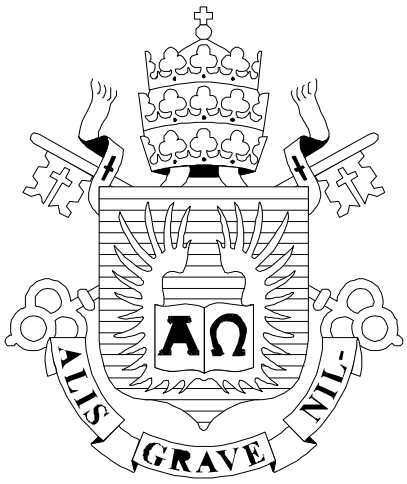

PUC

ISSN 0103-9741

Monografias em Ciência da Computação
n° 46/08

Brazilian Institute for Web Science Research

Nelson Maculan
Carlos José Pereira de Lucena
(editors)

Departamento de Informática

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO
RUA MARQUÊS DE SÃO VICENTE, 225 - CEP 22451-900
RIO DE JANEIRO - BRASIL

Monografias em Ciência da Computação, No. 46/08 ISSN: 0103-9741
Editor: Prof. Carlos José Pereira de Lucena October 2008

Brazilian Institute for Web Science Research[1]

Nelson Maculan[1], Carlos José Pereira de Lucena[2] (editors)

[1] Systems Engineering and Computer Science Program – UFRJ – Rio de Janeiro, Brazil

[2] Department of Informatics – PUC-Rio – Rio de Janeiro, Brazil

maculan@cos.ufrj.br, lucena@inf.puc-rio.br

Abstract. This technical report introduces the Brazilian Institute for Web Science Research, which will congregate 110 researchers from 10 Brazilian institutions. Investigations conducted within the Institute will range from understanding the impact the Web has on the daily lives of individuals to meeting the challenges of the Web graph. They will address the problems of developing software for Web-wide applications, of searching, retrieving and managing data stored in hundreds of millions of Web sites, and of proposing novel architectures that overcome the limitations of the current Web infrastructure.

Keywords: Web Science, People & Society, Web Applications, Web Data, Web Infrastructure.

Resumo. Este relatório técnico apresenta o Instituto Brasileiro de Pesquisa em Ciência da Web, que congregará 110 pesquisadores de 10 instituições brasileiras. A pesquisa a ser desenvolvida pelo Instituto cobre um amplo espectro, desde melhorar o entendimento do impacto da Web no quotidiano dos indivíduos a elicitar as propriedades do grafo da Web. Cobrem os problemas de desenvolver software para aplicações na Web, de pesquisar, recuperar e gerenciar dados armazenados em centenas de milhões de locais na Web, e de propor novas arquiteturas de redes de computadores que eliminarão as limitações da infraestrutura atual da Web.

Palavras-chave: Ciência da Web, Pessoas e Sociedade, Aplicações na Web, Dados na Web, Infraestrutura da Web.

[1] The authors of this technical report paper include those researchers that contributed to the proposal submitted to CNPq. The complete list can be found in the Annex.

In charge of publications:

Rosane Teles Lins Castilho
Assessoria de Biblioteca, Documentação e Informação
PUC-Rio Departamento de Informática
Rua Marquês de São Vicente, 225 - Gávea
22451-900 Rio de Janeiro RJ Brasil
Tel. +55 21 3527-1516 Fax: +55 21 3527-1530
E-mail: bib-di@inf.puc-rio.br
Web site: http://bib-di.inf.puc-rio.br/techreports/

Contents

1. Introduction

1.1. **Aims and Scope**

The World Wide Web (*the Web* for short) has had a tremendous impact on scientific research, technological development and human experience and society. The ways we communicate, collaborate and learn are being radically changed by it. As an entity, however, it remains surprisingly unstudied [Hendler et al. 2008].

This motivated the appearance, in 2006, of a new research domain – the so-called *Web Science* [Berners-Lee et al. 2006, Hendler et al. 2008]. In this new domain, the Web is the primary object of study and ceases to be viewed as a mere technology, based on computers, that helps people communicate and interact on a global basis. As such, it involves not only research on computing and technological aspects, but also on social and economic issues. More precisely, it is the science that investigates all issues around decentralized information systems, covering people, software and hardware, and their multiple, complex interactions.

Indeed, the significance of the Web depends not only on its computational properties, but also on the context of its use [O'Hara & Hall 2008]. Understanding the multiple relationships between the Web and society at a number of scales, from computer science and mathematics to emergent social behavior, has become an essential research agenda to ensure the Web's growth and contribution to society.

We hence propose the creation of the **Brazilian Institute of Web Science Research**, a reference center of excellence for state-of-art, high quality international standard research in Web Science. The mission of the Institute is four-fold:

1. To advance scientific research in a number of topics within Web Science, especially those topics that the Institute researchers have produced high-quality contributions to both the national and international scientific communities;
2. To educate qualified professionals and researchers to ensure innovation and promote interdisciplinary collaboration;
3. To transfer knowledge to the society at large through broad educational programs;
4. To transfer knowledge to the industry and service segments through specific programs, including residencies at the Institute.

The Institute will congregate 110 Ph.D. researchers from 10 Brazilian institutions, led by 6 senior researchers. The research team includes 3 recipients of the Great Cross of the Scientific Merit in Brazil, 6 recipients of the Medal of Scientific Merit in Brazil, 4 members of the Brazilian Academy of Sciences, and 55 recipients of research productivity grants from CNPq - the Brazilian Council for Technological and Scientific Development, including 7 that achieved the maximum level in the Brazilian researchers' rank. Internationally, the Institute will benefit from the collaboration with: DERI - Digital Enterprise Research Institute; L3S Research Center; The David R. Cheriton School of Computer Science, U. of Waterloo, Canada; and LIP6 – Laboratoire

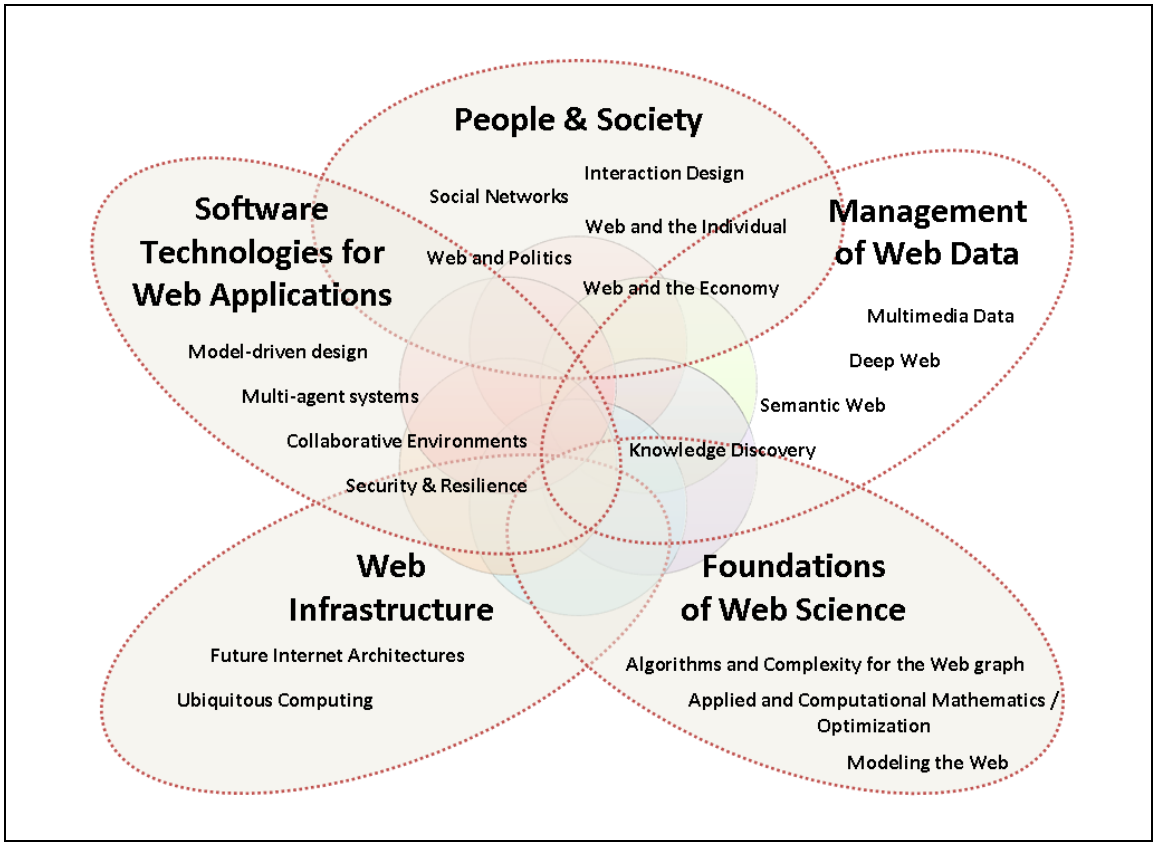

Figure 1. Web Science research program.

d'Informatique de Paris VI, France. The most prominent research effort closest to ours is the recently founded Web Science Research Initiative (WSRI) [2].

Following this worldwide trend, the Institute establishes a research program for the Web as a primary object of study. This program is organized along five axes, depicted in Figure 1. The first axis, "People & Society", will investigate the social, political and economic aspects of the Web, concerning for instance its influence in economic shifts and employment, and its role in securing the basic social values of trustworthiness, privacy, and respect for social boundaries. Studies will also deal with the need for new human interfaces – e.g., to allow the digital inclusion of elderly and semi-literate populations. This axis is intimately related with the fourth Grand Research Challenge proposed by the Brazilian Computer Society ("Participative and universal access to knowledge for the Brazilian citizen") [Medeiros 2008].

The "Software Technologies for Web Applications" axis will consider specific issues in the design, development and deployment of large distributed applications on the Web, involving millions of users. This includes investigating interaction aspects (and thus the "People & Society" axis), but also new execution paradigms, and models, tools and techniques to support novel application domains (such as those in e-Science, e-Learning and e-Engineering). As such, this axis is related with the second Grand Research Challenge of the Brazilian Computer Society ("Computational modeling of complex systems: artificial, natural, socio-cultural and human-nature interactions").

[2] http://webscience.org

The “Management of Web Data” axis will address access and management of heterogeneous, distributed data sources, from the Terabyte (10^{12}), through the Petabyte (10^{15}), to the Exabyte (10^{18}) levels. This involves combining these sources, to extract and generate new knowledge that is served to the “Software Technologies for Web Applications” axis, to the benefit of “People & Society” axis. While present Web knowledge management tools mainly deal with text, new kinds of results are needed to enhance semantics, and to access all kinds of multimedia data available. This axis contributes to efforts towards the first Grand Research Challenge (“Management of information over massive volumes of distributed multimedia data”).

The “Web Infrastructure” axis will deal with the Web as a technological means to ensure scientific, technological and societal progress, dealing primarily with the question of how to scale to meet performance or reliability expectations. It will thus advance results on areas such as computer networks, integrity and dependable computing, ensuring security in data transfer and communications, and distributed and parallel execution of the hundreds of thousands of processes needed by novel applications – therefore contributing to all other axis. It will borrow from, and contribute to research in the fifth Grand Research Challenge (“Technological development with quality: dependable, scalable and ubiquitous systems”).

Research on the “Foundations of Web Science” is needed to support research in all of the previous strata. Work conducted here will contribute, among others, to optimize the performance of systems that execute on the Web, to create mathematical models of the Web and the billions of links and interactions among its many systems (a new object of study in graph theory, called Web graph), and models that address the complex interactions between Web applications and their users.

In summary, the Institute is to play a pioneering role in this new research area, helping to understand what the Web is, to engineer its future, and to ensure its social benefit. The research to be conducted within the Institute will contribute to several scientific disciplines. Moreover, it has an enormous potential to impact society at large, firstly by achieving a larger and deeper understanding of the Web, arguably the biggest cultural revolution since the introduction of TV sets in the home sixty years ago, and secondly by engineering innovative technological tools that will further shape the Web of the future, and the way we learn ensuring its social benefit.

1.2. Proposal Structure

The proposal is structured as follows. Section 2 outlines the research program. Section 3 details the program. For each topic, it emphasizes the relevance and contributions, lists the objectives and goals, and shows a detailed plan for the first two years and a plan outline for the next three years. Section 4 lists the research team, describing its qualification and previous experience, and indicating to which topics each researcher will contribute. It also indicates the steering committee and the functional and organization structure of the institute. Section 5 describes the institute infrastructure. Section 6 covers the collaboration mechanisms. Finally, Section 7 discusses the expected benefits accrued from the project, including the educational programs and the knowledge transfer actions proposed.

2. Research Program Outline

As anticipated in Section 1, the research program comprises five axes: People & Society, Software Technologies for Web Applications, Management of Web Data, Web Infrastructure, and Foundations of Web Science (see Figure 1). These axes unfold into several topics, which in turn have specific goals, as outlined in what follows and summarized in Table 1 at the end of this section.

2.1. People & Society

Research in Web Science is multifaceted, covering all aspects in Computer Science—from device (hardware) design, through software development for managing data and services, to human-computer interaction design and evaluation and the impact of technology in the society, so much so that the expression "social machines" has been coined to define this machine-human-culture interplay.

As the world and our lives become increasingly more digital, people with special needs and low literacy skills in this world poses new challenges for both digital and social inclusion. As we all become more connected, issues of privacy and trust need to be responsibly dealt with. As we become frequent players in these systems, we want to adapt them or have them adapt themselves to better suit our needs and match our preferences.

From a more technological standpoint, people have access to different kinds of devices and user interfaces every day. New artifacts are developed in an astounding pace, requiring both design and analysis of novel user interaction techniques.

In this context, not only the systems act in concert, but also, and especially, people, forming social networks for various purposes, from keeping in touch with friends and communicating with the large audience in the Web to collaborating in "real world" projects. How these social networks emerge and whether they can be intentionally modeled with specific purposes and policies are also topics well worth of multidisciplinary investigation. Studies about the impact of social networks on society, culture and politics are also called for.

Web-mediated markets provide a valuable opportunity for economics research, both as subject of study and as a resource for unique data. Unlike other types of transactions, Web-based transactions leave a trace; in liquid markets, large amounts of individual transaction data accumulate in the databases available on the Web. These are data sources that provide the possibility of an in-depth analysis of economic behavior and the operation of market institutions.

Another important aspect of Web Science is the research on how the different social agents use the Web to achieve their goals. We intend to understand the political action of said agents, both within political parties as well as within social movements. It is believed that the Internet and in particular the Web are increasingly important in political mobilizations.

2.2. Software Technologies for Web Applications

Web applications today are very complex, and involve a understanding, specifying, designing and implementing a variety of concerns, such as specific domain models, navigation design to support a variety of tasks, interface models to support a diverse and heterogeneous user population accessing the Web through a multitude of devices, subject to a multitude to (often conflicting) security and privacy issues, often within the scope of social networks (be they intra-company or external). Furthermore, the traditional development cycle does not apply anymore (see [Hendler et al. 2008]); applications are nowadays in a permanent state of "in development", with updated and deployments occurring while the application itself is executing within the Web.

An interesting approach to the design of such applications is the model-driven approach, where the models are described using the Semantic Web formalisms. Each model addresses a specific aspect (or concern) of the application, by employing a suitable abstraction and providing a clear and concise notation to represent the model. Examples are Domain Models, Navigation Models, Interface Models, Personalization Models, Functionality Models, Security Models, etc. An associated concern is enabling reuse of existing design experience.

While traditional approaches to computer systems management are often centralized and hierarchical, today's large-scale computing systems are highly distributed with increasingly complex connectivity and interactions, rendering centralized management schemes infeasible. Multi-agent systems (MAS) approaches to software engineering typically decompose problems in terms of decentralized, autonomous agents that can engage in flexible, high-level interactions.

Autonomic Computing [Hennessy 2002, IBM 2003, Kephart et al. 2003] is a recent research theme that defines systems that manage themselves according to the goals of the administrator and without direct human intervention. The main idea is that the resources—either hardware or software — obey the self-CHOP rule, which means self-configuration, self-healing, self-optimization and self-protection, being thus able to observe and respond to changes in the environment where they are inserted. Jennings (2001) advocates an agent-based approach to software engineering. Many ideas developed in the MAS community, such as those pertaining to automatic group formation, emergent behavior, multi-agent adaptation, and agent coordination, among others, could likely be fruitfully adapted for autonomic computing.

Workflow management systems are a rich domain in which to apply autonomic computing. The constant unpredictability found in the market and in the business scenario, with new players, competitors, policies and rules, as well as unexpected problems, such as losing personnel, lack of resources, exceptions, among others, alters the execution of any process, making it necessary for the manager or administrator to update the workflow model accordingly. Taking into account that a workflow has associated sub-workflows, this update is costly. Because it is a manual task, even when domain specialists do the modeling, it becomes almost impossible to know or to remember all the rules involved in the business or foresee all flow alternatives, possible problems or incidents. The dynamic nature of processes is high and their impact in management is unquestionable.

Research areas such as e-Science, e-Learning and entertainment also present unique and interesting opportunities. Collaborative Virtual Environments (CVEs) on the Web appear as relevant environments, integrating Visualization and Collaboration, defined as simulation of real or imaginary worlds where the focus is to provide a common virtual space to distributed teams, where they can meet, co-exist and cooperate, while they interact with the environment, share information and manipulate artifacts in real time [Goebbels et al. 2003], [Raposo, Cruz et al. 2001]. Despite their increasing popularity, CVEs present a series of challenges for development.

As an increasing portion of our data lies on the Web, and Web-based enterprise applications deal with sensitive data (e.g. financial and medical data), the security and resilience of Web applications become a paramount concern. Failures and compromises, in addition to downtime, can lead to financial and personal damages. The current situation calls for better tools that help developers detecting vulnerabilities during the development cycle, such as the use of static analysis tools [Hulme, 2001][Livshits and Lam, 2005].

2.3. Management of Web Data

An important challenge to be met under the Web Science perspective is how to handle the increasing amount of multimedia data generated every second on the Web, for a wide scope of application domains – medicine, remote sensing, biodiversity, biometry and the arts, to name but a few. This has been identified by the Brazilian Computer Society as one of the five Grand Challenges in computer science in Brazil. The very many problems to be dealt with include, for example, the design of new data structures, indexing schemes, and query and retrieval mechanisms. Not only are there large volumes of data, one has now to deal with video, sound, images, to be processed in real time, across a multitude of sites. Tools and services are breaking domain boundaries and local limits as a consequence of progressive data and protocol interoperability, allied to an increasing connectivity.

The Web comprises both *static pages*, typically pre-built HTML pages, and *dynamic pages*, which are built from multimedia data stored in databases and other types of objects as they are requested by a user or an application. These two sets of pages comprise what is now called the “Surface Web” and the “Deep Web”, respectively. Estimates suggest that the size of the Deep Web greatly exceeds that of the Surface Web – with nearly 92,000 terabytes of data on the Deep Web versus only 167 terabytes on the Surface Web, as of 2003.

In particular, the information in Deep Web databases cannot be efficiently discovered through the simple traversal of hyperlinks, but rather require both people and applications to interact with (potentially) complex query interfaces. Hence the need for *query mediators*, software specifically designed to locate Web-enabled databases and to interact with their query interfaces, so as to uncover Deep Web databases and bring together large scales of information in a coherent, organized fashion, to help people make sense of and use the knowledge on the Web to solve their problems and achieve their information-related goals.

The Web has evolved until recently as a medium for information exchange among people, rather than machines. As a consequence, the semantic content, that is, the meaning of the data in a Web page, is coded in a way that is accessible to human beings alone. Enter the "Semantic Web", where ontologies will be responsible for making the semantics of pages and applications explicit, thus allowing electronic agents to process and integrate resources automatically. This society of knowledge-exchanging machines must now be dynamically orchestrated so as to provide useful and timely services to both machines and people in a secure and resilient fashion.

In the frontier between the Semantic Web and the more textual Web lie Machine Learning applications that draw on Computational Linguistics to extract semantic data that machines can process from the textual data written and maintained by human beings in a regular basis. The exponential growth of the total volume of contents accessible through the Web enables a new generation of innovative applications based on Machine Learning. In addition, the low-cost computational infrastructure based on new paradigms such as the Parallel External Memory (PEM) paradigm facilitates the scalability of learning algorithms, in an approach successfully employed by web-savvy companies such Google and Netflix.

2.4. Web Infrastructure

Since the formulation of the current IP technology almost 30 years ago, incremental changes have been made to extend its applicability to the demands which result from an ever increasing population of users, information resources and applications. The user population is now numbered in billions, and the major contribution to traffic comes from Web applications of increasing diversity.

In spite of this success, the basic IP technology contains the seeds of its own limitations, which are becoming more evident every day. The identical treatment for all information flows within the Internet at the IP packet level is neither desirable, nor particularly cost-effective, especially when certain classes of application, such as those involving interactive media, or remote access to scientific instruments require quality of service (QoS) guarantees which are unnecessary for most other applications. It is commonly believed that the solution to many, if not most, of today's problems will involve a fundamental redesign of the current Internet architecture, based on IP, and a major objective of the research on the topic of Web Infrastructure is the formulation and evaluation of alternative architectures to replace it.

Moreover, the Web is being increasingly used by mobile users. Nowadays, most such uses require a map-based Web application such as Google-maps, in which users can post photographs or textual descriptions associated with a particular geographic spot. It becomes important to enable the automatically upload of any kind of user-generated content (e.g. a voice recording, a video-clip, or a picture taken with the user's device) with a seamless association with the current user's position. Moreover, the user should be able to tag the geographic position (and its contributions) with semantic information, such as her current impressions, emotions, plans, etc. The Web will then consist of a Ubiquitous Location-aware Web (ULW), enabling new forms of sharing, communication and collaboration.

The use of the Web for sharing location-related information, especially by mobile users, opens a wide range of possibilities for mobile social networking, location-aware searching, resource sharing, workforce coordination and entertainment, as well as areas with direct economic and social impact such as: public security, transportation, logistics, tourism, health care, m-learning, entertainment, and e-government, to name but a few. However, in order to explore these new opportunities for user communication and collaboration, new services, middleware infrastructures, mobile applications and possibly even new mobile devices have to be developed. Hence, another major goal in the topic of Web infrastructure is to develop and experiment with innovative middleware services and applications for the Ubiquitous and Location-aware Web.

2.5. Foundations of Web Science

With the advent of the Internet, a new area emerged within theoretical computer science. This new field, algorithmic foundations of the Internet, was conceived to tackle the problems arisen from the need to deal with very large networks. It appeared during the first half of the nineties, coinciding with the popularization of the worldwide web. At present, the field consists mainly of topics from graph theory, algorithms, game theory, distributed computing, statistical physics, economics, information retrieval and probability theory, but topics have continuously been aggregated as new aspects of the field come to attention.

Therefore, optimization techniques are indispensable for an effective organization of the Web, thus constituting an essential part of Web Science. In order to contribute to the analysis of the Web and to the synthesis and enhancement of its mechanisms, the optimization methods must improve their algorithms as well as be able to make a quasi-perfect use of the most recent computational technologies: huge networks of heterogeneous computers. Speed enhancement in database information search and set covering problems are formulated as combinatorial optimization problems and models.

Modeling and understanding the Web, this vast, complex and highly dynamic system is also a challenge that requires a combination of different expertise and the development of new techniques. In addition, modeling the complex interactions between Web applications and users, as well as studying Web systems robustness and scalability are extremely import problems to address.

Table 1. Summary of Topics and Goals per Layer.

People & Society	
Interaction Design	G1. Study the fundamental properties of Web applications as social machines and the theories that could represent a foundation for the interaction design G2. Develop tools to support Web interaction and navigation for users with special needs such as visual impairment and low levels of functional literacy skills. G3. Develop methods and tools to support the development of interaction mechanisms that make the information sharing explicit and in accordance to the social policy expectation of a community of users. G4. Study the pragmatics of human interaction and communication throughout the Web to understand and explain Web languages, based on semiotics and semiotic engineering, but also linguistics, psychology and philosophy G5. Investigate current tools and interaction paradigms for empowering users to adapt and extend content and services on the Web G6. Study non-speech sound interfaces to improve interactions in the Person-to-Person Computer-Mediated Communications G7. Investigate new interaction techniques and paradigms on the Web G8. Investigate the socio-pragmatics of human interaction with information through Web application, the social and legal rules governing information uses.
Social Networks	G1. Ethnographic Analyses to Understand the social Web G2. Definition of models for social networks analysis and mining in order to discover patterns, needs and requirements for collaboration support G3. Definition of mechanisms for knowledge discovery and organization from narrative social content G4. Configuration and conceptualization of digital heritage in social networks G5: Analyze and balance social networks G6. Develop tools to support intentional social networks G7: Incorporate autonomic characteristics into the control of social networks G8: Investigate scientific social networks G9: Investigate temporal and mobile social networks
Web and the Individual	G1. Mapping the field G2. Data collection
Web and Politics	G1. Analysis of the use of the Web in Brazilian elections G2. Analysis of the Web use by social movements
Web and the Economy	G1. Development of Internet auction data collecting software. G2. Development and application of econometric techniques to investigate Internet auction data. G3. Development of a decision support framework based on data from the Semantic Web G4. Framework instantiation for the capital markets domain

Software Technologies for Web Applications	
A Multi-Agent Systems Approach for Developing Autonomic Web Applications	G1. Design techniques to improve the development of autonomic Web applications G2. Define a framework to support the development of autonomic Web applications
Model-driven Design and Implementation of Web Applications	G1. Define a model-driven framework to support specification, design and implementation of Web applications, seen as part of men-machine teams. G2. Develop case studies (software product lines) for the Web context G3. Develop methodologies, empirical studies and tools to support the development of software product lines for the Web context
Design and Implementation of Autonomic Workflows	G1: Investigate and prepare the scenario G2: Develop the autonomic element G3: Integrate sub-workflows G4: Develop autonomic data approach G5: Specialize solution for Web services G6: Analyze the applicability of the solution in other scenarios
Security and Resilience of Web Applications	G1. Define static analysis techniques supporting development approaches for detecting possible threats to application security G2. Propose a development approach to improve Web-application resilience to faults, enabling them to coexist with faults
Web-based Collaborative Virtual Environments	G1. Investigate the Web as a means for eScience and eLearning through CVEs G2. Interactive storytelling for Web-based collaborative virtual worlds G3. Interactive storytelling for cultural content production and e-learning in the Web
Management of Web Data	
Managing Web Multimedia Data	G1. Definition of new multimedia data descriptors and of storage structures to support their indexing, reuse and composition on the Web G2. Design and development of mechanisms to annotate multimedia data, and to index and retrieve such annotations, having in mind educational uses on the Web
Accessing the Deep Web	G1. Definition of heuristics for database identification, mediated schema definition and schema matching G2. Development of query mediators
Developing the Semantic Web	G1. Definition of a method to develop reference domain ontologies based on the composition and alignment of existing (upper level) ontology fragments G2. Definition of a method for the description, discovery and automated composition of semantic Web services G3. Design techniques to improve the resilience of service-oriented architectures based on Web services G4. Define a framework to support the development of dynamic and resilient Web applications G5. Design and development of a specification and a model to

	evaluate, and personalize search results according to users' quality level perspectives
Knowledge Discovery on the Web	G1. Develop learning algorithms for classification and regression problems G2. Develop machine learning frameworks with EPM implementation for large volumes of data
Web Infrastructure	
Future Internet Architectures	G1. Design and deploy a testbed for R&D in network architectures G2. Explore new design approaches aiming at Web Applications G3. Design and implementation of a measurement infrastructure for the testbed G4. Design, deployment and use of a wireless testbed for research on the impact of mobility on web applications
Ubiquitous and Location-aware Web	G1. Develop prototype location-aware Web services G2. Develop frameworks and middleware services that ease the development of such ULW applications
Foundations of Web Science	
Algorithms and Complexity for the Web graph	G1. Define algorithms in graphs, and / or negative results for combinatorial problems
Applied and Computational Mathematics / Optimization	G1. Implementation of a continuous optimization library with and without constraints. G2. Development of new methods for implicit enumeration in combinatorial optimization, column generation and cutting planes. G3. Implementation of metaheuristics aiming to search for lower bounds (in maximization) and upper bounds (in minimization).
Modeling the Web	G1. Study robustness/scalability issues of P2P systems and develop robustness models for those. G2. Development of methods, tools and techniques for e-learning G3. Development of algorithms for P2P systems G4. Model the future Web graph

3. Research Program

In this section, we detail the research topics including, for each topic, the topic description, its goals, activities, expected results and deliverables in the project timeframe, the multi-institutional team that will address the topic, and their qualifications.

3.1. People & Society

3.1.1. Interaction Design

The Web is being considered one of the most influential creations in the history of Information and Communication Technology. It has impacted life in society transforming the way academia, industry and government work, relate to knowledge and to people in their everyday life. Considering the Web as an object of study in its own right, demands understanding Web design in a context of scalable and open applications development, its architectural and data requirements, and enabled interaction. Interaction design is a research topic which requires understanding the Web as a computational structure which shapes interaction with and among its users.

Together with its own expansion and evolution, the Web has changed the way people achieve many tasks. When faced with a problem to solve, a person may, typically, try to solve it on her own. Before the advent of the Web, she would perhaps look up some references in a library, or consult friends and co-workers, in writing or by phone. Nowadays, a frequent approach is to search for the information on the Web, either by looking up in a search engine, or by interacting with other persons, through the Web. In many instances, part of the tasks to be performed is executed by a program running in some Web server; the person uses the results of this processing as part of her solution to the problem. In other situations, the solution is achieved through collaboration between several persons, supported by an application running on one or more Web servers. Yet another possibility is the collaboration between several persons, supported by applications, whose work is integrated with the processing of several applications that also collaborate to achieve the final result.

A Web Application must be understood as a “social machine”, in that it includes an underlying technology, but also the rules, policies and organizational structures used to manage technology. For example, *Wikipedia* is built on top of *mediaWiki*, free software available and used by other Web applications besides *Wikipedia*. Nevertheless, not all *mediaWiki*-based applications succeeded in the same way; the success or failure of the Web applications seems to rely on the social features enabled by the interaction of the users and the communities they constitute through the whole system. The ways people will interact and transform the usage of a particular technology is not easy to explain and to predict with the traditional methods and frames of reference we have today for software development; a better understanding of the features and functions of the social aspects of the systems are needed.

In the general case, one can think of men-machine teams that collaborate to achieve tasks. Within such teams, there is formal and informal knowledge that is

shared between the participants. Informal knowledge is processed by humans, whereas formal knowledge is processed by computers. While computers cannot directly process informal knowledge, they can still support human teams in sharing informal knowledge, as well as help them to integrate this informal knowledge with the formal knowledge manipulated by the computer.

The various types of communication that occur within such teams can be summarized in Figure 2.

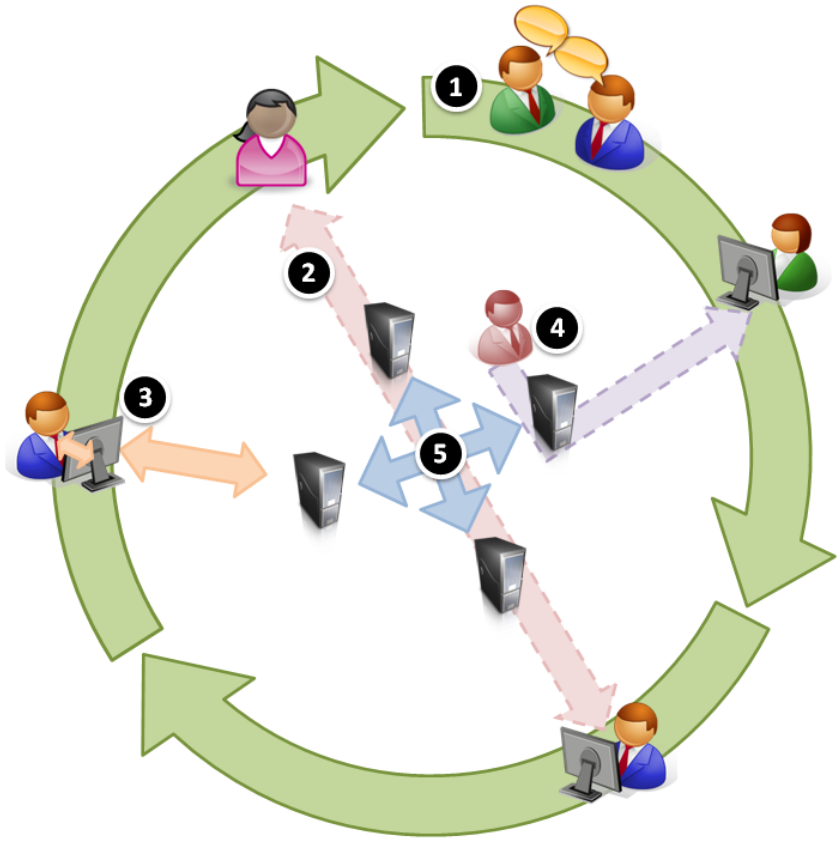

Figure 2. Perspectives of interaction on the Web.

Figure 2 indicates the various types of communications that can occur:

1. Person-to-Person Direct Communication – This includes all communications between people in which there is no involvement of the computer – face to face, telephone etc. Notice that we are not including here cases in which the communication leaves some trace that may be processed by the computer.
2. Person-to-Person (or People) Computer-Mediated Communication (CMC) – This includes all communication between persons that occurs through the computer. This implies that there must be some computer-processable representation of the entire communication process – the actors, media and content involved. This category includes a vast number of applications, such as blogs, wikis, bulletin board systems, instant messaging, email, interactive digital TV, digital phones etc.
3. Person-to-Application (or Human-Computer) Communication – This includes all interactions between persons and the applications, when manipulating and integrating the parts of the problem that are processed by the application and those processed by the human being. Examples of this communication are query and

search applications, special purpose applications such as e-stores, scientific applications etc.

4. Designer-to-User Communication – Within the theory of semiotic engineering [de Souza 2005], user interfaces consist of designer-to-user metacommunication artifacts. Bringing the designer into the picture allows us to shed new light on HCI and CMC phenomena and provide new methods for evaluation and design of computational artifacts.
5. Application-to-Application Communication – This includes all interactions between programs, without human participation. Examples are Web services, grid applications, distributed systems etc.

We may refer to the range of these perspectives as Web-Human Interactions (WHI). We suppose that there should be theoretical foundations that could explain the limitations of Web languages due to the different nature of humans and computers. An important goal is therefore to investigate the dynamics of design, construction and evolution of code and languages in the social context of Web application.

Within this framework, some issues are worthy of investigation in Web Science, such as the design and implementation of Web applications to support men-machine teams. Here we find a range of applications from traditional Web sites to Semantic Wikis, which support non-structured, semantically annotated content produced and shared by groups of users. An interesting approach to the design of such applications is model driven, where the models are described using the Semantic Web formalisms. Each model addresses a specific aspect (or concern) of the application, by employing a suitable abstraction and providing a clear and concise notation to represent the model. Examples are Domain Models, Navigation Models, Interface Models, Personalization Models, Functionality Models, Security Models, etc. An associated concern is enabling reuse of existing design experience. As the wiki evolves and increasingly more semantic data is available, the application starts to become more similar to structured, database-backed applications. This is an example of collective, user-generated content which enables the emergence of structure that can be integrated into the existing models.

A particular kind of application, which does not fit the classical model of existing Web applications today, is one that supports *knowledge discovery* by human beings. As the Web grows, there are vast amounts of encoded knowledge (e.g., Semantic Web data) that is available for humans to use in solving their problems. The difficulty arises when they must discover what the relevant knowledge for a given task is and where it is available. Traditional search interfaces are not satisfactory, first because they require the user to state exactly what she is looking for, which is often not well defined enough to enable effective search. Second, since current approaches are mostly syntactically based, they miss many relevant results, or return irrelevant results (e.g., Paris the person, Paris the city or Paris the perfume?). Another type of interface, such as faceted search, allows the user to incrementally refine her query, while examining intermediate results at each step. Starting with the entire search space, the user applies successive filters, which reduce the set of candidates to those that satisfy all the filters specified so far. Eventually, the set is reduced to a manageable size where the user can directly inspect each remaining item. We are investigating exploration metaphors that allow the user to combine search, faceted navigation and traditional navigation when exploring Semantic Web data. This metaphor is realized in the interface through a mix

of direct manipulation and command interfaces. In this context, it is easy to see that the investigation of such applications is not limited to the People & Society layer, but goes across the Applications and Data strata as well.

Going further in the goal of empowering users, we find on the Web a large potential for end-user development in the form of *mashups* and other forms of scripting. Mashups are made possible by the application programming interfaces (API) provided by Google, Flickr, Yahoo, and YouTube, among others. Scripting on the Web is provided by plug-ins such as CoScripter [Leshed 2008, de Souza & Cypher 2008]. In this context, it is interesting to investigate how users employ the tools that are currently available to adapt and extend the content and services on the Web to better meet their needs, looking for patterns of usage and exploring the novel design issues brought about by this kind of user empowerment.

Web languages have evolved from textual-based to graphical and multimedia. What lessons can be learnt about them that could help the design of new languages in the Web context? How can we design new languages to improve Web-Human Interactions? We are also interested in understanding how multi-modal interfaces can enhance the Person-to-Computer communication. In particular, non-speech sound is a modality which has received little attention so far, but has great potential, especially when considering access in mobile environments. Likewise, Web languages and Web scripts can be explored to support navigation and interaction of people with special needs, such as visually impaired users and users with lower functional literacy skills.

With the development of virtual reality, computer vision and game technologies, several types of interaction and output devices were created and are at each day more accessible to the mass use at the Web, augmenting the possibility to realize practical experiments with innovative devices, such as optical trackers, touch tables, among others. There is a category of interaction devices that is becoming more popular due to the increasing availability of digital cameras [Stefani et al. 2003]. This kind of optical device uses the image captured by a set of digital cameras to track the position of one or more objects and provide position and orientation information to the application. The Wiimote is an example of optical device increasingly used in many applications

Another form of interaction to be studied are the hybrid interfaces. The concept of Hybrid User Interfaces was initially approached by Feiner and Shamash [Feiner and Shamash 1991] characterizing a heterogeneous environment, rich in terms of interaction techniques and composed by different devices types used in a complementary and advantageous way. Interfaces like that could end up with different interaction environments (Real World, WIMP, Augmented Reality, Virtual Reality) and then some sort of transition between them could be necessary (transitional interfaces [Butz et al. 1999]). The seamless transition between those environments would improve interaction continuity by providing focus on the task.

This perspective brings about the goal to study how these new interaction devices and paradigms will affect Web application and their users.

Research in interaction design needs to investigate and articulate Web applications by at least two perspectives: as a social phenomenon in its macro scale and as an artifact to be engineered in its micro scale. As a consequence of the micro and

macro dimensions of the Web, the development cycle Computer Science has traditionally viewed as software engineering best practice (specify, design, build, test) is being redefined for Web applications development. In the same way, concepts and practices within the tradition of the Human-Computer Interaction field, such as usability and accessibility, needs to be put into perspective in this new scenario of development and the social phenomena of the Web. The essence of how to develop Web applications considering the micro and macro dimensions requires, besides the analysis of technological issues, consideration of the social dynamics enabled to potentially millions of users and its profound potential influence on social structures, i.e. political and educational systems, commercial organizations, etc.

Goals, Activities, and Deliverables

G1. Study the fundamental properties of Web applications as social machines and the theories that could represent a foundation for the interaction design

Activities:

- Elaborate a survey on the evolution of Web applications and their underlying technologies with a focus on the types of social interaction they enable;
- Analyze the limitations and the values that particular theories could bring to the process of Web application interaction design;
- Define architectural principles and conceptual model to guide the articulation of the micro and macro dimensions of Web development regarding interaction design;
- Elaborate scientific papers.

Deliverables:

- international journal paper
- international conference paper
- national conference paper

G2. Develop tools to support Web interaction and navigation for users with special needs such as visual impairment and low levels of functional literacy skills.

Activities:

- To survey existing studies about the challenges of using the Web for visually-impaired users and users with low functional literacy skills.
- To explore the use of CoScripter as a basis for Web-navigation aiders for this population of users.
- To develop working prototypes and carry out empirical studies.

Deliverables:

- proofs of concept that implement proposed solutions
- national conference paper

international conference paper

international journal paper

G3. Develop methods and tools to support the development of interaction mechanisms that make the information sharing explicit and in accordance to the social policy expectation of a community of users.

Activities:

Elaborate a survey on interaction mechanisms of new Web applications;

Develop an efficient engineering of interaction components for the social software in the Web; in particular norm-based and presence awareness mechanisms;

Select existing Web applications for analysis and experimentation;

Develop proofs of concept to experiment solutions motivated by the selected Web applications;

Evaluate the proposed methods and tools

Elaborate scientific papers

Deliverables:

proofs of concept that implement proposed solutions

national conference paper

international conference paper

international journal paper

G4. Study the pragmatics of human interaction and communication throughout the Web to understand and explain Web languages, based mostly on semiotics and semiotic engineering, but also linguistics, psychology and philosophy

Activities:

Analyze the evolution of WHI throughout the Web in the last 20 years – languages and code for design, interaction and communication to investigate how they were created and used. This topic should investigate:

- Web application design languages
- Web-User interaction languages
- Human-Human communication languages throughout the Web

Identify theoretical foundations to explain the phenomena of Web languages creation and evolution in a social context.

Define hypothesis about Web Languages to unify all perspectives of WHI.

Develop tools to design and use Web languages in some perspectives of WHI.

Realize experiments to validate the hypothesis using the Web language tools.

Propose new Web languages

Realize experiments to evaluate the proposed languages.

Deliverables:

- M.Sc. dissertations
- D.Sc. Theses
- Papers in international and national conferences
- Tools for Web language design

G5. Investigate current tools and interaction paradigms for empowering users to adapt and extend content and services on the Web

Activities:

- Study of mechanisms and models to support collective production of content and emergent structured models and allow smooth progress and integration between the unstructured, the partially structured and the fully structured portions of users' and designers' provided models and content.
- Study existing programming by demonstration concepts, methods, techniques, and tools
- Study existing mashup-enabling APIs
- Prepare and conduct empirical studies with CoScripter and mashup-related applications
- Analyze the results of the empirical studies
- Draw conclusions from the results and the semiotic engineering theory of human-computer interaction
- Redesign CoScripter based on the study results and proposing the incorporation of additional extension mechanisms
- Develop a revised version of CoScripter
- Evaluate the proposed extensions to CoScripter

Deliverables:

- Formal specifications of the methods and models, available as processable documents from servers at the institute.
- Public domain tools supporting the models and generating final running applications, available for download from the Web site of the institute
- Example applications publicly available for download
- Redesigned CoScripter prototype
- International and national journal and conference papers describing the studies, the redesign, and extension proposal
- M.Sc. dissertations and D.Sc. theses

G6. Study non-speech sound interfaces to improve interactions in the Person-to-Person Computer-Mediated Communications

Activities:

- Study of existing theories, models and tools related to non-speech sound in Web (and other) interfaces.

Development of a model, expressed as ontologies, capable of characterizing non-speech sound with respect to its relations to visual information, and its roles in interactions.

Design and implementation of an environment (or integration with existing one) to support the specification and use of non-speech sound in Web applications

Design and implementation of example applications incorporating non-speech sound, generated using the models proposed, with the support of the development environment

Evaluation of these application with respect to communicability, usability and user performance

Deliverables:

Formal specifications of the methods and models, available as processable documents from servers at the institute.

Public domain tools supporting the models and generating final running applications, available for download from the Web site of the institute

Example applications publicly available for download

Journal and conference papers, describing the methods, models, tools and user studies conducted

M.Sc. dissertations and D.Sc. theses, describing the methods, models, tools and user studies conducted

G7. Investigate new interaction techniques and paradigms on the Web

Activities:

Study virtual reality, computer vision and game technologies

Investigate new interaction and output devices

Investigate hybrid user interfaces

Develop hybrid user interface prototypes

Plan and conduct experiments to evaluate the prototypes

Elaborate models to further the development of hybrid user interfaces

Deliverables:

Hybrid user interface prototypes publicly available for download

Specifications of the design and development models

Example applications publicly available for download

Journal and conference papers, describing the interaction techniques and paradigms, prototypes, and conducted experiments

M.Sc. dissertations and D.Sc. theses

G8. Investigate the socio-pragmatics of human interaction with information through Web application, the social and legal rules governing information uses.

Activities:

Elaborate a survey on ways of representing and reasoning over attributes such as trustworthiness, reliability, privacy, copyright and other formal and informal norms;

Develop structures to formally represent them and compute over them;

Select Web applications for experimentation;

Analyze experimentation results;

Elaborate scientific paper

Deliverables:

conference paper

international journal paper

Timeframe

		Y1		Y2		Y3	Y4	Y5
Goal	Activity	S1	S2	S3	S4			
Properties of social machines and Theories for foundation of ID	Elaborate a survey on the evolution of Web applications and their underlying technologies with a focus on the types of social interaction they enable	X						
	Analyze the limitations and the values that particular theories could bring to the process of Web application interaction design		X					
	Define architectural principles and conceptual model to guide the articulation of the micro and macro dimensions of Web development regarding interaction design			X				
	Elaborate scientific papers		X	X	X	X	X	X
Develop tools to support Web interaction and navigation for users with special needs such as visual impairment and low levels of functional literacy skills.	To survey existing studies about the challenges of using the Web for visually-impaired users and users with low functional literacy skills.	X	X					
	To explore the use of CoScripter as a basis for Web-navigation aiders for this population of users.		X	X				
	To develop working prototypes and carry out empirical studies.				X	X		
Methods and tools to support the development of social machines interaction mechanisms	Elaborate a survey on design processes and interaction mechanisms of new Web applications			X				
	Develop an efficient engineering of interaction components for the social software in the Web				X			
	Select existing Web applications for analysis and experimentation				X			
	Develop proofs of concept to experiment solutions motivated by the selected Web applications					X		

	Evaluate the proposed methods and tools					X	X	
	Elaborate scientific papers				X	X	X	X
Study the pragmatics of human interaction and communication through the Web to understand and explain Web languages	Analyze the evolution of WHI throughout the Web in the last 20 years – languages and code for design, interaction and communication to investigate how they were created and used. This topic should investigate: Web application design languages User-Web interaction languages Human-Human communication languages through the Web	X	X	X	X			
	Identify theoretical foundations to explain the phenomena of Web languages creation and evolution in a social context		X	X	X			
	Define hypothesis about Web Languages to unify all perspectives of WHI		X	X	X			
	Develop tools to design and use Web languages in some perspectives of WHI					X	X	
	Realize experiments to validate the hypothesis using the Web language tools					X	X	
	Propose new Web languages						X	X
	Realize experiments to evaluate the proposed languages							X
Investigate current tools and interaction paradigms for empowering users to adapt and extend content and services on the Web	Study of mechanisms and models to support collective production of content and emergent models in Web applications, which integrate smoothly with existing underlying models		X	X	X	X	X	
	Study existing programming by demonstration concepts, methods, techniques, and tools	X						
	Study existing mashup-enabling APIs	X						
	Prepare and conduct empirical studies with CoScripter and mashup-related applications		X	X				
	Analyze the results of the empirical studies			X				
	Draw conclusions from the results and the semiotic engineering theory of human-computer interaction			X	X			
	Redesign CoScripter based on the study results and proposing the incorporation of additional extension mechanisms					X		
	Develop a revised version of CoScripter					X	X	
	Evaluate the proposed extensions to CoScripter							X
Study of the use of non-speech sound	Study of existing theories, models and tools related to non-speech sound in Web (and other) interfaces	X	X					

in Web applications	Development of a model, expressed as ontologies, capable of characterizing non-speech sound with respect to its relations to visual information, and its roles in interactions		X	X			X	X
	Design and implementation of an environment (or integration with existing one) to support the specification and use of non-speech sound in Web applications			X	X	X		
	Design and implementation of example applications incorporating non-speech sound, generated using the models proposed, with the support of the development environment				X	X	X	
	Evaluation of these application with respect to communicability, usability and user performance					X	X	X
Investigate new interaction techniques and paradigms on the Web	Study virtual reality, computer vision and game technologies	X						
	Investigate new interaction and output devices	X	X					
	Investigate hybrid user interfaces		X	X				
	Develop hybrid user interface prototypes			X	X	X	X	X
	Plan and conduct experiments to evaluate the prototypes				X	X	X	X
	Elaborate models to further the development of hybrid user interfaces					X	X	X
Socio-pragmatics of human interaction through Web applications	Elaborate a survey on ways of representing and reasoning over attributes such as trustworthiness, reliability, privacy, copyright and other formal and informal norms				X			
	Develop structures to formally represent them and compute over them					X		
	Select Web applications for experimentation					X		
	Analyze experimentation results						X	
	Elaborate scientific papers					X	X	X

Team

Name	**Affiliation**	**Degree**	**CNPq level**
Alberto Barbosa Raposo	PUC-Rio	D.Sc., UNICAMP, 2000	–
Clarisse Sieckenius de Souza	PUC-Rio	Dr., PUC-Rio, 1987	1C
Cláudia Maria Bauzer Medeiros	UNICAMP	PhD, U Waterloo 1985	1B
Daniel Schwabe	PUC-Rio	PhD, UCLA,	1B
Hugo Fuks	PUC-Rio	PhD, Univ. of	1D

		London, Imperial College, 1991	
Jair Cavalcanti Leite	UFRN	D.Sc., PUC-Rio, 1998	—
Marcelo Gattass	PUC-Rio	PhD, Cornell U., 1982	1A
Maria Cecilia Calani Baranauskas	UNICAMP	Dr., UNICAMP 1993	—
Ricardo da Silva Torres	UNICAMP	Dr., UNICAMP 2004	2
Simone Diniz Junqueira Barbosa	PUC-Rio	D.Sc., PUC-Rio, 1999	2

Recent publications

BARANAUSKAS, M. C. C. ; BONACIN, R. . Design - Indicating Through Signs. MIT Press Design Issues, v. 24, p. 30-45, 2008.

BARANAUSKAS, M.C.C., LIU, K., CHONG, S. Website Interfaces as Representamina of Organisational Behavior. In H.W.M. Gazendam, R.J.Jorna, R.S.Cijsouw (eds) Dynamics and Change in Organization, Dordrecht: Kluwer Academic Publishers, 2003, pp. 63-88.

CARVALHO, F. G., DAM, P., SOARES, L. P., RAPOSO, A. B., CORSEUIL, E. T. L., SANTOS, I. H. F. Virtual Presence Mixing 3D Video and Interactive Scenes. **X Simposium on Virtual and Augmented Reality – SVR 2008**, p. 307-312. João Pessoa, PB, Brasil.

CARVALHO, F. G., RAPOSO, A. B., GATTASS, M. An Approach for Enabling the Use of Immersive Virtual Reality in Desktop Hybrid Interfaces. **Proceedings of the Second IASTED Int. Conf. on Human-Computer Interaction**, p. 68-73. Chamonix, France, 2007.

CARVALHO, F. G., RAPOSO, A. B., GATTASS, M. Uma Interface Híbrida para Desktop Integrando Realidade Virtual, Realidade Aumentada e 2D WIMP. **IX Simposium on Virtual and Augmented Reality – SVR 2007**, p.152-161. Petrópolis, RJ, Brasil, 2007.

ESCALONA, M.J. ; TORRES-ZENTENOo, A. ; GUTIERREZ, J. ; MARTINS, E. ; TORRES, R.S. ; BARANAUSKAS, M. C. C. . A DEVELOPMENT PROCESS FOR WEB GEOGRAPHIC INFORMATION SYSTEM. In: Proceedings of the Tenth International Conference on Enterprise Information Systems, ICEIS 2008. Barcelona. v. HCI. p. 112-117.

HABIB, J H., RAPOSO, A. B. ARMsg: Mensageiro Digital usando Realidade Aumentada. **X Simposium on Virtual and Augmented Reality – SVR 2008**, p. 90-94. João Pessoa, PB, 2008.

KOLARIC, S., RAPOSO, A. B., GATTASS, M. Direct 3D Manipulation Using Vision-Based Recognition of Uninstrumented Hands. **X Simposium on Virtual and Augmented Reality – SVR 2008**, p. 212-220. João Pessoa, PB, Brasil.

LOAIZA, M. E., RAPOSO, A. B., GATTASS, M., An Interaction Device using Computer Vision Techniques. **VIII Symposium on Virtual Reality – SVR 2006**, p.259-270. Belém, PA Brasil, 2006.

LOAIZA, M. E., WAGNER, G. N., BASTOS, T. A., RAPOSO, A. B., GATTASS, M., Sistemas de Rastreamento de Baixo Custo para Desktops. **XXXII Seminário Integrado de Software e Hardware – SEMISH 2005**, **Anais do XXV Congresso da Sociedade Brasileira de Computação,** p. 1676-1690. São Leopoldo, RS, Brasil, 2005.

LOAIZA, M., RAPOSO, A. B., GATTASS, M. A Novel Optical Tracking Algorithm for Point-Based Projective Invariant Marker Patterns. **3rd International Symposium on Visual Computing – ISVC 2007**, Lake Tahoe, Nevada/California, EUA. Advances in Visual Computing - Lecture Notes in Computer Science, vol. 4841, p.160-169. Springer-Verlag, 2007.

PEREIRA, A., RAPOSO, A. B., FUKS, H. Ambiente Legal em TAMPO: Aprendizagem Colaborativa em Educação Infantil. **XVIII Simpósio Brasileiro de Informática na Educação – SBIE 2007** (CD-ROM). São Paulo, SP, Brasil, 2007.

SANTANA, V.F. ; BARANAUSKAS, M. C. C. . A Prospect of Websites Evaluation Tools Based on Event Logs. In: IFIP 20th World Computer Congress, 1st TC 13 Human-Computer Interaction Symposium (HCIS 2008), 2008, Milão. Book Series IFIP International Federation for Information Processing. Boston : Publisher Springer Boston, 2008.

SIMONI, C. A. C. ; MELO, A. M. ; BARANAUSKAS, M. C. C. . Towards a Social-based Process for Information System Development: a Case Study. In: Pierre-Jean Charrel; Daniel Galarreta (eds.) Project Management and Risk Management in Complex Projects: Studies in Organizational Semiotics. Heidelberg: Springer, 2007, p. 177-192.

TORRES, Ricardo da Silva ; SILVA, C. ; MEDEIROS, C. M. B. ; ROCHA, Heloisa V . Visual Structures for Image Browsing. In: 12th ACM International Conference on Information and knowledge Management, 2003, Nova Orleans. Proceedings 12th CIKM conference. New York : ACM, 2003. p. 49-55.

Relevant theses and dissertations

Ongoing

Vânia Paula de Almeida Neris. Estudo e Proposta de um Framework para o Design de Interfaces de Usuário Ajustáveis. Início: 2006. Tese (Doutorado em Ciência da Computação) - UNICAMP. (Orientador Maria Cecilia C. Baranauskas).

Leonelo Dell Anhol Almeida. Percepção do Espaço de Trabalho em Ambientes Colaborativos Inclusivos. Início: 2007. Tese (Doutorado em Ciência da Computação) - UNICAMP. (Orientador Maria Cecilia C. Baranauskas).

Vagner Figuerêdo de Santana. Identificação de Padrões de Utilização de Websites Mediada por Tecnologias Assistivas. Início: 2007. Dissertação (Mestrado em Ciência da Computação) – UNICAMP. (Orientador Maria Cecilia C. Baranauskas).

Frederico José Fortuna. Aspectos Pragmáticos e Sociais no Design e Desenvolvimento de Ambientes Web Inclusivos. Início: 2008. Dissertação (Mestrado em Ciência da Computação) - UNICAMP. (Orientador Maria Cecilia C. Baranauskas).

Chantal Intrator. Usando *Web scripts* para auxiliar a navegação e interação de usuários com necessidades especiais: Um experimento com o CoScripter. Inicio: 2008. Dissertação (Mestrado em Informática) – PUC-Rio. (Orientadora Clarisse Sieckenius de Souza)

Completed

Manuel Eduardo Loaiza Fernandéz. Implementação de um Dispositivo de Rastreamento Óptico com 6 Graus de Liberdade para Interação com Aplicações de Realidade Virtual. MSc. Dissertation, Depto. de Informática, PUC-Rio, April 2005. Advisors: Marcelo Gattass and Alberto Raposo

Sinisa Kolaric. Towards direct spatial manipulation of virtual 3D objects using vision-based tracking and gesture recognition of unmarked hands. MSc. Dissertation, Depto. de Informática, PUC-Rio, April 2008. Advisors: Marcelo Gattass and Alberto Raposo

3.1.2. Social Networks

The Web has proven to be an impressive platform for interaction. The expressiveness and interconnectivity enabled by the Web allowed individuals to produce massive amounts of data and information. The combined interactions of thousands of individuals effectively create value, as can be seen by success stories such as Wikipedia, Flickr, IMDB and many open source projects. With the Web, individuals have found a new medium to express themselves, interact and collaborate. Tools such as blogs, photologs, wikis, and social networking sites (e.g., Orkut, Facebook, MySpace) have flourished, and are still in high demand. These services allow individuals to create and exchange content. Though the use of these tools, individuals create and manage online identities, connect to others and share their ideas, opinions or work. Through interactions and contributions, they build rich online personas with which others can interact.

Different forms of interaction are possible, and diverse organizational structures emerge as a result of the group's needs. Thus far, developers have pretty much adopted the approach of building software and watching to see what users would do with it. A better understanding of how individual choose which groups to join and whom to associate with is necessary if we are to build appropriate tools for these people. Additionally, an investigation should be conducted into what types of interaction and organization tools are necessary for effective participation and generation of value. Issues of trust, reputation, authorship, verification and participation are of prime importance for the success of future Web applications. Understanding how users build trust online, how they decide which source to rely on and what drives them to participate is important if we are to design new applications for these users.

Social networks are sets of links that may equally and democratically organize people, groups, and institutions around a common goal [Barnes 1987]. In other words, a dynamic model is flexible, with liberty and spontaneity among links, respect to people's individuality and based mostly on mutual trust. Members of a social network may gather and disclose data, information, and knowledge.

To deal with this problem, it is important to acquire a better understanding of Web users. Expected contributions from this subtopic include a set of models along three main axes:

1. Interpersonal: modeling trust and reputation building on the Web; modeling the formation and dissolution of social networks and workgroups;

2. Personal: modeling motivation to contribute content to group or personal projects;

3. Interactional: modeling preferred methods and tools for contributing, interacting and group organization in different situations.

To reach these models, we will select groups and individuals with desired characteristics and conduct ethnographic analyses of their work. This will involve close monitoring of ongoing work, email exchanges via public channels, structures and semi-structured interviews, mapping social networks and analyzing work objects.

It is important to analyze how people interact—with other people, data, applications and other resources—on the Web, as well as to provide mechanisms for analyzing such interactions.

Considering the Web from a social phenomena perspective, Hendler and co-authors have coined the expression "social machine" [Hendler et al. 2008], establishing as an important challenge "How can we extend the current Web infrastructure to provide mechanisms that make the social properties of information-sharing explicit and guarantee that the use of this information conforms to relevant social-policy expectations?". The same challenge was explored by Weitzner and colleagues (2008), from which we highlight the following excerpts:

"Transparency and accountability make bad acts visible to all concerned."

"Data quality is protected by giving all consumers the right to see the data held about them (transparency)."

"The typical consumer appreciates the paradox associated with protecting privacy and other information policy values through increased transparency."

"They could also present accountability reasoning in human-readable ways and allow annotation, editing, and publishing of the data and reasoning being presented [...]. This aspect of the accountability and transparency perspective is closely related to the issue of maintaining provenance for scientific data [...]."

We have investigated the concept of transparency in software production [Cappelli et al. 2007, Leite & Cappelli 2008], using intentional modeling [Mylopoulos et al. 1999] as a way to model social networks [Leite et al. 2007]. In the context of the institute, we will emphasize modeling from the citizen's point-of-view, allowing conflicts among softgoals to be expressed by cross-cutting characteristics [SILVA 2006].

Many discussions enforce the potential of the Web to empower collaboration and participation in society through multidirectional, decentralized, non-hierarchical and volunteered interactions. It is argued that this power is just beginning to show its

face since information access by citizens is gaining several different possibilities and their participation in public matters and problems will continuously arise. What is being called the 'Citizen Era' comprises volunteer and participative action of people connected through an open system with the potential for interchanging information, ideas, opinions and multiple discoveries in different areas of interest.

Another kind of collaboration is the creation of productive and innovation arrangements, which are defined as the set of cooperation agreements among independent organizations pointing to a common technological project. As per DeBresson & Amesse (1991), the concept of network is used on different meanings, from the grouping of individuals on a research project, through technical artifacts to innovation enterprises which work together. Freeman (1991), by a research which was based on DeBresson & Amesse (1991), shows ten subdivisions of productive and innovation arrangements, and one of them was the informal networks. One of the goals of this research is to analyze how social networks which integrate organizations – of research or not – can be organized to generate innovation.

Research in this area analyzes social interaction through the Web and its associated tools, called social software: collaborative applications such as wikis, blogs, social bookmarking and so on. These tools, due to their simplicity and flexibility, have been the fundamental platform for creating and sharing knowledge into communities and social networks through the Web, as well as enabling their actions. However, although social software enables high connectivity among participants, it presents very basic interaction support. Networks with a high articulation level (e.g., open source software communities, social networks for social activism (like WWF), press and news networks) require more advanced coordination and collaboration functionality. In these scenarios, existing social tools are not adequate to guarantee higher levels of collective action, collaboration, productivity and results quality.

The intense use of web tools by social networks has contributed to the augmentation of the content available in the Web, in the form of collective knowledge. This content, however, lacks structure and mechanisms for discovering, capturing and reusing this knowledge for collective and social use. The typical narratives in social networks – stories, posts, messages, etc – bring challenges for the representation/modeling, extraction and interconnection of concepts and information in the web. This would allow the dynamic understanding of the network itself and make this knowledge more widely available for reuse by the network and by the usual citizen, thus increasing information credibility and consistency.

Knowledge generation in a participative, volunteer and in a high-connective manner opens the discussion about the condition of heritage in contemporary days. It becomes a challenge to understand the concept of digital heritage as a new category of organization and keep for the memories created in or transferred to the cyberspace; changing the idea that heritage must always be a product of government and institutional intervention. The digital medium favors the understanding of patrimonial good as an informational object in continuous development, circumstantial, unique and virtual.

Thus, the topic of social networks brings about diverse research opportunities, among which we will investigate the following:

Understanding Social Networks – When we analyze social networks, we verify that they are living organisms, changing, evolving and allowing information and knowledge to circulate among their members. The motivation for this work is to always keep alive the knowledge flow, so that the social networks structure is kept in balance. This way, structural problems in the formation of the social network shall be detected and new relationships (among the members of a social network) will be proposed. The recommendation of new relationships will be based on the characteristics and profiles of each member, as well as on social interaction style and preferences.

Intentional Modeling of Social Networks – Considering the variety of models involved in social networks (e.g., emergent models of reputation, trust, motivation and group formation on the Web), it becomes necessary to build a framework for tool development that integrates all models, as well as to provide mechanisms for reputation, incentive, authoritativeness assessment, and group/partnership recommendation. Moreover, the complexity makes it necessary to also provide means to assist users with media and technology choices. Moreover, intentional modeling may be seen as a way to provide "information accountability" [Weitzner et al. 2008] and software transparency.

Incorporation of Autonomic Characteristics in the Control of Social Networks – Considering that social networks evolve and may become complex, to analyze and manage them may become an impossible task. Possible problems, such as the disconnection among members and the lack or distortion of communication, may affect interaction, leaving the network an inflexible mechanism. Analyzing the Information Technology (IT) scenario, where the number and complexity of systems has enormously grown in the recent decades, we find invaluable lessons to deal with such challenges. Autonomic Computing, which arose due to the increase in complexity of current computational solutions and the difficulty of managing them, is a good example to draw on. Autonomic Computing [Hennessy 2002, IBM 2003, Kephart et al. 2003] is a recent research theme that defines systems that manage themselves according to the goals of the administrator and without direct human intervention. The main idea is that the resources—either hardware or software—obey the self-CHOP rule, that means self-configuration, self-healing, self-optimization and self-protection, being thus able to observe and respond to changes in the environment where they are inserted. The goal of this subtopic is to create mechanisms for a social network to become autonomic, acquiring self-CHOP characteristics, in order to preserve the relationships and members of a social network.

Scientific Social Networks – Scientific collaboration has a great potential for solving complex scientific problems and promote various political, economic and social commitments, such as democracy, sustainable development, besides cultural integration and understanding. This area of study aims to identify—both quantitatively and qualitatively—how professionals in this environment

interact, creating mechanisms for identifying, evaluating, and analyzing social networks built in the scientific context. Moreover, we will study as social networks can be structured to create innovation, involving research organizations, suppliers, universities and the society.

Temporal and Mobile Social Networks – Some relationships and interactions occur in a certain period of time, in a certain place. To keep track, understand, and analyze such interactions, the spatio-temporal information related to such networks, as well as to provide mechanisms for socialization is another goal of this topic.

Goals, Activities, and Deliverables

G1. Ethnographic analyses to understand the social Web

Activities:

Definition of the desired profile of future participants by means of informal pre-interviews with different types of Web users.

Ethnographic analyses to understand how individuals use the World Wide Web, what motivates them and what they expect from it.

Development of models of reputation, trust, motivation and group formation on the Web.

Development of model of preferred technology/media for different scenarios.

Evaluation of the models.

Deliverables:

Interview scripts and results of pilot interviews.

Models that represent how individuals develop trust, reputation, motivation, group formation and technology/media choice. Models will be published on the Web site of the institute.

Journal papers, as well as conference papers, describing the role played by virtual communities in present day societies, the elaborated models, and their evaluation.

Master's dissertations and Doctoral theses based on the investigations, models and their evaluations.

G2. Definition of models for social networks analysis and mining in order to discover patterns, needs and requirements for collaboration support

Activities:

Survey and study of methods for social network analysis

Development of techniques for collaboration patterns discovery and analysis in social networks

Experiments plan and execution

Development of social software adaptive architecture/components for social network dynamic collaboration support

Deliverables:

- Formalization of analysis techniques
- Prototypes for collaboration analysis in social networks
- Method for requirements specification of collaboration patterns support
- Prototypes of adaptive social networks based on collaboration dynamic requirements
- Scientific papers in national and international conferences
- Scientific papers in national and international journals
- MSc dissertations

G3. Definition of mechanisms for knowledge discovery and organization from narrative social content

Activities:

- Study of styles of discourse and narratives from social network interactions
- Development of conceptual models for representing social interaction and its associated elements and contexts
- Study and identification of knowledge discovery techniques (text mining, process mining) for modeling social network phenomena and search for people (social combination) and content (information, actions and knowledge) recommendations for both the network and its dynamic sub-groups
- Development of a model for information quality evaluation through attributes such as reputation and trust
- Experiments plan and execution

Deliverables:

- Conceptual models of discourse and narrative in social networks
- Methodologies, techniques and software prototypes for knowledge discovery in social networks
- Scientific papers in national and international conferences
- Scientific papers in national and international journals
- MSc dissertations

G4. Configuration and conceptualization of digital heritage in social networks

Activities:

- Study of processes for creation and keeping of digital heritage
- Observation of the social dynamics in knowledge appropriation and value aggregation
- Simulation of models for knowledge sharing and digital heritage access

Deliverables:

- Formalization of reading strategies for patrimonial objects

Conceptual domain of knowledge production in social networks

Theoretical construction about virtual social memory and the meanings of collective knowledge accumulation and transformation

Knowledge production and organization models created from the dynamics of information transmission in social networks environments (production – transmission – re-appropriation)

Scientific papers in national and international conferences

Scientific papers in national and international journals

MSc and DSc dissertations

G5: Analyze and balance social networks

Activities

Study and analysis of the research in the area

Extract large volumes of real data for experiments

Develop the approach

Plan an prepare the experiments

Analyze results and validate solution

Deliverables

international journal paper submissions

international conference papers

developed prototypes

M.Sc. dissertations and D.Sc. theses

graduate course and material about the Analysis of Social Networks

G6. Develop tools to support intentional social networks

Activities:

Definition of a framework for tool development that integrates the aforementioned models.

Design of an incentive mechanism based on the contribution motivation model.

Design of a reputation mechanism based on the reputation model.

Design of an authoritativeness assessment mechanism based on the trust model.

Design of a group/partnership recommendation mechanism based on the group formation model.

Design of an intelligent system to assist users with media and technology choices, based on the media choice model.

Instantiation of the framework in the form of a prototype system that integrates the mechanisms designed.

Evaluation of the instantiated prototype through user studies.

Deliverables:

A framework for construction of Web based social network based tool.

An incentive mechanism.

A reputation mechanism.

An authoritativeness assessment mechanism.

A group/partnership recommendation mechanism.

Heuristics for and intelligent system to assist media and technology choices.

Implemented prototype combining mechanisms designed.

Journal and conference papers, describing the framework, the models and corresponding mechanisms, the intelligent agent heuristics, the final implemented prototype and evaluation results.

M.Sc. dissertations and D.Sc. theses describing the framework, the mechanisms the intelligent system their implementation and results.

G7: Incorporate autonomic characteristics into the control of social networks

Activities

Study and analysis of the research in the area

Extract large volumes of real data for experiments

Develop the approach

Plan an prepare the experiments

Analyze results and validate solution

Deliverables

international journal paper submissions

international conference papers

developed prototypes

M.Sc. dissertations and D.Sc. theses

G8: Investigate scientific social networks

Activities

Study and analysis of the research in the area

Extract large volumes of real data for experiments

Develop the approach

Plan an prepare the experiments

Analyze results and validate solution

Deliverables

international journal paper submissions

international conference papers

developed prototypes

M.Sc. dissertations and D.Sc. theses

graduate course and material about Scientific Social Networks

G9: Investigate temporal and mobile social networks

Activities

Study and analysis of the research in the area

Extract large volumes of real data for experiments

Develop the approach

Plan an prepare the experiments

Analyze results and validate solution

Deliverables

international journal paper submissions

international conference papers

developed prototypes

M.Sc. dissertations and D.Sc. theses

graduate course and material about Mobile Social Networks

Timeframe

		Y1		Y2		Y3	Y4	Y5
Goal	Activity	S1	S2	S3	S4			
Understanding the social Web	Definition of the desired profile of future participants by means of informal pre-interviews with different types of Web users.	X	X	X				
	Ethnographic analyses to understand how individuals use the world wide Web, what motivates them and what they expect from it.	X	X					
	Development of models of reputation, trust, motivation and group formation on the Web.	X	X					
	Development of model of preferred technology/media for different scenarios.	X	X					
	Evaluation of the models.		X	X				
Analyze and balance social networks	Study and analysis of the research in the area	X						
	Extract large volumes of real data for experiments	X						
	Develop the approach		X	X				
	Plan an prepare the experiments		X	X				
	Analyze results and validate solution			X	X			
Development of tools to support intentional social networks	Definition of a framework for tool development that that integrates the aforementioned models.		X	X				
	Design of an incentive mechanism based on the contribution motivation model.				X	X		
	Design of a reputation mechanism based on				X	X		

	the reputation model.							
	Design of an authoritativeness assessment mechanism based on the trust model.				X	X		
	Design of a group/partnership recommendation mechanism based on the group formation model.				X	X		
	Design of an intelligent system to assist users with media and technology choices, based on the media choice model.				X	X		
	Instantiation of the framework in the form of a prototype system that integrates the mechanisms designed.					X	X	
	Evaluation of the instantiated prototype through user studies.							X
Incorporate autonomic characteristics into the control of social networks	Study and analysis of the research in the area			X				
	Extract large volumes of real data for experiments			X				
	Develop the approach				X	X		
	Plan an prepare the experiments					X		
	Analyze results and validate solution					X		
Investigate scientific social networks	Study and analysis of the research in the area					X		
	Extract large volumes of real data for experiments					X		
	Develop the approach					X	X	
	Plan an prepare the experiments						X	
	Analyze results and validate solution						X	
Investigate temporal and mobile social networks	Study and analysis of the research in the area						X	
	Extract large volumes of real data for experiments						X	
	Develop the approach						X	X
	Plan an prepare the experiments							X
	Analyze results and validate solution							X
Definition of models for social networks analysis and mining in order to discover patterns, needs and requirements for collaboration support	Survey and study of methods for social network analysis	X						
	Development of techniques for collaboration patterns discovery and analysis in social networks		X	X	X			
	Experiments plan and execution					X	X	X
	Development of social software adaptive architecture/components for social network dynamic collaboration support						X	X
Definition of mechanisms for	Study of styles of discourse and narratives from social network interactions	X						

knowledge discovery and organization from narrative social content	Development of conceptual models for representing social interaction and its associated elements and contexts		X	X				
	Study and identification of knowledge discovery techniques (text mining, process mining) for modeling social network phenomena and search for people (social combination) and content (information, actions and knowledge) recommendations for both the network and its dynamic sub-groups		X	X				
	Development of a model for information quality evaluation through attributes such as reputation and trust				X	X		
	Experiments plan and execution				X	X	X	X
Configuration and conceptualization of digital heritage in social networks	Study of processes for creation and keeping of digital heritage	X	X					
	Observation of the social dynamics in knowledge appropriation and value aggregation			X	X			
	Simulation of models for knowledge sharing and digital heritage access					X	X	X

Team

Name	Affiliation	Degree	CNPq Level
Adriana S. Vivacqua	ADDLabs/UFF	D.Sc., COPPE/UFRJ, 2007	—
Anne Marie Delaunay Maculan	Engenharia de Produção/COPPE	Ph.D., Universite Du Quebec a Montreal, 1989	—
Fernanda Araujo Baião Amorim	UNIRIO	D.Sc., COPPE-UFRJ, 2001	—
Flávia Maria Santoro	UNIRIO	D.Sc., COPPE-UFRJ, 2001	—
Geraldo Bonorino Xexéo	PESC/COPPE	D.Sc., COPPE, 1994	—
Jano Moreira de Souza	PESC/COPPE	Ph.D., University of East Anglia, 1986	1D
Jonice de Oliveira Sampaio	UERJ	D.Sc., COPPE, 2007	—
Julio Cesar Sampaio do Prado Leite	PUC-Rio	PhD, University of California at Irvine, 1988	1C
Renata Mendes de	UNIRIO	D.Sc., COPPE-UFRJ,	—

Araujo		2000	
Vera Lúcia Doyle Louzada de Mattos Dodebei	UNIRIO	D.Sc., ECO-UFRJ, 1997	2

Recent publications

ARAUJO, R. M.; BORGES, M. R. S. The role of collaborative support to promote participation and commitment in software development teams. Software Process Improvement and Practice, v. 12, p. 229-246, 2007.

BAIÃO, F. ; SANTORO, F. M. ; IENDRICK, H ; CAPPELLI, C. ; LOPES, M ; NUNES, V ; DUMONT, A. . Towards a Data Integration Approach based on Business Process Models and Domain Ontologies. In: International Conference on Enterprise Information Systems (ICEIS), 2008, Barcelona. International Conference on Enterprise Information Systems, 2008. v. ISAS-1. p. 338-442.

CAPPELLI *Requirements Engineering Conference, 2007. RE '07. 15th IEEE International* , vol., no., pp.389-390, 15-19 Oct. 2007, C., LEITE, http://ceur-ws.org/Vol-322/paper13.pdfJ.C.S.P., OLIVEIRA, A.P.A. "Exploring Business Process Transparency Concepts,"

DODEBEI, V., GOUVEIA, I. Organisation et préservation de l´information : questions à l´expérience de virtualisation des musées brésiliens. In : 6eme Colloque du Chapitre français de L´ÍSKO. Toulouse, LERASS- Laboratoire d´études et de recherches appliquées en sciences sociales. Université Paul Sabatier Toulouse III, 2007. p. 384-398.

FERNANDES, L. ; LEITE, J. C. S. P. . Generating Requirements Views: A Transformation-Driven Approach. Electronic Communications of the EASST, v. 3, p. 1/7-14, 2006.

GONZÁLEZ-BAIXAULI, B., LEITE, J.C.S.P., MYLOPOULOS, J., Visual Variability Analysis for Goal Models, Proceedings of the 12th IEEE International Conference on Requirements Engineering : IEEE Computer Society, pp. 198-207 (2004)

LEITE, J.C.S.P., CAPPELLI, C., Exploring i* Characteristics that Support Software Transparency. iStar 2008: 51-54 Proceedings of the 3rd International i* Workshop. CEUR Workshop Proceedings 322, 2008

LEITE, J.C.S.P., WERNECK, V.M.B, OLIVEIRA, A.P.A., CAPPELLI, C., CERQUEIRA, A.L.A., CUNHA, H.S., GONZALEZ-BAIXAULI, B. Understanding the Strategic Actor Diagram: an Exercise of Meta Modeling.. Anais do WER07 - Workshop em Engenharia de Requisitos, Toronto, Canada, May 17-18, 2007, pp 2-12 (http://www.inf.puc-rio.br/~wer/WERpapers/artigos/artigos_WER07/Cwer07-leite01.pdf)

LEITE, J.C.S.P., WERNECK, V.M.B, OLIVEIRA, A.P.A., CAPPELLI, C., CERQUEIRA, A.L.A., CUNHA, H.S., GONZALEZ-BAIXAULI, B. Understanding the Strategic Actor Diagram: an Exercise of Meta Modeling.. Anais do WER07 - Workshop em Engenharia de Requisitos, Toronto, Canada, May 17-18, 2007, pp 2-12 (http://www.inf.puc-rio.br/~wer/WERpapers/artigos/artigos_WER07/Cwer07-leite01.pdf)

MONCLAR, R.S. ; OLIVEIRA, J. ; SOUZA, J. M. . A New Approach to Balance Social Networks. In: UK Social Network Conference, 2007, Londres. Proceedings of UK Social Network Conference, 2007. p. 141-142.

MONCLAR, R.S. ; OLIVEIRA, J. ; SOUZA, J. M. . Use of Space and Time Information for Context Identification. In: 11th International Conference on Computer Supported Cooperative Work in Design, 2007, Melbourne. Proceedings of the 2007 11th International Conference on Computer Supported Cooperative Work in Design, 2007. v. II. p. 709.

MONCLAR, R.S. ; TECLA, A. ; OLIVEIRA, J. ; SOUZA, J. M. . MEK: Using Spatial-Temporal Information to Improve Social Networks and Knowledge Dissemination. Journal of Information Sciences, 2008.

SANTORO, F. M. ; BORGES, Marcos R.B. ; REZENDE, E. A. Collaboration and knowledge sharing in network organizations. Expert Systems With Applications, v. 31, p. 715-727, 2006.

SANTORO, F. M. ; BREZILLON, P. . The Role of Shared Context in Group Storytelling. Computing and Informatics, v. 25, p. 1001-1026, 2006.

STROELE et al., Mining and Analyzing Organizational Social Networks Using Minimum Spanning Tree, Cooperative Information Systems (CoopIS 2008)

Relevant theses and dissertations

Completed

Andréa Carreiro Kubitschek Lopes. **Da possibilidade de exercício da memória criativa: Internet, blogs e bloggers.** 2005. Dissertação (Mestrado em Programa de Pós-Graduação em Memória Social) - Universidade Federal do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico. Orientadores: Josaida Gondar e Vera Lucia Doyle Louzada de Mattos Dodebei.

Camila Guimarães Dantas. **O passado em bits: memórias e histórias na internet.** 2008. Dissertação (Mestrado em Programa de Pós-Graduação em Memória Social) - Universidade Federal do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico. Orientador: Vera Lucia Doyle Louzada de Mattos Dodebei.

Inês Gouveia. **A persistência da memória no mundo virtualizado dos museus.** 2008. Dissertação (Mestrado em Programa de Pós-Graduação em Memória Social) - Universidade Federal do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico. Orientador: Vera Lucia Doyle Louzada de Mattos Dodebei.

SILVA, L.F.: Uma Estratégia Orientada a Aspectos para Modelagem de Requisitos, Tese de Doutorado, Março de 2006, Departamento de Informática, PUC-Rio.

CUNHA, H. S. Uso de estratégias orientadas a metas para modelagem de requisitos de segurança, Dissertação de Mestrado, Abril de 2007, PUC-Rio.

Ongoing

CAPPELLI, C. Características Transversais de Transparência, Tese de Doutorado em Andamento, Previsão, Abril 2009, PUC-Rio.

Rafael Monclar Studart. **Identificação, Análise e Balanceamento de Redes Sociais Científicas** (Título Provisório). Início: 2006. Dissertação (Mestrado em Linha de Banco de Dados) - COPPE/Sistemas - Universidade Federal do Rio de Janeiro. Orientadores: Jano Moreira de Souza e Jonice Oliveira.

Ricardo Tadeu da Silva. **Redes Sociais Autonômicas** (Título Provisório). Início: 2008. Tese (Doutorado em Linha de Banco de Dados) - COPPE/Sistemas - Universidade Federal do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico. Orientadores: Jano Moreira de Souza e Jonice Oliveira.

Alessandro Tecla. **Troca de Conhecimento Móvel utilizando Telefones Celulares** (Título Provisório). Troca de Conhecimento Móvel utilizando Telefones Celulares (Título Provisório). Início: 2007. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) - Universidade Federal do Rio de Janeiro. Orientadores: Jano Moreira de Souza e Jonice Oliveira.

Flávio Leal da Silva. **Sobrevivência e tensões: uma análise sobre o patrimônio digital e a memória social entre sociedades culturalmente diferentes.** Início: 2007. Tese (Doutorado em Programa de Pós-Graduação em Memória Social) - Universidade Federal do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico. Orientador: Vera Lucia Doyle Louzada de Mattos Dodebei.

João Carlos Gonçalves. **Identificação automática de Processos baseada em Group Storytelling e Mineração de Textos.** Início: 2008. Dissertação (Mestrado em Informática) - Universidade Federal do Estado do Rio de Janeiro. Orientador: Flávia Maria Santoro.

Leonardo Gatti. **Raciocínio de Conhecimento Contextual em um Modelo de Suporte à Gestão do Conhecimento.** Início: 2007. Dissertação (Mestrado em Informática) - Universidade Federal do Estado do Rio de Janeiro. Orientador: Flávia Maria Santoro.

Lícia Nascimento. **Um Modelo para Gestão de Conhecimento em Comunidades de Software Livre**. Início: 2008. Dissertação (Mestrado em Informática) - Universidade Federal do Estado do Rio de Janeiro, Financiamento: Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro. Orientador: Flávia Maria Santoro.

Mônica Rizzo Soares Pinto. **A questão das publicações científicas eletrônicas e sua preservação pelo depósito legal.** Início: 2008. Dissertação (Mestrado em Programa de Pós-Graduação em Memória Social) - Universidade Federal do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico. Orientador: Vera Lucia Doyle Louzada de Mattos Dodebei.

Nadia Lobo. **Memória virtual do Instituto de Física da UERJ.** Início: 2008. Dissertação (Mestrado em Programa de Pós-Graduação em Memória Social) - Universidade Federal do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico. Orientador: Vera Lucia Doyle Louzada de Mattos Dodebei.

Raphael Crerie da Silva. **Mineração de Processos para descoberta de regras de negócio.** Início: 2007. Dissertação (Mestrado em Informática) - UNIRIO. Orientador: Fernanda Baião. Co-orientação: Flávia Santoro.

Ricardo Túlio Gandelman. **Ambiente de Suporte à Aprendizagem Organizacional baseado em Comunidades Virtuais**. Início: 2008. Dissertação (Mestrado em Informática) - Universidade Federal do Estado do Rio de Janeiro. Orientador: Flávia Maria Santoro.

Taísa Alves dos Santos. **Apoio Colaborativo ao Processo de Desenvolvimento de Software Livre com base na dinâmica de Redes Sociais**. Início: 2008. Dissertação (Mestrado em Informática) - Universidade Federal do Estado do Rio de Janeiro, Financiamento: Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro. Orientador: Renata Mendes de Araujo.

3.1.3. Web and the Individual

Communities, be they virtual or not, are networks of relationships based on one or more sets of affinities.

The importance of these networks cannot be underestimated. In the pre-industrial period, before the emergence of capitalism, for instance, communities were the pillars of social organization. Then, the basis for their formation was mainly the tripod consisting of ties of consanguinity, ties created by life in the same locality or neighborhood and ties generated by shared religious beliefs. Those communities were, therefore, most of the times “natural” in the sense that the criteria for joining them were almost all given by birth (in a specific family, which lived in a specific locality and had specific religious beliefs). As a consequence, they formed closed circles which seldom admitted new members.

In the last two centuries, with the expansion of capitalism and the emergence of individualism, this type of social communal organization was eroded and such erosion was deeply resented by many social and human scientists. The negative consequences of its disappearance, such as widespread selfishness and solitude, were also emphasized.

The arrival of the Internet introduced a radical change in this scenario. From the very first moments of its popularization, the vocation it has for connecting *people* scattered around the world in the same way as it connects international computers became obvious to its users.

In the beginning, such vocation for connecting human beings found its materialization in the new many-to-many synchronous interaction environments – still very simple by present standards – such as the IRC and the Web Chats and in the asynchronous discussion lists. Even though in all of these the available tools and functionalities were still comparatively few and rudimentary, the Web’s potential for the formation of spontaneous communities was soon realized. Soon after, virtual relationship networks became widely popular among the Web’s ever increasing numbers of users.

Nowadays, we can clearly observe that the Web is making possible a revival of communal life. In fact, it is allowing networks of sociability and mutual support to reemerge from the neglected position they occupied during Modernity. The big difference is that the formation of such networks is no longer based almost exclusively on family, neighborhood or religious ties as they were in the pre-industrial era.

It is well known that, in the Web, there are several environments specifically developed to support the formation of networks of relationships (of a variety of types). A few examples are Orkut, MySpace and Second Life, among other similar ones. Recent and still unpublished observation, nevertheless, shows that these environments are frequently used in association with others, developed for different purposes. Among the reasons for this joint use of the resources offered by more than one environment, we strongly suspect that the following play an important role:

(a) Individually, the environments do not fulfill the desires, expectations or needs of their users because they lack specific tools or functionalities which are immediately searched in other environments (Orkut, for instance, is known for its lack of privacy, which makes its users look for other spaces, such as MSN or email programs, for their more intimate interactions);

(b) Also individually, the environments are not popular enough to embrace the users' social networks in their entirety. An illustration makes this clear. When asked why he preferred MSN to Google Talk, a Young interviewee answered emphatically: "My friends are there!". In other words, the network of people with whom this young man interacted regularly was much more important to him than the functionalities or the attractiveness of the alternative environment. In order to conquer him, it would be necessary to conquer his whole social network.

There is, however, another more important reason for the joint use of several environments in the spontaneous formation of virtual communities. Having absorbed what could be called the "network logic", once connected to the Internet, users, mainly the younger ones, seem not to perceive that they are making use of different programs, conceived and developed with different objectives. The impression one has when observing them is that the Web is one single environment, in which several resources are available for different purposes, it doesn't matter where (i.e., in which program or environment). Thus, it was possible for us to find out that community formation may be initiated in environments not conceived for that purpose, as is the case with blogs.

Meanwhile, spaces specifically created for the formation of communities may also be used for different purposes. The space dedicated to communities in Orkut is a good example. There, one can create or join a community with the sole intention to add a taste, a dislike, a preference or a fantasy to one's user profile. Alternatively, one can also create or become a member of an Orkut community in search of a space of solidarity and mutual support. Examples of the first type are communities such as "I love chocolate ice cream" or "I hate going to the beach on Sundays". Examples of the second type are the communities devoted to the discussion of rare diseases by those who have them or by their relatives, to the discussion of difficult life situations such as

divorce and children's custody, and even to the elaboration of mourning among people who are facing the same experience.

From the point of view of its social relevance, one should not immediately conclude that only the instances of the second type are important. To construct his/her social identity, it may be very important to a youngster that others get to know his likes and dislikes in a joyful, informal manner, without having to say a word about them, thus avoiding unwanted exposure. The examples of the second type obviously need no further explanations; they have really become spaces for the exchange of experiences, for mutual support, for manifestations of solidarity and the like. This, however, does not mean to say that they are self-sufficient and that their members do not resort to other complementary environments available on the Web (one good example is the use of environments that allow synchronous interactions which complement the asynchronous ones that characterize Orkut).

In any case, it is clear that certain personal characteristics, interests or life circumstances are among the elements which trigger the spontaneous formation of virtual communities, usually open to new members. And, judging from what we have been able to see in investigations not conducted with this purpose, this seems to happen having any Web program or environment as a starting point.

On the basis of these observations and considerations, a few questions can be formulated to make the initial objectives of the present project more concrete. These objectives are those of investigating the following points by means of open-ended online interviews with Web users:

(1) what do they know about virtual communities,

(2) in their view, how are they created (using what programs/environments as points of departure, what procedures are used to make them public and to recruit their first members),

(3) in their view, what are the motivations behind the creation of communities,

(4) with which virtual communities have they had contact,

(5) to which virtual communities (if any) do they belong, how did they learn about them,

(6) which types of relationships do they harbor,

(7) in their view, what is the efficacy of such communities in respect to:

   a. the feeling of solitude,

   b. mutual support to face suffering,

   c. solidariety,

   d. making public ways of life, beliefs or points of view considered to be alternative,

(8) if these communities have a similar version in "real life",

   a. if so, which are they,

   b. if not, why not,

(9) what role do these communities play in the lives of their members,

(10) when one is a member of a virtual community, what happens if, for some reason, one cannot have access to it,

(11) how do members of virtual communities deal with their constant reconfigurations.

It has already become a standard procedure for studies which deal with new behaviors, new psychological configurations and new social arrangements which result from new technologies that emerge constantly to have initial objectives that lead to results upon which new objectives are built. This is exactly the case of the present proposal, whose further goals will have to be set once the results of this first investigation are known.

Goals, Activities, and Deliverables

G1. Mapping the field.

Activities:

Survey of recent bibliography on virtual communities and social networks in the Human and Social Sciences.

Definition of the desired profile of future participants by means of informal pre-interviews with different types of Web users.

Elaboration of a script of open-ended questions on which the interviews will be based.

Results:

Recruitment of participants.

Pilot interviews to test the script.

G2. Data collection.

Activities:

Training of interviewers.

Carrying out of online interviews.

Qualitative analysis of collected discourse.

Evaluation of psychological and social benefits of online communities.

Results:

Journal papers, as well as conference papers, describing the role played by virtual communities in present day societies.

M.A. dissertations and Doctoral theses that carry out further investigation on the subject.

Time frame

		Y1		Y2		Y3	Y4	Y5
Goal	Activity	S1	S2	S3	S4			
Mapping the field	Survey of recent bibliography on virtual communities and social networks in the Human and Social Sciences.	X	X					
	Elaboration of a script of open-ended questions on which the interviews will be based.	X	X	X				
Data collection	Training of interviewers			X	X			
	Execution of online interviews.					X		
	Qualitative analysis of collected discourse.						X	X
	Evaluation of psychological and social benefits of online communities.						X	X

Researcher

Name	**Affiliation**	**Degree**	**CNPq level**
Ana Maria Nicolaci da Costa	PUC-Rio	PhD, U. of London, 1983	1C

Recent publications

DI LUCCIO, Flavia.: NICOLACI-DA-COSTA, A.M., Hipertexto, blogs e leitores-escritores. Em Miguel Rettenmaier e Tânia K. Rösing (Orgs.), Questões de leitura no hipertexto. Passo Fundo, UFP Editora, pp. 92-110, 2007.

NICOLACI-DA-COSTA, Ana Maria. O campo da pesquisa qualitativa e o Método de Explicitação do Discurso Subjacente (MEDS). Psicologia: Reflexão e Crítica, v. 20, n. 1, pp. 65-73, 2007.

NICOLACI-DA-COSTA, Ana Maria (org) Cabeças digitais: o cotidiano na Era da Informação. Rio de Janeiro, Editora PUC-Rio / Edições Loyola, 2006.

NICOLACI-DA-COSTA, Ana Maria. Revoluções tecnológicas e transformações subjetivas. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-37722002000200009&lng=en&nrm=iso&tlng=pt NICOLACI-DA-COSTA, Ana Maria Na malha da Rede: Os impactos íntimos da Internet. Rio de Janeiro, Editora Campus, 1998.

Relevant theses and dissertations

Completed

Daniela Romão Barbuto Dias, **Brincando de ser na realidade virtual: uma visão positiva da subjetividade contemporânea.** 2007 Tese (doutorado) Pontifícia Universidade católica do Rio de janeiro, Departamento de Psicologia.

Flavia di Luccio, **As múltiplas faces dos blogs: Um estudo sobre as relações entre escritores, leitores e textos.** 2005. Dissertação (Psicologia) – Pontifícia Universidade Católica do Rio de Janeiro.

Erika Falcão Ramalho, **Par Perfeito: um novo espaço virtual para a procura de parceiros amorosos.** 2005. Dissertação (Psicologia) – Pontifícia Universidade Católica do Rio de Janeiro.

Daniela Romão-Dias, **Nossa plural realidade: um estudo sobre a subjetividade na era da Internet.** 2001 Dissertação (Psicologia) Pontifícia Universidade Católica do Rio de Janeiro.

Ongoing

Betty Carkushansky Wainstock: **Filhos que vão, pais que ficam: a web como recurso de "sobrevivência" durante o luto.** 2010? Tese (doutorado) - Pontifícia universidade católica do Rio de janeiro, Departamento de Psicologia

Mariana Matos. **Teclando com os mortos: um estudo sobre mensagens enviadas pelo Orkut a perfis de pessoas falecidas.** 2009?. Tese (doutorado) – Pontifícia universidade católica do Rio de janeiro, Departamento de Psicologia

3.1.4. Web and Politics

An important aspect of Web Science is the research on how the different social agents use the Web to achieve their goals. Within this subtopic we intend to understand the political action of said agents, both within political parties as well as within social movements. It is believed that the Internet and in particular the Web are increasingly important in political mobilizations.

In Brazil, the Web is an increasingly important tool in electoral campaigns, but in general its use is still primitive (Fernandez 2005). The goal of this research line is to follow the use of the Internet in the municipal elections of 2008 and 2012, and in the national election of 2010. With the increasing penetration of the Internet in Brazilian households, it seems reasonable to forecast that it will become a very important tool in the Brazilian elections. It is interesting to compare the importance of the Internet and the Web in the 2008 US elections (Pew Internet 2008) with its relative lack of importance in the 2008 Brazilian elections (so far). The goal of this line of research is to map the evolution of the use of the Web in the Brazilian elections.

Since the Zapatista movement in Chiapas, Mexico (1994) and the anti-globalizations movements in Seattle (1999), it is believed that social movements can and should use the Internet to organize and attract their militants. In Brazil, it is said that the MST uses the Internet (but not the Web) to organize its militants into coordinated action throughout the country. The NGO and social movements that joint to create the Social Forum use the Web to promote their views and to organize them across different organizations. This line of research will measure the quantity and quality of the use of the Web as a form of promotion and organization of different social movements in Brazil.

This subtopic is central to the understanding of social aspects of Web Science. Organization of political movements, both within political parties and in social movements, is an important though partial aspect of the social life of a country.

In the light of the aforementioned issues, the research in this subtopic comprises 4 basic paths. The first one aims at the specification and the design of social tools which will be adaptable to the social network dynamics, leveraging the need of formalism in collaboration and interaction among participants in the network. The second path addresses knowledge discovery and organization from the web content accessible through social software, taking into account its typical characteristics of being free, unstructured and non-intentional. The third path searches for understanding the condition of the collective knowledge generated through social networks as a patrimonial good, and how this heritage can be identified, visible and accessed by all. Finally, the fourth analyzes the Web in Brazilian elections and by social movements.

Goals, Activities, and Deliverables

G1. Analysis of the use of the Web in Brazilian elections

Activities:

- Analysis of the Brazilian literature on uses of the Internet in elections.
- Development of metrics and taxonomies to classify the Web pages from candidates and political parties.
- Evaluation of the 2008, 2010 and 2012 elections in Brazil according to the metrics above
- Elaboration of scientific papers

Deliverables:

- national journal paper
- international conference paper

G2. Analysis of the Web use by social movements

Activities:

- Analysis of the Brazilian and international literature of the uses of the Web as an organizational tool for social movements.
- Develop metrics and taxonomies to analyze the quality and quantity of the use of the Web by these movements
- Selection of some movements for an in-depth analysis (such as Forum Social and MST)
- Evaluation of the uses of the Web by these movements.
- Elaboration of scientific papers

Deliverables:

- national journal paper

Development of one post-graduate course and materials on schema alignment

Timeframe

		Y1		Y2		Y3	Y4	Y5
Goal	Activity	S1	S2	S3	S4			
Web use in elections	Review of the literature	X						
	Develop metrics and taxonomies		X					
	Analysis of elections		X			X		X
	Elaboration of scientific papers		X	X	X	X	X	X
Web use by social movements	Review of the literature	X						
	Development of metrics and taxonomies		X					
	Selection of which social movement to analyze		X					
	Analysis of the movements			X	X	X	X	
	Elaboration of scientific papers		X	X	X	X	X	X

Team

Name	**Affiliation**	**Degree**	**CNPq level**
Jacques Wainer	IC-Unicamp	PhD, Pennsylvania State U. - 1991	1D
Thomas Dwyer	IFCH-Unicamp	Dr., École des Hautes Études en Sciences Sociales, 1978	1C

3.1.5. Web and the Economy

The Web has revolutionized economic life in many ways. A complex of industries, providing infrastructure, information technology and content was created. Corporate life changed dramatically, as workers learned to take advantage of the new technology to gather, communicate and use information.

In many areas, such as banking, stock trading, commerce of consumer products, and business to business procurement, Web-based transactions became common place, and have drastically altered the organization and efficiency of these markets.

Web-mediated markets provide a valuable opportunity for economics research, both as subject of study and as a resource for unique data. Unlike other types of transactions, Web-based transactions leave a trace; in liquid markets, large amounts of individual transaction data accumulate in the databases of the Deep Web. These are data sources that provide the possibility of an in-depth analysis of economic behavior and the operation of market institutions.

Decision support in complex systems requires the conceptual modeling of such systems and the implementation of inference mechanisms, such as simulations, that work upon the conceptual model. These inference mechanisms use data from several different sources, requiring these data to be filtered (according to its source's reliability and relevance), structured, and finally submitted to the simulation machine. The Semantic Web and the growing use of ontologies tend to make available several repositories of semantically described data. Such semantic descriptions make it easier to filter and structure the information. Moreover, semantic web-services describe actions related to the same ontologies used to describe the data they manipulate, allowing decision support systems to be built upon their descriptions and, thus, leading to a better usage of the enormous information repositories available in the Web.

Some domains are already using such data interchange and processing strategies. For instance, developers in the capital markets domain have established XML based languages to allow the interchange of market data. Examples of such languages include MDDL (Market Data Definition Language) and FpML (Financial Products Markup Language). Company balance sheets, equity prices, trading operations, financial related news, among other information relevant to capital markets tend to be described and interchanged in forms dictated by domain aware languages. Once such information becomes standardized and supported by ontologies, inference mechanisms can be enacted to support decision on trading, hedging, and investing.

When an agent plays a role in a complex system, it may be required to make decisions under uncertainty. Usually, the system cannot be fully observed; thus, decisions are made with partial knowledge. On the other hand, the effects of the selected actions are not totally predictable, contributing to the already high uncertainty on how the system shall behave. Given this scenario, agents must rely on models of the world to properly evaluate the risks they are incurring by taking decisions, and so be able to select the actions that will improve their position toward success.

We propose an approach to develop decision support systems by combining Artificial Intelligence techniques with mathematical simulation techniques, like Monte Carlo simulation and Markov-chain based simulation. Automatic plan generation algorithms, like planning algorithms based on Markov decision processes, will be used to generate policies that maximize the benefits an agent may obtain given an acceptable risk exposure. Machine learning techniques will be investigated to support the estimation of the probability distributions that describe the values assigned to parameters of the simulation processes.

Decision support systems based on data collected from the Semantic Web can be useful to several domains, such as capital markets, complex engineering structures, logistics, software development, and so on. We intend to develop a methodology and implement a framework to facilitate the implementation of such decision support systems in different domains. We also intend to create an instance of the framework to support investment decisions in capital markets, according to the analysis techniques most used in this domain: equity technical analysis, equity fundamental analysis, and arbitrage-based analysis.

Equity technical analysis is a financial markets technique that consists in analyzing charts showing short-term historical prices for a given equity, along with information related to the trading volume of this asset along time. The technique aims to identify hilltops (highest price observed in two or more price candlestick bars before a reverse), valleys (lowest price observed in two or more price candlestick bars before a reverse), supports (price level in which buying investors are strong enough to stop and possibly revert a falling trend in the equity price, creating a reverse), and resistances (price level in which selling investors are strong enough to stop and revert a rising trend in the equity price, creating a reverse). Based on such points, traders make decisions on buying or selling positions on equities.

Equity fundamentalist analysis works with information derived from the balance sheets of companies under interest and the forces of supply and demand that regulate the market where these companies are inserted. While technical analysis use trading related information, fundamentalist analysis studies the position of a company within its market and projects its future earnings and losses according to different micro and macro economic scenarios. Finally, arbitrage analysis techniques derive their decisions from mathematical relations between a particular equity and other financial assets derived from it, known as derivatives. By searching for differences between the price observed in the market and the fair price estimated by the mathematical model that links equities and their derivatives, arbitrage analysis techniques support buy-or-sell decisions for such financial assets.

We believe that our proposed decision support framework can take advantage of the availability of semantically described information in the Web and also can incorporate domain specificities. The instantiation of the framework for the capital markets domain will support the validation of such an assumption.

One such market is eBay, the leading auction Web site in the United States. eBay congregates more than 100 million users and, while it is occasionally mentioned in the popular press as a venue for the sale of unusual items, it is an important marketplace for many products, such as collectibles and electronic devices. It has been the subject of several books in the popular press as well as scholarly papers in the economics literature.

Behind the Web site lies a vast database of auctions, both current and past, with bid, seller and buyer histories. This is a rich environment to investigate economic behavior in detail.

Econometric tools that leverage on auction theory have been developed to explore auction data in order to investigate the properties of this market, as well as identify underlying consumer preferences over product characteristics. The method proposed in “Econometrics of Auctions by Least Squares” is particularly suitable to databases from the Web, as it requires relatively large amounts of data in exchange for simplicity and robustness. This paper contains an application using a small sample of 2000 eBay auctions, collected over a month. A second paper, “Measuring Technological Progress in Plam PDAs using Auction Data”, applies the method again to eBay data, and highlights the possibility of using this information to identify underlying consumer preferences over product characteristics. Such features, known in the economics

literature as hedonic price regressions, are an important piece of information in the development of more precise price and cost-of-living indices, as well the impact of technological progress on consumer welfare. This second paper investigates consumer valuations of several generations of palm pilot devices, and find that the introduction of newer products have had a small impact on consumer welfare. In both papers, the datasets used were tiny fractions of the vast amount of information public available from eBay, collected with simple ad hoc data retrieval code.

This topic therefore proposes to develop tools to gather Internet auction data in a more flexible, robust and systematic way, and to organize the data in a way that makes structural econometric analysis practical. Wider data collection efforts would allow for a characterization of consumer preferences over a wider array of products as well as more precise estimates from larger samples. Also, it may in the future permit to investigate patterns of behavior of users across auctions.

The theory of economic behavior in interrelated auctions is in its infancy, and the eBay data is an excellent source of empirical evidence to test and guide the development of this theory.

Goals, Activities, and Deliverables

G1. Development of Internet auction data collecting software.

Activities:

Specification of the desired characteristics of the software.

Development of auction data collecting software.

Testing and debugging.

Deliverables:

auction data collecting software, published in the Web site of the institute.

G2. Development and application of econometric techniques to investigate Internet auction data.

Activities:

Development of structural econometric techniques that recognize the specifics of auction data.

Application of the techniques to investigate online auctions.

Deliverables:

Journal papers and conferences presentations describing the results of the research.

M.A. dissertations and Ph.D. theses on the subject of Internet auctions.

G3. Development of a decision support framework based on data from the Semantic Web

Activities:

Studies on the specification, search, retrieval, and filtering of ontology described data and actions to support decisions;

Bibliographical research on decision support for complex systems;

Research on Artificial Intelligence techniques and simulation-based decision support;

Detailed description of the decision support methodology;

Implementation of the decision support framework;

Production of scientific papers.

Deliverables:

Master's Degree Theses;

Papers submitted to national and international journals or conferences;

Semantic Web based decision support framework.

G4. Framework instantiation for the capital markets domain

Activities:

Bibliographical research on decision support systems for capital markets;

Bibliographical research on ontologies describing capital markets;

Framework instantiation for a specific problem in the capital markets domain;

Framework enhancement based on the former instantiation experience;

Production of scientific papers.

Deliverables:

Master's Degree Theses;

Papers submitted to national and international journals or conferences;

Semantic Web decision support system for a specific problem within the capital markets domain;

Enhancement of the Semantic Web based decision support framework.

Timeframe

		Y1		Y2		Y3	Y4	Y5
Goal	Activity	S1	S2	S3	S4			
Development of Internet auction data collecting software	Specification of the desired characteristics of the software	X						
	Development of auction data collecting software		X	X				
	Testing and debugging			X	X			
Dev. and application of econometric techniques to investigate	Development of structural econometric techniques that recognize the specifics of auction data					X	X	
	Application of the techniques to investigate online auctions							X

Internet auction data								
Development of a decision support framework based on data from the Semantic Web	Studies on the specification, search, retrieval, and filtering of ontology described data and actions to support decisions	X						
	Bibliographical research on decision support for complex systems	X	X					
	Research on Artificial Intelligence techniques and simulation-based decision support		X	X				
	Detailed description of the decision support methodology			X	X			
	Implementation of the decision support framework			X	X			
	Production of scientific papers			X	X	X	X	X
Framework instantiation for the capital markets domain	Bibliographical research on decision support systems for capital markets		X	X				
	Bibliographical research on ontologies describing capital markets			X	X			
	Framework instantiation for a specific problem in the capital markets domain				X	X	X	
	Framework enhancement based on the former instantiation experience					X	X	X
	Production of scientific papers				X	X	X	X

Team

Name	**Affiliation**	**Degree**	**CNPq level**
Leonardo Bandeira Rezende	PUC-Rio	PhD, Stanford University, 2003	—
Márcio de Oliveira Barros	UNIRIO	D.Sc., COPPE-UFRJ, 2001	2

Recent publications

BARROS, M.O.; WERNER, C.M.L.; TRAVASSOS, G.H. Supporting Risk Analysis on Software Projects. Journal of Systems and Software, 2004.

COSTA, H. R.; BARROS, M. O. ; TRAVASSOS, G.H. Evaluating Software Project Portfolio Risks. Journal of Systems and Software, v. 80, p. 16-31, 2007

REZENDE, L. (2003a): "Essays on the Theory and Estimation of Auction Models," Ph.D. thesis, Stanford University.

REZENDE, L. (2003b): "Measuring Technological Progress in Palm PDAs using Auction Data," in Rezende (2003a), chap. 6.

REZENDE, L. (2008): "Econometrics of Auctions by Least Squares," Journal of Applied Econometrics, forthcoming.

Relevant theses and dissertations

Ongoing

Augusto César Espíndola Baffa. (Título Provisório) **Geração Automática de Políticas de Investimento com base em Análise Técnica e Planejamento com Incerteza**. Início: 2008. Dissertação (Mestrado em Informática) - Universidade Federal do Estado do Rio de Janeiro. Orientador: Angelo Ernani Maia Ciarlini.

3.2. Software Technologies for Web Applications

3.2.1. A Multi-Agent Systems Approach for Developing Autonomic Web Applications

The vision of autonomic computing [Tesauro 2004] denotes computing systems that manage themselves to a far greater extent than they do today. To achieve this vision, we believe that interacting sets of individual computing elements must regulate and adapt their own behavior in response to widely changing conditions, with only high-level directions from humans. While traditional approaches to computer systems management are often centralized and hierarchical, today's large-scale computing systems are highly distributed with increasingly complex connectivity and interactions, rendering centralized management schemes infeasible. We propose instead a multi-agent systems (MAS) approach, which is much better suited for autonomic computing. Jennings (2001) advocates an agent-based approach to software engineering based on decomposing problems in terms of decentralized, autonomous agents that can engage in flexible, high-level interactions. This approach is particularly well-suited for autonomic computing systems, which must self-configure, self-protect, self-heal, and self-optimize on both local and system levels. Many ideas developed in the MAS community, such as those pertaining to automatic group formation, emergent behavior, multi-agent adaptation, and agent coordination, among others, could likely be fruitfully adapted for autonomic computing. The practical challenges of automatic computing may likewise spur basic research advances within the MAS community.

Current Web applications have several tasks that need human interaction to be executed. The addition of new functionalities that provide self-* properties to these Web applications can bring a lot of advantages, providing the automation (or semi-automation) of tasks that need human intervention and the incorporation of intelligent services. There are plenty of examples that illustrate this scenario, such as: (i) automatic discovering of new data sources when the current used data sources fails; and (ii) automatic discovering and incorporation of new services that can be released in dynamic environments.

Goals, Activities, and Deliverables

G1. Design techniques to improve the development of autonomic Web applications

Activities:

Elaborate a survey of the autonomic computing approaches

Investigate the adoption of bio-inspired autonomic approaches in the context of Web applications

Define a domain-specific language for the autonomic Web domain

Develop case studies to evaluate the proposed techniques

Identify patterns to be reused in the design and implementation of Autonomic Web application

Create a patterns library

Elaborate scientific papers

Deliverables:

Dissertations and Theses

Papers in international and national conferences

Survey

Techniques to improve the development of autonomic Web applications

G2. Define a framework to support the development of autonomic Web applications

Activities:

Analyze commonalities in the developed case studies to define a kernel of a framework

Design and implementation of the framework

Instantiate the proposed framework for different case studies

Elaborate scientific papers

Deliverables:

M.Sc. dissertations

D.Sc. Theses

Papers in international and national conferences

A framework for Autonomic Web Applications

Timeframe

		Y1		Y2		Y3	Y4	Y5
Goal	Activity	S1	S2	S3	S4			
Design techniques to improve the development of autonomic Web applications	A survey of the autonomic computing approaches	X						
	Investigate the adoption of bio-inspired autonomic approaches in the context of Web applications	X	X					
	Define a domain-specific languages for the autonomic Web domain		X	X				
	Development of case studies to evaluate the			X	X			

	proposed techniques							
	Identify patterns to be reused in the design and implementation of Autonomic Web application				X	X		
	Create a patterns library						X	X
	Elaboration of scientific papers		X	X	X	X	X	X
Define a framework to support the development of autonomic Web applications	Analyze commonalities in the developed case studies to define a kernel of a framework				X			
	Design and implementation of the framework				X	X		
	Instantiate the proposed framework for different case studies						X	X
	Elaboration of scientific papers					X	X	X

Team

Name	**Affiliation**	**Degree**	**CNPq level**
Carlos José P. Lucena	PUC-Rio	PhD, UCLA, 1974	1A
Flávia C. Delicato	UFRN	D.Sc. UFRJ,2005	
Paulo de Figueiredo Pires	UFRN	D.Sc. UFRJ, 2002	
Renato Cerqueira	PUC-Rio	D.Sc., PUC-Rio, 2000	
Thais Vasconcelos Batista	UFRN	D.Sc., PUC-Rio, 2000	
Uirá Kulesza	UFRN	D.Sc., PUC-Rio, 2007	

Recent publications

GATTI, M. ; LUCENA, C. J. P. . A BIO-INSPIRED REPRESENTATION MODEL FOR ENGINEERING SELF-ORGANIZING. In: Brazilian Symposium on Software Engineering., 2008, Campinas (to appear).

NUNES, I. O. ; KULESZA, U. ; NUNES, C. P. B. , CIRILO, E. J. R., LUCENA, C. J. P. . Extending Web-Based Applications to Incorporate Autonomous Behaviour. In: WebMedia 2008, Vila Velha (to appear).

DELICATO, F. C. ; REZENDE, J. F. ; PIRMEZ, L. ; PIRES, P. . Exploiting Web Technologies to Build Autonomic Wireless Sensor Networks. In: 8th IFIP/IEEE International conference on Mobile and Wireless Communication Networks (MWCN 2006), 2006, Santiago. Mobile and Wireless Communication Networks WCC 2006. New York : Springer, 2006. v. 1. p. 99-114.

Relevant theses and dissertations

Completed

COSTA, A. D. Um Sistema Híbrido de Diagnóstico e Recomendação para Sistemas Multi-Agentes. Master Dissertation, PUC-Rio (2008).

Ongoing

GATTI M. A Bio-Inspired Method and Representation Model for Engineering Self-Organizing Multi-Agent Systems PhD Thesis, PUC-Rio.

NUNES, I. O. Incorporating Agency Features in Software Product Lines for the Web Context. Master Dissertation, PUC-Rio.

3.2.2. Model-driven Design and Implementation of Web Applications

Within the framework described in 3.1.1, some issues are worthy of investigation in Web Science, such as the design and implementation of Web applications to support men-machine teams. Web applications today are very complex, and involve a understanding, specifying, designing and implementing a variety of concerns, such as specific domain models, navigation design to support a variety of tasks, interface models to support a diverse and heterogeneous user population accessing the Web through a multitude of devices, subject to a multitude to (often conflicting) security and privacy issues, often within the scope of social networks (be they intra-company or external). Furthermore, the traditional development cycle does not apply anymore (see [Hendler et al. 2008]); applications are nowadays in a permanent state of "in development", with updated and deployments occurring while the application itself is executing within the Web.

An interesting approach to the design of such applications is the model-driven approach, where the models are described using the Semantic Web formalisms. Each model addresses a specific aspect (or concern) of the application, by employing a suitable abstraction and providing a clear and concise notation to represent the model. Examples are Domain Models, Navigation Models, Interface Models, Personalization Models, Functionality Models, Security Models, etc. An associated concern is enabling reuse of existing design experience.

In general, model-driven software development (MDSD) [Greenfield & Short 2005, Stahl & Voelter 2006] aims at capturing essential aspects of a software system through appropriate models. In MDSD, models are not just auxiliary documentation artifacts; rather, they are source first class artifacts that can be used for different automated tasks such as: code generation, and system verification, test and analysis.

Orthogonally, software product lines (SPLs) [Pohl et al. 2005, Clements & Northrop 2002] have emerged as a new trend of software reuse investigating methods and techniques in order to build and customize families of applications. SPLs refer to engineering techniques for creating similar software systems (families of applications) from a shared set of software assets using a systematic method. SPL is a promising approach to improve the productivity and reducing both cost and time-to-market in the software development process. A feature [Czarnecki & Helsen 2006] in a SPL is a system property that is relevant to some stakeholder and is used to capture commonalities or discriminate variabilities among products in a product line. The main aim of SPL engineering is to analyze the common and variable features of applications from a specific market segment, and to develop a reusable infrastructure that supports the software development.

SPLs and MDSD are not only complementary, but their integration holds the potential for valuable synergies. The abstraction power of MDSD can be used to represent different aspects of a product line in a more flexible and convenient way. On the other hand, SPL provides a well-defined application scope, which puts the development and selection of appropriate modeling languages on a sound basis. Furthermore, MDSD naturally aids in the automation of the building product line artifacts. Therefore, comparing to a traditional SPL development, a Model-driven Software Product Line approach can narrow the gap between the problem space and the solution space of software systems. This feature has the potential of speeding up the development of systems that requires mass customization as it is the case of current Web applications.

On the one hand, Web users are getting more demanding, desiring Web contents customized for them based on their profiles. In addition, they also want these Web applications to be provided in shorter time-to-market, reduced costs and better quality. On the other hand, the Web context is extremely dynamic. This implies that Web applications evolve frequently in the time, so they need to be constructed in such way that they can be rapidly adapted. In this context, a Model-driven Software Product Line can be successfully applied to produce a product line of customizable Web applications.

Goals, Activities, and Deliverables

G1. Define a model-driven framework to support specification, design and implementation of Web applications, seen as part of men-machine teams.

Activities:

- Study of methods for heterogeneous model integration, which still allow a clear path to execution. All models will be based on the Semantic Web formalisms.
- Development of complementary models to existing ones (Domain, Navigation, Interface), such as Transactions, Security, Personalization, Social Networking, etc.
- Evolution of existing models to support richer functionalities, such ephemeral or transient applications, rich Internet clients, end-user development, etc.
- Development of implementation environments supporting the models, allowing direct code generation to running environments.
- Development of industrial scale test applications to allow evaluation of the proposed methods and models.
- Empirical evaluation of the tools and methods through case studies.

Deliverables:

- Formal specifications of the methods and models, available as processable documents from servers at the institute.
- Public domain tools supporting the models and generating final running applications, available for download from the Web site of the institute
- Example applications publicly available for download

Journal and conference papers, describing the methods, models, tools and user studies conducted

M.Sc. dissertations and D.Sc. theses, describing the methods, models, tools and user studies conducted

G2. Develop case studies (software product lines) for the Web context

Activities:

Determinate the domain (e.g. content management) of the case studies

Analyze the domain to specify the product lines

Implement the case studies

Choose an existing Web application (legacy system) that has several releases

Analyze the variability among the different versions of the Web application

Refactor the Web application to characterize a software product line

Elaborate scientific papers

Deliverables:

Software product lines for the Web context

Papers in international and national conferences

G3. Develop methodologies, empirical studies and tools to support the development of software product lines for the Web context

Activities:

Make a comparative study among the different SPL methodologies

Define a model driven methodology for the development of SPL in the Web context

Validate the proposed methodology using the case studies previously constructed

Evaluate the case studies using different implementation techniques according empirical engineering

Analyze the results of the empirical studies to determine directives for the development of SPLs for the Web context

Design model driven tools to automate the process of customization of Web applications

Validate the developed tools using the case studies previously constructed

Elaborate scientific papers

Deliverables:

Master dissertations

D.Sc. Theses

Papers in international and national conferences

TimeFrame

		Y1		Y2		Y3	Y4	Y5
Goal	Activity	S1	S2	S3	S4			
Definition of a model-driven framework to support specification, design and implementation of Web applications	Study of methods for heterogeneous model integration, which still allow a clear path to execution. All models will be based on the Semantic Web formalisms	X	X					
	Development of complementary models to existing ones (Domain, Navigation, Interface), such as Transactions, Security, Personalization, Social Networking, etc.	X	X	X				
	Evolution of existing models to support richer functionalities, such ephemeral or transient applications, rich Internet clients, end-user development, etc.	X	X	X	X	X		
	Development of implementation environments supporting the models, allowing direct code generation to running environments		X	X	X	X		
	Development of industrial scale test applications to allow evaluation of the proposed methods and models			X	X	X		
	Empirical evaluation of the tools and methods through case studies				X	X	X	X
Develop case studies (software product lines) for the Web context	Determinate the domain (e.g. content management) of the case studies	X						
	Analyze the domain to specify the product lines	X						
	Implement the case studies	X	X					
	Choose an existing Web application (legacy systems) that have several releases		X					
	Analyze the variability among the different versions of the Web application		X					
	Refactor the Web application to characterize a software product line			X	X			
	Elaboration of scientific papers		X		X			
Develop methodologies, empirical studies and tools to support the development of software product lines for the Web context	Make a comparative study among the different SPL methodologies			X				
	Define a model driven methodology for the development of SPL in the Web context			X	X			
	Validate the proposed methodology using the case studies previously constructed				X	X		
	Evaluate the case studies using different implementation techniques according empirical engineering			X	X	X	X	
	Analyze the results of the empirical studies to determine directives for the development of SPLs for the Web context				X	X	X	X
	Design model driven tools to automate the				X	X	X	

	process of customization of Web applications							
	Validate the developed tools using the case studies previously constructed					X	X	X
	Elaboration of scientific papers				X	X	X	X

Team

Name	Affiliation	Degree	CNPq level
Daniel Schwabe	PUC-Rio	PhD, UCLA,	1B
Carlos José Pereira de Lucena	PUC-Rio	PhD, UCLA, 1974	1A
Renato Cerqueira	PUC-Rio	D.Sc., PUC-Rio, 2000	
Thais Vasconcelos Batista	UFRN	D.Sc., PUC-Rio, 2000	
Paulo de Figueiredo Pires	UFRN	D.Sc., UFRJ, 2002	
Flávia Coimbra Delicato	UFRN	D.Sc., UFRJ, 2005	
Uirá Kulesza	UFRN	D.Sc., PUC-Rio, 2007	
Roberta de Souza Coelho	UFRN	D.Sc., PUC-Rio, 2008	

Recent publications

ALVES, M P ; PIRES, P. ; DELICATO, F. C. CAMPOS, M.L.. CrossMDA: a model-driven approach for aspect management. Journal of Universal Computer Science (Online), v. 14, p. 1314-1343, 2008.

BATISTA, T.; BASTARRICA, C.; SOARES, S.; FERNANDES, L. A Marriage of MDD and Early Aspects in Software Product Line Development. Early Aspects Workshop at SPLC 2008 (EA@SPLC 2008): Aspect-Oriented Requirements and Architecture for Product Lines. co-located with the 12th International Software Product Line Conference 2008, Limerick - Ireland, September 2008.

CIRILO, E.; KULESZA, U.; COELHO, R.; LUCENA, C.; STAA, A. Integrating Component and Product Lines Technologies. In Proceedings of 10th International Conference on Software Reuse (ICSR 2008), Beijing – China, 2008.

CIRILO, E.; KULESZA, U.; LUCENA, C.. A Product Derivation Tool Based on Model-Driven Techniques and Annotations. Special Issue on "Software Components, Architectures and Reuse". JUCS (Online), v. 14, n. 8, pp. 1344-1367, 2008

COELHO, R. S. ; RASHID, A. ; GARCIA, A. ; FERRARI, F. ; CACHO, N. ; KULESZA, U. ; STAA, A. V. ; LUCENA, C. . Assessing the Impact of Aspects on Exception Flows: An Exploratory Study. In: 22nd European Conference on Object-Oriented Programming, 2008, Paphos, Cyprus. Proceedings of 22nd European Conference on Object-Oriented Programming, 2008. v. LNCS. p. 207-234.

FIGUEIREDO, E. ; CACHO, N. ; SANTANNA, C. ; MONTEIRO, M. ; KULESZA, U. ; GARCIA, A. ; SOARES, S. ; FERRARI, F. ; KHAN, S. ; CASTOR, F. ; DANTAS, F. . Evolving

Software Product Lines with Aspects: An Empirical Study on Design Stability. In: Proceedings of International Conference on Software Engineering (ICSE´2008) (to appear), 2008, Leipzig.

GOMES, MARIA CRISTINA FERREIRA ; PIRES, P. F. ; CAMPOS, LINAIR MARIA ; CAMPOS, MARIA L M . Certificação da Utilização de Padrões de Projeto no Desenvolvimento Orientado a Modelos. In: Simpósio Brasileiro de Engenharia de Software, 2006, Florianópolis. Anais do XX Simpósio Brasileiro de Engenharia de Software (SBES), 2006. p. 65-80.

LEAL, L.N. ; PIRES, P. ; CAMPOS, MARIA L M ; DELICATO, F. C. . Natural MDA: Controlled Natural Language for Action Specifications on Model Driven Development. In: 14th International Conference on COOPERATIVE INFORMATION SYSTEMS (COOPIS 2006), 2006, Montpelier. On the Move to Meaningful Internet Systems 2006: CoopIS, DOA, GADA, and ODBASE. Heidelberg, Alemanha : Springer Verlag, 2006. v. 1. p. 551-568.

LEAL, LUCIANA NUNES ; PIRES, P. F. ; CAMPOS, MARIA L M . Uma Linguagem de Especificação de Ações baseada em Linguagem Natural Controlada. In: Simpósio Brasileiro de Linguagens de Programação, 2006, Itatiaia. Anais do 10º Simpósio Brasileiro de Linguagem de Programação (SBLP), 2006. p. 161-174.

LOBATO, C. ; GARCIA, A. ; KULESZA, U. ; LUCENA, C. ; STAA, A. V. . Evolving and Composing Frameworks with Aspects: The MobiGrid Case. In: Proceedings of 7th IEEE International Conference on Composition-Based Software Systems (ICCBSS´2008), 2008, Madrid. pp. 53-62

NUNES, C.; KULESZA, U..; Sant Anna, C.; NUNES, I..; LUCENA, C. On the Modularity Assessment of Aspect-Oriented Multi-Agent Systems Product Lines: a Quantitative Study. In Simpósio Brasileiro de Componentes, Arquiteturas e Reutilização de Software, 2008. p. 122-135.

NUNES, I. O. ; KULESZA, U. ; NUNES, C. P. B. ; LUCENA, C. J. P. . Documenting and Modeling Multi-Agent System Product Lines. In: International Conference on Software Engineering and Knowledge Engineering (SEKE), 2008, San Francisco. 20th International Conference on Software Engineering and Knowledge Engineering (SEKE 2008), 2008. p. 745-751.

NUNES, I. O. ; NUNES, C. ; KULESZA, U. ; LUCENA, C. Developing and Evolving a Multi-Agent System Product Line: An Exploratory Study. In: 9th International Workshop on Agent-Oriented Software Engineering, 2008, Estoril. 9th International Workshop on Agent-Oriented Software Engineering, 2008. p. 177-188.

PIRES, P. ; DELICATO, F. C. ; FIORIO, M. ; OLIVEIRA, C. E. . ArchiMDAs: Um Arcabouço de Segurança Baseado em Transformações MDA. In: Simposio Brasileiro de Engenharia de Software (SBES), 2008, Campinas. Anais do XXII Simposio Brasileiro de Engenharia de Software. Porto Alegre : SBC, 2008.

Relevant theses and dissertations

Completed

CIRILO, E. J. R. GenArch: Uma Ferramenta Baseada em Modelos para Derivação de Produtos de Software. Master Dissertation, PUC-Rio, 2008.

MEDEIROS, A. L. F. MARISA-MDD: Uma Abordagem para Transformações entre Modelos Orientados a Aspectos: dos Requisitos ao Projeto Detalhado, Master Dissertation, Universidade Federal do Rio Grande do Norte (UFRN), 2008.

ALVES, M.P. CrossMDA: Arcabouço para Integração de Interesses Transversais no Desenvolvimento Orientado a Modelos. 2007. Dissertação (Mestrado em Informática) - Universidade Federal do Rio de Janeiro.

FIORIO, M.; ArchMDAs: Um Arcabouço de Segurança Baseado em Transformações de Modelos em MDA. 2007. Dissertação (Mestrado em Informática) - Universidade Federal do Rio de Janeiro.

LEAL, L.N. Natural MDA: Uma Linguagem de Especificações de ações Baseada em Linguagem Natural Controlada. 2006. 120 f. Dissertação (Mestrado em Informática) - Universidade Federal do Rio de Janeiro, Núcleo de Computação Eletrônica.

GOMES, M.C.F. Certificação da Utilização de Padrões de Projeto no Ambiente da MDA. 2005. 165 f. Dissertação (Mestrado em Informática) - Universidade Federal do Rio de Janeiro,

Ongoing

NUNES, C. P. B. On the Modularity Assessment of Aspect-Oriented Multi-Agent Systems Product Lines. Master Dissertation, PUC-Rio.

NUNES, I. O. Incorporating Agency Features in Software Product Lines for the Web Context. Master Dissertation, PUC-Rio.

MEDEIROS, A. L. F. Combining MDD and AOSD to Support Software Product Lines Development, PhD Thesis, Universidade Federal do Rio Grande do Norte (UFRN).

FILGUEIRA, G.C.A. Applying Aspects and Models to Build a Practical Product Line Approach. MSc. Thesis, Universidade Federal do Rio Grande do Norte (UFRN).

COBE, R. M.O., An Ontology-based MDA Framework for Requeriments Engineering. MSc. Thesis, Universidade Federal do Rio Grande do Norte (UFRN).

3.2.3. Design and Implementation of Autonomic Workflows

The new economy presents several challenges to organizations. The highly dynamic nature of markets demands companies capable of facing constant change, always aiming at a better strategic position and competitive advantage. This forces process optimizations that compose the company's value chains, so that the company survives this unstable scenario, keeping its partners and acquiring new clients. Such a scenario requires new approaches for managing companies' processes [Harmon 2003] and, consequently, more efficient workflow management systems.

Workflow management systems are a kind of tool for automating processes and improve services. Workflow models represent the functionalities, flows, requirements and priorities that constitute the tasks in a business process.

Although, during the last decade, concepts and technologies related to workflow management have been applied in many enterprise systems [van der Aalst et al. 2000, van der Aalst & van Hee 2002, Fischer 2001, Jablonski & Bussler 1996, Leymann & Roller, 1999], there are many problems faced by their users that may become great research opportunities in Computer Science.

One of these problems is that such systems need a workflow designer, i.e., a professional responsible for building a detailed and precise model describing work routines [Weijters & van der Aalst 2002]. Modeling a workflow is not a trivial task, requiring some understanding about the domain, knowledge about some formalism or language for the model specification, besides the execution of tasks that help the process, such as interviews and discussions with workers and managers involved in the process to be modeled in order to capture the business logic. Because it is a manual task, even when domain specialists do the modeling, it becomes almost impossible to know or to remember all the rules involved in the business or foresee all flow alternatives, possible problems or incidents.

Another limiting factor is the constant unpredictability found in the market and in the business scenario, with new players, competitors, policies and rules, as well as unexpected problems, such as losing personnel, lack of resources, exceptions, among others. Such factors alter the execution of any process, making it necessary for the manager or administrator to update the model. Taking into account that a workflow has associated sub-workflows, this update is costly. The dynamic nature of processes is high and their impact in management is unquestionable.

As mentioned in section 3.2.1, Autonomic Computing brings a promising approach to such challenges. Making an analogy between a new economy and new systems, companies need autonomic characteristics to survive the complex and agile economy. We believe that the principles used in Autonomic Computing may be adapted to help organizations to survive in the new business scenario. Thus, organizations could rely on processes capable of self-management and self-adaptation to changes, i.e., autonomic processes.

This research topic is based on the hypothesis that workflows may acquire autonomic behavior, thus improving their performance and ensuring their complete execution. The applicability of this work is not limited to the business sphere. Autonomic workflows may be used in scientific scenarios for managing scientific workflows, such as those executed in computational grid environments.

Another applicability context is that of personal computing. Our activities are continuous, being executed through mobile devices—such as PDAs, cell phones and notebooks—, traversing from company computers to our personal computers at home. We carry out some of our work while waiting in some line or when stuck in traffic. In other words, workflows— be they personal or professional —may be executed anywhere and anytime. Autonomic workflows would help in this scenario, obtaining some information about the person's current context, such as geographical location, time,

current appointments and goals, as well as the level of attention, for them to manage their own execution.

A third scenario is the medical area, where unstable conditions such as new symptoms or patient reactions, as well as other adversities (lack of a certain medicine, professional or equipment defects) may alter the expected activity flow.

In this topic, we will use the concept of Autonomic Element [Brittenham et al. 2007], abstraction of the main component in an autonomic system architecture. The main function of the autonomic element is to manage a resource, be it software or hardware, as well as take actions on it. The managed elements are equivalent to those found in non-autonomic systems, but they can be adapted to allow the autonomic element to monitor and control them. The autonomic element, besides monitoring the resource it manages, needs also to monitor its external environment. This external monitoring is done by communicating with other autonomic elements. Each element maintains its internal behavior and its relationship with other autonomic elements according to policies established by a human agent (resource administrator, such as, for instance, a data base administrator) or by other elements. Using an analogy of an insect colony, a large part of the autonomic network management is carried out through the intelligence obtained from the relationships among the autonomic elements.

In this work, each workflow and sub-workflow will be managed by an autonomic element, in addition to having each element in communication with the others.

Goals, Activities, and Deliverables

G1: Investigate and prepare the scenario

Activities:

- Study and analysis of different kinds and levels of workflows
- Gather large volumes of real data for experiments, verification of patterns and possible anomalies in the models or in the workflow execution
- In-depth survey of the research work in the area, as well as the concepts and techniques related to the following themes: workflow (especially reuse, simulation, flexibility and dynamic remodeling), Web services, and autonomic computing
- Define the solution architecture

Deliverables:

- international journal paper with a survey of the area
- international conference paper

G2: Develop the autonomic element

Activities:

- Identify attributes used in monitoring
- Develop methods for monitoring, analysis, planning, and execution (actions of the autonomic element)
- Plan and prepare experiments

Analyze results and validate solution

Deliverables

implemented prototype

international journal paper

international conference paper

graduate course and material on process-based autonomic computing

G3: Integrate sub-workflows

Activities:

Develop sub-workflows

Update self-CHOP characteristics, taking into account changes in the environment, in the workflows and in the sub-workflows

Plan and prepare experiments

Analyze results and validate solution

Deliverables:

implemented prototype

Master dissertations or D.Sc. theses

international conference papers

G4: Develop autonomic data approach

Activities:

Create an approach for the data used in a process to have self-CHOP characteristics and for it to self-configure according to the activities being executed

Develop the approach

Plan and prepare experiments

Analyze results and validate solution

Deliverables:

implemented prototype

international journal paper

Master dissertations or D.Sc. theses

international conference papers

G5: Specialize solution for Web services

Activities:

Insert autonomic properties in Web services management, considering that many processes are integrated using this kind of service

Develop the approach

Plan and prepare experiments

Analyze results and validate solution

Deliverables:

implemented prototype

international journal paper

Master dissertations

international conference papers

G6: Analyze the applicability of the solution in other scenarios

Activities:

Analyze medical, personal computing, and scientific computing scenarios

Investigate differences and subtleties of each scenario

Develop and specify solution

Deliverables:

specification of a solution for management in autonomic processes in the medical, personal computing, and scientific computing scenarios

international conference paper

TimeFrame

		Y1		Y2		Y3	Y4	Y5
Goal	Activity	S1	S2	S3	S4			
Investigate and prepare the scenario	Study and analyze different kinds and levels of workflows	X						
	Gather large volumes of real data for experiments, verification of patterns and possible anomalies in the models or in the workflow execution	X						
	In-depth survey of the research work in the area, as well as the concepts and techniques related to the following themes: workflow (especially reuse, simulation, flexibility and dynamic remodeling), Web services, and autonomic computing	X	X					
	Define the solution architecture		X					
Develop the autonomic element	Identify attributes used in monitoring	X	X					
	Develop methods for monitoring, analysis, planning, and execution (actions of the autonomic element)		X	X				
	Plan and prepare experiments				X	X		
	Analyze results and validate solution					X		
Integrate sub-workflows	Develop sub-workflows					X		
	Update self-CHOP characteristics, taking into account changes in the environment, in the workflows and in the sub-workflows					X	X	
	Plan and prepare experiments						X	X

	Analyze results and validate solution							X
Develop autonomic data approach	Create an approach for the data used in a process to have self-CHOP characteristics and for it to self-configure according to the activities being executed					X		
	Develop the approach					X	X	
	Plan and prepare experiments						X	X
	Analyze results and validate solution							X
Specialize solution for Web services	Insert autonomic properties in Web services management, considering that many processes are integrated using this kind of service					X		
	Develop the approach					X	X	
	Plan and prepare experiments						X	X
	Analyze results and validate solution							X
Analyze the applicability of the solution in other scenarios	Analyze medical, personal computing, and scientific computing scenarios							X
	Investigate differences and subtleties of each scenario							X
	Develop and specify solution							X

Team

Name	Affiliation	Degree	CNPq Level
Geraldo Bonorino Xexéo	PESC/COPPE	D.Sc., COPPE, 1994	
Jano Moreira de Souza	PESC/COPPE	Ph.D., University of East Anglia, 1986	1D
Jonice de Oliveira Sampaio	UERJ	D.Sc., COPPE, 2007	

Recent publications

MONTEIRO, P., ABPM: Autonomic Business Process Manager, LAACS 2008

TERRES, L. et al. Autonomic Business Process Simulation In A Rock Analysis Laboratory, LAACS 2008

PINHEIRO, W. et al. Using Active Rules and Petri Nets to Composite Event Detection in Autonomic Systems, LAACS 2008

SILVA, M.; OLIVEIRA, J., SOUZA, J.M., Autonomic Data: Use of Meta Information to Evaluate, Classify and Reallocate Data in a Distributed Environment, LAACS 2008

MARZULLO Fabio Perez; SOUZA, J. M. ; BLASCHEK, J. R. . A Domain-Driven Development Approach, using MDA, SOA and Web Services. Journal of Eletronics and Computer Science, v. 10, p. 1-6, 2008.

RODRIGUES NT., J.A. ; MONTEIRO JR. ; SAMPAIO, J. O. ; SOUZA, J. M. ; SILVA, G. Z. . Autonomic Business Processes Scalable Architecture. Business Process Management

Workshops - Post Conference Book. Berlin / Heidelberg: Springer, 2008, v. 4928, p. 78-83.

PINHEIRO, W. A. ; XEXEO, G. B. ; SOUZA, J. M. . Autonomic Patterns: Modelling Data Killing Patterns Using High-Level Petri Nets. In: The Fourth International Conference on Autonomic and Autonomous Systems, 2008, Gosier. The Fourth International Conference on Autonomic and Autonomous Systems, 2008. p. 198-203.

Rodrigues Nt., J.A. ; Monteiro Jr. ; SAMPAIO, J. O. ; SOUZA, J. M. ; SILVA, G. Z. . Autonomic Business Processes Scalable Architecture. In: Workshop on Business Process Intelligence, 2007, Brisbane. Workshop on Business Process Intelligence - BPM 2007 - Pós-Conference Proceedings. Berlin : Springer, 2007. v. 4928. p. 78-83.

Rodrigues Nt., J.A. ; Monteiro Jr. ; SAMPAIO, J. O. ; SOUZA, J. M. ; SILVA, G. Z. . Towards an Autonomic Enterprise: From Autonomic Business Processes to Autonomic Balanced Scorecard. In: Brazilian Workshop on Business Process Management, 2007, Gramado. Brazilian Workshop on Business Process Management. Porto Alegre : SBC, 2007.

SAMPAIO, J. O. ; Monteiro Jr. ; Rodrigues Nt., J.A. ; SOUZA, J. M. ; PERAZOLO, M. . The Autonomic Balanced Scorecard. In: Latin American Autonomic Computing Symposium, 2007, Petrópolis. Latin American Autonomic Computing Symposium - 2007. Petrópolis : LNCC, 2007.

Relevant theses and dissertations

Ongoing

RODRIGUES, NT; Balanced Scorecard Autonômico, D.Sc. M.Sc. - COPPE/Sistemas - Universidade Federal do Rio de Janeiro. A ser defendida em Março/2009

MONTEIRO JR.; Processos Autonômicos, M.Sc. - COPPE/Sistemas - Universidade Federal do Rio de Janeiro. A ser defendida em Março/2009

TERRES, L. Processo Autonômico de Testes de Amostras (Título Provisório). M.Sc. - COPPE/Sistemas - Universidade Federal do Rio de Janeiro. A ser defendida em Março/2009

ZUQUIM, RAFAEL; A adição de semântica A Serviços Web como suporte A Serviços Web Autonômicos, B.Sc. – Engenharia de Sistemas e Computação- Universidade Federal do Rio de Janeiro. A ser defendida em Março/2009A ser defendida em Dezembro/2008

3.2.4. Security and Resilience of Web Applications

The emergence of the Internet as a ubiquitous computing infrastructure has stimulated the creation of a new class of applications. Such applications are usually structured as a collection of distributed components (accessible over the Web), that must deal with inputs from a variety of sources, provide real-time responses, often adapt to user-defined preferences - and do all this while providing a positive user experience (e.g., the system should not abruptly crash while performing any of its tasks). Thus, if on the one hand Web creates a new realm of service, design and implementation possibilities, most of which are based on multiple threads of control and the asynchronous exchange

of data; on the other hand, the Web environment makes systems fundamentally more complex, and consequently brings many threats to application *security* and *resilience*.

The *security* of Web applications has become increasingly important, as more and more Web-based enterprise applications deal with sensitive data (e.g. financial and medical data), which, if compromised, in addition to downtime can lead to financial and personal damages. Although it is crucial to protect these applications from security threats, the current approaches to achieve application security leave much to be desired. A great deal of attention has been paid to network-level attacks (e.g., port scanning), but according to a survey [Hulme, 2001] about 75% of all attacks against Web servers target Web-based applications. Thus, a simple programming mistake can leave a Web application vulnerable to unauthorized data access, unauthorized updates or deletion of data, and application crashes leading to denial-of-service attacks. Many threats can be avoided by adapting adequate coding standards [Howard and LeBlanc, 2001]. Furthermore, many threats to application security can be detected with code reviews [Larus et al, 2004]. Although code reviews are recognized as one of the most effective verification techniques, they are time-consuming, costly, and are therefore performed infrequently.

The current situation calls for better tools that help developers detecting vulnerabilities during the development cycle, such as the use of static analysis tools [Hulme, 2001][Livshits and Lam, 2005]. Static analysis tools have been used to improve software development by systematically detecting programming faults and helping programmers fixing such faults early in the development process. Example of such tools are: PREFix [Howard and LeBlanc, 2001], developed and used by Microsoft, which finds a fixed set of bugs in C and C++ programs; and FindBugs[3], developed by the University of Maryland and used by the open source community, which statically finds bugs on Java programs. Such tools traverse a program's representation looking for pattern of faults. Current static analysis tools find "general" faults (e.g., null pointers and buffer overflows), they do not tackle faults specific to security threats.

However, security is just one aspect of software dependability [Avizienis et al, 2004], [Avizienis, 1997]. This is an issue that still is requiring much research. Citing [PITAC, 1999] one key issue is that: "*... We have become dangerously dependent on large software systems whose behavior is not well understood and which often fail in unpredicted ways*. ..." Other authors state similar observations. Just an example: [Thomas, 2002] criticizes modern class libraries mentioning that they are not dependable, often due to shoddy work. With such libraries, are we able to develop dependable software? Currently most of the proposals aim at defect free software. Certainly this is desirable. However, it is utopist and unfortunately does not suffice. Despite our efforts the software we are capable to deliver today will contain defects. If it is defect free, we will never know for sure. But even perfect software might fail, since the hardware it runs on might fail, other software it depends on (e.g. operating system, network servers, remote services etc.) might fail or users might err whether willfully or not [Brown and Patterson, 2001]. Hence we should aim at developing recoverable software [Fox 2002]. Software failures are more often than not due to defects in seldom

[3] http://findbugs.sourceforge.net/

exercised parts [Humphrey, 2008]. These parts tend to be responsible for run-time error detection and handling. Thus they tend to be insufficiently or not at all tested.

Hence, as important as trying to detect and remove faults in Web-applications, is to write software that can coexist with them. Since there is no available approach that can remove all faults from an application [Patterson et al, 2002], it is not a matter of "*if your software will fail*", but "*what will happen when it fails*". As the Web-applications usually have rigid dependability requirements, we also need to development approaches that enable the application to coexist with the remaining faults, minimizing the consequences when they come. We need, therefore, development approaches to make applications more *resilient*[4] to such faults. The recovery-oriented programming [Patterson et al, 2002] proposed recently takes the perspective that hardware faults, software faults, and operator errors are facts to be coped with, since they cannot be completely solved - only minimized through verification approaches such as static analysis and testing. In distributed environments it is even more important to develop software that is self-checking, easing the effort of understanding the causes of possible failures. Recent studies have shown that self-checking software not only helps diagnosing failures, but contributes to reducing the amount of defects that are accidentally inserted into the software [Magalhães et al, to appear].

Goals, Activities, and Deliverables

G1. Define static analysis techniques supporting development approaches for detecting possible threats to application *security*

Activities:

- Identify the main threats to Web-applications security.
- Analyze existing static analysis tools
- Propose and static analysis tool or an extension of an existing tool to detect faults that affect the application security.
- Design and implement the static analysis technique.
- Perform extensive tests to validate the technique.
- Conduct case studies to evaluate the technique: (i) select a list of existing industrial Web applications, (ii) use the static analysis technique, and (iii) measure the its effectiveness.
- Evaluate the case studies results.
- Propose a combination between the proposed static analysis technique and existing testing approaches for Web-applications.
- Write technical reports and papers describing the technique and the study results.

Deliverables:

- A static analysis technique - that discovers threats to application security.
- Papers in international and national conferences and periodicals

[4] Some aspects concerning *resilience* are fault tolerance, self-checking, self-recovering and self-healing.

G2. Propose a development approach to improve Web-application *resilience* to faults, enabling them to coexist with faults

Activities:

Evaluate the current techniques of recovery-oriented software.

Define guidelines for developing recovery-oriented software in the Web context.

Define metrics to assess the effectiveness of the proposed approach e.g., time spent to implement, test and debug the application, remaining fault density.

Evaluate the methodology though empirical studies.

Compare the proposed approach with existing development approach – in terms of the metrics defined previously.

Evolve the approach according to the feedback received along the empirical studies.

Write technical reports and papers describing the technique and the study results.

Deliverables:

An approach for developing resilient Web applications.

Master dissertations and a D.Sc. Theses

Papers in international and national conferences and periodicals

Timeframe

		Y1		Y2		Y3	Y4	Y5
Goal	Activity	S1	S2	S3	S4			
Define a static analysis techniques and supporting development approaches for detecting possible threats to application *security*	Identify the main threats to Web-applications security.	X						
	Analyze existing static analysis tools.	X						
	Propose and static analysis tool or an extension of an existing tool to detect faults that affect the application security.		X					
	Design and implement the static analysis technique.		X	X				
	Perform extensive tests to validate the technique.			X				
	Conduct case studies to evaluate the technique: (i) select a list of existing industrial Web applications, (ii) use the static analysis technique, and (iii) measure the its effectiveness.			X	X			
	Evaluate the case studies results.			X	X			
	Propose a combination between the proposed static analysis technique and existing testing approaches for Web-applications.			X	X			
	Write technical reports and papers describing the technique and the study results			X	X	X		
Propose a development	Evaluate the current techniques of recovery-oriented computing.			X				

approach to improve Web-application *resilience* to faults, enabling them to coexist with faults	Define a methodology to application development in the Web context.			X	X			
	Define metrics to evaluate the effectiveness of the proposed approach e.g., time spent to implement, test and debug the application.				X			
	Evaluate the methodology though empirical studies.				X	X	X	
	Compare the proposed approach with existing development approach – in terms of the metrics defined previously.				X	X	X	
	Evolve the approach according to the feedback received along the empirical studies.					X	X	
	Write technical reports and papers describing the technique and the study results.					X	X	X

Team

Name	**Affiliation**	**Degree**	**CNPq level**
Arndt von Staa	PUC-Rio	PhD, Waterloo, 1974	
Roberta de Souza Coelho	UFRN	D.Sc., PUC-Rio, 2008	
João Magalhães	PUC-Rio	PhD candidate, PUC-Rio, to conclude in 2009	
Uirá Kulesza	UFRN	D.Sc., PUC-Rio, 2007	

Recent publications

COELHO, R.; RASHID, A.; GARCIA, A.; FERRARI, F.; CACHO, N.; KULESZA, U.; STAA, A.V.; LUCENA., C. Assessing the Impact of Aspects on Exception Flows: An Exploratory Study. In: European Conference on Object Oriented Programming (ECOOP 2008), 2008, pp 207-234.

COELHO, R; RASHID, A; KULESZA, U; STAA, A.; LUCENA, C; Unveiling and Taming the Liabilities of Aspect Libraries Reuse, Brazilian Symposium on Software Engineering, (SBES´08), 2008.

COELHO, R.; CIRILO, E.; KULESZA, U.; STAA, A., RASHID, A.; LUCENA, C.; JAT: A Test Automation Framework for Multi-Agent Systems, International Conference on Software Maintenance (ICSM 2007), 2007.

MAGALHÃES, J. A. P. ; STAA, A. V. ; LUCENA, C. J. P. . Evaluating the Recovery Oriented Approach through the Systematic Development of Real Complex Applications. Software, Practice & Experience, 2009.

STAA, A. V.; OLIVEIRA, J. P. M. . Desenvolvimento tecnológico de qualidade: sistemas disponíveis, corretos, seguros, escaláveis, persistentes e ubíquos. Computação Brasil, Porto Alegre, RS, p. 8 - 8, 01 nov. 2006.

Relevant theses and dissertations

Completed

COELHO, R., Analyzing Exception Flows of Aspect-Oriented Programs. PhD Thesis, PUC-Rio, 2008.

Ongoing

MAGALHÃES, J. A. P.; Evaluating the Recovery Oriented Approach through the Systematic Development of Real Complex Applications, PhD Thesis, PUC-Rio, a ser finalizada em 2009.

3.2.5. Web-based Collaborative Virtual Environments

The term "e-Science" denotes the systematic development of research methods that exploit advanced computational thinking. As a means of making e-science, the Web can be conceived as a Problem Solving Environment (PSE), i.e., a specialized software system that provides all the computational facilities needed to solve a target class of complex scientific problems. These features include advanced solution methods, automatic and semiautomatic selection of solution methods, and ways to easily incorporate novel solution methods. Moreover, PSEs use the language of the target class of problems, so users can run them without specialized knowledge of the underlying computer hardware and software technology [Houstis et al.1997]. Therefore, High Performance Computing, Visualization and Collaboration technologies are being heavily used to improve the Web capabilities as a PSE.

Collaborative Virtual Environments (CVEs) appear as relevant environments, integrating Visualization and Collaboration, defined as simulation of real or imaginary worlds where the focus is to provide a common virtual space to distributed teams, where they can meet, co-exist and cooperate, while they interact with the environment, share information and manipulate artifacts in real time [Goebbels et al. 2003], [Raposo, Cruz et al. 2001]. CVEs may be seen as the result of a convergence of research interests within the Virtual Reality and Computer Supported Cooperative Work (CSCW) communities. CVEs are finally leaving academic and military spheres and becoming increasingly popular, partially due to the Second Life phenomenon [Second Life 2008]. In spite of their increasing popularity, CVEs present a series of challenges in their development, being a challenging research theme.

CVEs also constitute a link between e-science and e-learning, being important tools for training and learning. The area of education is undergoing a process of adaptation since the appearance of the Web. The use of the Web makes cooperative learning easier to implement, allowing for a rich exchange of information between members of a knowledge community. With the new possibilities brought by CVEs and related technologies a new stage on e-learning is starting to be envisioned.

In recent years, the convergence of games and filmmaking has been seen as an opportunity to create Interactive Storytelling systems by means of which authors, audience, and virtual agents engage in a collaborative process to dynamically create and tell narratives. This process can be useful for different purposes, such as writing literary

texts, education and training, modeling and decision making and, of course, entertainment. Different approaches have been proposed, using techniques and concepts from many areas such as Computer Graphics, Artificial Intelligence, Cognitive Science, Literature and Psychology. In the Web, Interactive Storytelling systems can benefit from contributions of many different authors. Nevertheless, due to the distributed nature of the Web, special techniques for sharing and reusing (parts of) narrative contexts are needed.

In Digital Entertainment, one of the most promising applications for Interactive Storytelling corresponds to Virtual Worlds where users can interfere actively or passively in the course of a narrative. To catch users' attention, these applications should be based on conceptual models of the characters and of the narrative genre, which should guarantee the coherence of the stories. Since these models describe both the typical events that might occur and the emergence of goals (of the characters and of the narrative as whole), as well as graphical attributes of the scenario and of the characters, goal-inference and automated planning techniques can be applied to coherently combine events and then dramatize them by means of graphical animations in a virtual world. The sharing of contexts through the Web is an important feature for these systems. Many different users should be able to interact within a single virtual world, either directly (e.g. by means of avatars) or indirectly, by inserting autonomous characters modeled by them to play a role in the narratives.

The use of Interactive Storytelling techniques has also a great potential for the development of Web applications for cultural content production and e-learning. In these applications, narratives have also to be coherently created and presented in order to convey a message. If various (cultural or e-learning) objects are formally modeled, specifying pre-conditions for their understanding and their effects on the audience, it is possible to use planning algorithms to generate specific presentation flows for each user. Different Physics lessons about forces, for instance, could be generated for specific students, depending on their preferences and previous skills. Students who do not know Cinematic could have this content presented previously, students who already know Differential Calculus could have a more detailed explanation, etc. For cultural production, a documentary about musical theater at the beginning of the 20$^{th.}$ century could, for instance, show details about Music, History or Fashion according to the user's profile. If, additionally, there are models for inferring the goals of different types of users, after the presentation of a partial narrative, the composition of the rest of the narrative could be guided by these goals, both in cultural and e-learning contexts.

Whether the interactive stories are graphical animations within a virtual world, interactive lessons produced on demand or documentaries about a certain topic, the advent of the Semantic Web allows the distributed specification of their components according to ontologies and their coherent integration. Since graphical and logical contexts of virtual worlds are specified by means of ontologies, the elements of a virtual world can be incorporated into another virtual world, greatly increasing the number of different stories that can be generated. Characters modeled to play a role in a specific virtual world can be, for instance, "taken" to other virtual worlds in order to generate new narratives. On the other hand, in e-learning and cultural production applications, the reuse of semantically described content is crucial. In e-learning, the development of

learning objects, in particular those that incorporate multimedia objects, is considered expensive and time consuming. In this way, the reuse of parts of objects, already developed by various authors, and their combination are considered a valid alternative to the creation of new objects. This approach is even more interesting due to the huge volume of documents in the Web. People who are responsible for the development of learning objects could build and describe components in a granularity level that is smaller than the usual one of a learning object. Such components could be recombined, in accordance with their descriptions, to create various narratives that would be more contextualized and personalized.

In this topic we will investigate the use of Interactive Storytelling techniques to support Web applications that are based on the concept of narratives. The research will focus on supporting Web applications in two main areas: Virtual Worlds and production of cultural and e-learning content. In both areas, components are semantically described in order to be reused in the composition of new artifacts, using automated planning and content personalization.

Goals, Activities, and Deliverables

G1. Investigate the Web as a means for eScience and eLearning through CVEs

Activities:

- Study CVEs
- Study eScience and eLearning technologies on the Web
- Develop CVEs prototypes
- Plan and conduct experiments to evaluate the prototypes
- Elaborate models to further the development of the Web for enabling eScience and eLearning

Deliverables:

- CVEs publicly available
- Specifications of the design and development models
- Example applications publicly available for download
- Journal and conference papers, describing the prototypes and conducted experiments
- M.Sc. dissertations and D.Sc. theses

G2. Interactive storytelling for Web-based collaborative virtual worlds

Activities:

- Bibliographical research on approaches for modeling virtual worlds and interactive stories;
- Studies and definition of models for describing, by means of Semantic Web ontologies, the logical and graphical elements of interactive stories in virtual worlds, including scenarios, characters, goals and events;

Implementation of a tool for specifying the context of interactive stories in accordance with the proposed model;

Implementation of an Interactive Storytelling tool for composing interactive stories in accordance with the proposed model;

Planning and execution of experiments;

Production of scientific papers.

Deliverables:

Conceptual models for describing, in the Semantic Web, the context of interactive stories in virtual worlds;

Tool for specifying the context of interactive stories in accordance with the proposed model;

Interactive Storytelling tool for composing interactive stories in accordance with the proposed model;

Papers submitted to national and international journals or conferences;

Master's Dissertations.

G3. Interactive storytelling for cultural content production and e-learning in the Web

Activities:

Bibliographical research on proposals for modeling content for e-learning and cultural production;

Development of model(s) for specifying e-learning and cultural content by means of Semantic Web ontologies and taking into account their automatic composition;

Specification and implementation of algorithms for composing e-learning and cultural content based on automated planning and content personalization;

Implementation of prototype(s);

Planning and execution of experiments;

Production of scientific papers.

Deliverables:

Methodology(ies), technique(s) and software prototype(s) for composing content based on automated planning and content personalization

Papers submitted to national and international journals or conferences;

Master's Dissertations

Timeframe

		Y1		Y2		Y3	Y4	Y5
Goal	Activity	S1	S2	S3	S4			
Interactive storytelling for Web-based	Bibliographical research on approaches for modeling virtual worlds and interactive stories	X						

collaborative virtual worlds	Studies and definition of models for describing, by means of Semantic Web ontologies, the logical and graphical elements of interactive stories in virtual worlds, including scenarios, characters, goals and events	X	X					
	Implementation of a tool for specifying the context of interactive stories in accordance with the proposed model			X	X			
	Implementation of an Interactive Storytelling tool for composing interactive stories in accordance with the proposed model				X	X		
	Planning and execution of experiments						X	X
	Production of scientific papers			X	X	X	X	X
Interactive storytelling for cultural content production and e-learning in the Web	Bibliographical research on proposals for modeling content for e-learning and cultural production	X						
	Development of model(s) for specifying e-learning and cultural content by means of Semantic Web ontologies and taking into account their automatic composition		X	X				
	Specification and implementation of algorithms for composing e-learning and cultural content based on automated planning and content personalization			X	X			
	Implementation of prototype(s)				X	X	X	
	Planning and execution of experiments						X	X
	Production of scientific papers			X	X	X	X	X
Investigate the Web as a means for eScience and eLearning through CVEs	Study CVEs	X	X					
	Study eScience and eLearning technologies on the Web	X	X					
	Develop CVEs prototypes			X	X			
	Plan and conduct experiments to evaluate the prototypes				X	X	X	X
	Elaborate models to further the development of the Web for enabling eScience and eLearning					X	X	X

Team

Name	**Affiliation**	**Degree**	**CNPq level**
Alberto Barbosa Raposo	PUC-Rio	D.Sc., UNICAMP, 2000	–
Angelo E. M. Ciarlini	UNIRIO	D.Sc., PUC-Rio, 1999	—
Antonio Luz Furtado	PUC-Rio	PhD, U. Toronto, 1974	—
Bruno Feijó	PUC-Rio	PhD, U. of London, Imperial College,	1C

		England, 1988	
Cesar Tadeu Pozzer	UFSM	D.Sc., PUC-Rio, 2005	—
Hugo Fuks	PUC-Rio	PhD, Univ. of London, Imperial College, 1991	1D
Marcelo Gattass	PUC-Rio	PhD, Harvard University, 1979	1A
Sean Wolfgand Matsui Siqueira	UNIRIO	D.Sc., PUC-Rio, 2005	—
Simone Diniz Junqueira Barbosa	PUC-Rio	D.Sc., PUC-Rio, 1999	2

Recent publications

CAMANHO, M. M. ; CIARLINI, A. E. M. ; FURTADO, A. L. ; POZZER, C.T. ; FEIJÓ, B. Conciliating Coherence and High Responsiveness in Interactive Storytelling. In: Proc. 3rd, ACM International Conference on Digital Interactive Media in Entertainment and Arts (DIMEA 2008), Atenas, 2008.

CASTRO, T., FUKS, H., SPÓSITO, M. & CASTRO, A. The Analysis of a Case Study for Group Programming Learning. ICALT - Proc. of the 8th IEEE International Conference on Advanced Learning Technologies, July 1-5, 2008, Santander, Spain.

CIARLINI, A. E. M. ; CAMANHO, M. M. ; DORIA, T. R. ; FURTADO, A. L. ; POZZER, C.T. ; FEIJO, B. Planning and Interaction Levels for TV Storytelling. In: Proc. 1st. Joint International Conference on Interactive Digital Storytelling, Erfurt, Alemanha, 2008. (a ser publicado)

CUNHA, M., RAPOSO, A., FUKS, H. Educational Technology for Collaborative Virtual Environments. **Proceedings of the 12th International Conference on CSCW in Design – CSCWD 2008**, Vol. II, p. 716-720. Xi'an, China.

FUKS, H., RAPOSO, A. B., GEROSA, M.A., LUCENA, C. J. P. Applying the 3C Model to Groupware Development. **International Journal of Cooperative Information Systems (IJCIS)**, v.14, n.2-3, p.299-328, June-September 2005. World Scientific (ISSN 0218-8430).

FUKS, H., PIMENTEL, M. & LUCENA, C.J.P. R-U-Typing-2-Me? Evolving a chat tool to increase understanding in learning activities. International Journal of Computer-Supported Collaborative Learning, Volume 1, Issue 1. ISSN: 1556-1607 (Paper) 1556-1615 (Online). Springer: Mar 2006. pp. 117-142.

FUKS, H., RAPOSO, A. B., GEROSA, M.A., PIMENTEL, M., FILIPPO, D., LUCENA, C. J. P. Inter- e Intra-relações entre Comunicação, Coordenação e Cooperação. **Anais do IV Simpósio Brasileiro de Sistemas Colaborativos – SBSC**, p.57-68 (ou **Anais do XXVII Congresso da Sociedade Brasileira de Computação**, p. 1789-1800). Rio de Janeiro, RJ, Brasil, 2007.

FUKS, H., RAPOSO, A. B., GEROSA, M.A., PIMENTEL, M., FILIPPO, D., LUCENA, C. J. P. Inter- and Intra-Relationships between Communication Coordination and Cooperation in the Scope of the 3C Collaboration Model. **Proceedings of the 12th**

International Conference on CSCW in Design – CSCWD 2008, Vol. I, p. 148-153. Xi'an, China.

FUKS, H., RAPOSO, A., GEROSA, M.A., PIMENTEL, M. & LUCENA, C.J.P. The 3C Collaboration Model. **In:** Ned Kock (org), **Encyclopedia of E-Collaboration**, p. 637-644, Information Science Reference, New York, 2008 (ISBN 978-1-59904-000-4).

GEROSA, M. A., PIMENTEL, M. G., RAPOSO, A. B., FUKS, H., LUCENA, C. J. P., Towards an Engineering Approach for Groupware Development: Learning from the AulaNet LMS Development. **The 9th International Conference on CSCW in Design – CSCWD 2005**, p. 329-333. Coventry, U.K., 2005.

LIMA, L. S., KARLSSON, B., RAPOSO, A. B., SANTOS, I. H. F. Uma Ferramenta de Videoconferência para apoiar Múltiplas Sessões de Trabalho Colaborativo. **XIII Brazilian Symposium on Multimedia and the Web – WebMedia 2007**, p.282-289. Gramado, RS, Brasil, 2007.

LUCENA, C. J. P., FUKS, H., RAPOSO, A. B., GEROSA, M. A., PIMENTEL, M. G., Communication, Coordination and Cooperation in Computer-Supported Learning: The AulaNet Experience. **In:** F. M. Mendes Neto and F. V. Brasileiro (eds.), **Advances in Computer-Supported Learning**, p.274-297. Information Science Publishing, Hershey, PA, USA, 2007 (ISBN 1-59904-356-4).

OLIVEIRA, E. W. ; SIQUEIRA, S. W. M. ; BRAZ, M.H.L.B. Structuring Segments of E-Learning Multimedia Objects with LOM and MPEG-7. In: The 11th IASTED International Conference on Computers and Advanced Technology in Education (CATE 2008), 2008

OLIVEIRA, E. W. ; SIQUEIRA, S. W. M. ; BRAZ, M.H.L.B. . Using OWL to Represent Metadata of Multimedia Learning Objects. In: Webmedia 2008, (a ser publicado)

PIMENTEL, M., ESCOVEDO, T., FUKS, H. & LUCENA, C.J.P. Investigating the assessment of learners' participation in asynchronous conference of an online course. 22nd ICDE - World Conference on Distance Education: Promoting Quality in On-line, Flexible and Distance Education (CD-ROM), Publisher: ABED, Rio de Janeiro, Brazil, Sep, 3-6, 2006.

RUSSO, E. E. R., RAPOSO, A. B., FERNANDO, T., GATTASS, M. A Multiple-Perspective Architecture for CSCW Applications. **7th International Conference on the Design of Cooperative Systems – COOP 2006**, Proceedings Supplement, p.131-138. Carry-le-Rouet, France, 2006.

SANTOS, I. H. F., GÖBEL, M., RAPOSO, A. B., GATTASS, M. A Multimedia Workflow-Based Collaborative Engineering Environment for Oil & Gas Industry. **ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry - VRCAI2004**, p.112-119. Cingapura, 2004.

SANTOS, I. H. F., RAPOSO, A.B., GATTASS, M. A Service Oriented Architecture for a Collaborative Engineering Environment in Petroleum Engineering. **In**: X. Fischer and D. Coutellier (eds.), **Research in Interactive Design - Vol. 2** (Proceedings of the Virtual Concept Conference 2006), Springer Verlag, France, 2006, p. 110 (full paper in CD-ROM). (ISBN-10 2-287-48363-2 / ISBN-13 978-2-287-48363-9).

SANTOS, I. H. F., RAPOSO, A. B., GATTASS, M. A Software Architecture for an Engineering Collaborative Problem Solving Environment. **32nd Annual IEEE Software Engineering Workshop**. Kassandra, Greece. To be published.

SIQUEIRA, S.W.M. ; BRAZ, M.H.L.B. ; MELO, R.N. Modeling e-Learning Content. International Journal of Web Information Systems, v. 3, p. 140-152, 2007.

Relevant theses and dissertations

Completed

Cesar Tadeu Pozzer. **Um Sistema para Geração, Interação e Visualização 3D de Historias para TV Interativa**. 2005. Tese (Doutorado em Informática) - Pontifícia Universidade Católica do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico. Orientador: Bruno Feijó. Co-Orientadores: Antonio Luz Furtado e Angelo Ernani Maia Ciarlini.

Diva de Souza e Silva. **Modelagem de Composição de Objetos de Aprendizagem**. 2006. Tese (Doutorado em Informática) - Pontifícia Universidade Católica do Rio de Janeiro. Co-Orientador: Sean Wolfgand Matsui Siqueira.

Fábio Wanderley Guerra. **Engenharia De Estórias: Um Estudo sobre a Geração e Narração Automática de Estórias.** 2008. Dissertação (Mestrado em Informática) - Pontifícia Universidade Católica do Rio de Janeiro.

FILIPPO, D. Suporte à Coordenação em Sistemas Colaborativos: uma pesquisa-ação com aprendizes e mediadores engajados em fóruns de discussão. Tese de Doutorado, Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), 26 de março de 2008. Advisor: Hugo Fuks.

Raphael Ghelman. **Processamento de Consultas distribuídas à Versões de Materiais de Ensino**. 2006. 0 f. Dissertação (Mestrado em Informática) - Pontifícia Universidade Católica do Rio de Janeiro. Co-Orientador: Sean Wolfgand Matsui Siqueira.

Ongoing

Thiago de Rhamnusia Dória. **Planejamento na Dramatização de Enredos**. Início: 2007. Dissertação (Mestrado em Informática) - Universidade Federal do Estado do Rio de Janeiro. Orientador: Angelo Ernani Maia Ciarlini.

Marcelo de Mello Camanho. **Conciliando Coerência e Responsividade em Storytelling Interativo**. Início: 2007. Dissertação (Mestrado em Informática) - Universidade Federal do Estado do Rio de Janeiro. Orientador: Angelo Ernani Maia Ciarlini.

Edmar Welington Oliveira. (Título provisório) **Estruturando Objetos de Informação a partir de Objetos de Aprendizagem Multimídia**. Início: 2007. Dissertação (Mestrado em Informática) - Universidade Federal do Estado do Rio de Janeiro, Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro. Orientador: Sean Wolfgand Matsui Siqueira.

Fábio Araujo Guilherme da Silva. (Título provisório) **Composição de Conteúdos de Aprendizagem a Partir de Técnicas de Planejamento**. Início: 2008. Dissertação

(Mestrado em Informática) - Universidade Federal do Estado do Rio de Janeiro. Orientador: Sean Wolfgand Matsui Siqueira.

3.3. Management of Web Data

3.3.1. Managing Web Multimedia Data

An important challenge to be met under the Web Science perspective is how to handle the increasing amount of multimedia data generated every second on the Web, for a wide scope of application domains – medicine, remote sensing, biodiversity, biometry and the arts, to name but a few. This has been identified by the Brazilian Computer Society as one of the five Grand Challenges in computer science in Brazil. The very many problems to be dealt with include, for example, the design of new data structures, indexing schemes, and query and retrieval mechanisms. Not only are there large volumes of data, one has now to deal with video, sound, images, to be processed in real time, across a multitude of sites. Tools and services are breaking domain boundaries and local limits as a consequence of progressive data and protocol interoperability, allied to an increasing connectivity.

For this reason, the production, consumption, composition, reuse, customization and management of multimedia content has been attracting increasing attention. Such mechanisms are needed to help spatially distributed groups to work together. There is intensive research on all aspects of multimedia data management (from investigation of data structures, at the storage level, to the design of interaction mechanisms, at the user interface level). Many of the solutions proposed, even when formulated for distributed environments, do not escalate well to the Web. Hence, there are several kinds of research initiatives that investigate how to extend operations on multimedia data to the Web context.

Indeed, the Web is evolving from an environment for publication/consumption of documents to a distributed environment for collaborative work involving any kind of digital content. As a consequence, the “document-oriented'” approach that is still the most common in the Web must be revised to face the increasing diversity of digital content formats and producers. Heterogeneity is an issue that is being daily aggravated – not only can anyone make multimedia data available on the Web, but such data can be reused and composed to produce new data. In summary, the Web infrastructure must be extended to support collaborative work requirements to handle multimedia data, such as content semantics, composition frameworks, replication and modification, configuration management, annotation mechanisms, and version control.

Another important research topic is concerned with mechanisms to retrieve multimedia data on the Web efficiently (performance issues) and effectively (finding relevant data). In particular, two issues will be exploited: how to describe such data in a compact way (i.e., extracting from a given data item the so-called *data descriptions*), and how to annotate it to enhance semantics.

Thanks to the Web, data may be retrieved and combined in an *ad hoc* way, from any source available in the world, extrapolating their local context. Usually, the search for these data is done by their syntactic content, focusing primarily in keyword

matching. This can lead to retrieval of irrelevant multimedia data. Since keywords are user- and context-dependent, many alternative solutions are being explored to allow multimedia data retrieval on the Web, for subsequent reuse. One such direction is the so called *content-based retrieval*, an area with an increasing number of applications. Originally restricted to images, it is being progressively extended to other kinds of media, in special video (including research in computer vision). This kind of research is based on the notion of similarity – given a large set of multimedia data items (usually images), the user wants to retrieve the items which are most similar to a given input item.

Image descriptors can be characterized as a pair of: (i) an algorithm that extracts relevant characteristics from an image, based on image processing, producing a point in a multi-dimensional space; and (ii) some kind of similarity function which computes the distance between two descriptors. This basic concept is being extended in many ways. First, many descriptors can be extracted from the same image (e.g., concerning color, shape or texture) and combined to improve the system effectiveness. Second, distinct user profiles, and even user contexts, can give origin to different retrieval strategies (thereby supporting context-based retrieval, in which distinct sets of images can be returned for the same input, depending on the user context - e.g., classroom, home, leisure – and feedback). Third, content-based analysis has recently started to be adopted for other kinds of media, such as video and sound, presenting the additional challenges of temporal and dynamic evolution.

Semantic interoperability is also a key issue in discovery, access and effective search for multimedia data in different application contexts. While extracted multimedia data descriptions support semantic retrieval up to a certain level, they are not enough. The proposal, here, is to describe multimedia data through a combination of content descriptors and annotations. Distinct kinds of annotation will be considered – (1) the so-called textual *semantic annotations*, (2) sketches (e.g., using pen interfaces), and (3) images (i.e., having an image annotate a multimedia data item). While research in semantic annotations started a couple of years ago, little has been done about other kinds of annotation, in spite of their wide use in real life.

This term semantic annotation has several meanings. Here the definition used will be: a semantic annotation is a set of one or more metadata fields, where each field describes a given multimedia item using ontology terms. An ontology formally describes the elements of a domain and the relationships among them, providing a common understanding of the domain.

Semantic annotations are subject of much research, in distinct contexts. Their use has many goals, such as data discovery, integration and adding meaning to data. Most of this research focuses on annotation of textual resources, without considering multimedia content. As ontologies can change, and new versions be created, there is a need of tools for their management. It is also important to provide a tool for the annotation activity itself, supporting the creation, edition and storage of annotations. This tool must consider that annotations can be attached to entire multimedia items, or parts thereof. In this scenario, a given multimedia content can be summarized by a set of descriptions, and be furthermore annotated in many ways.

The research effort under this subtopic has two major expected contributions. First, it will define algorithms to extract descriptions of multimedia data, especially images and video, and storage structures to support the reuse and composition of such data on the Web. Second, it will design and develop mechanisms to annotate multimedia data, for the three kinds of annotation envisaged, and to index and retrieve these data using such annotations, having in mind educational uses on the Web.

Goals, Activities, and Deliverables

G1. Definition of new multimedia data descriptors and of storage structures to support their indexing, reuse and composition on the Web

Activities:

- Design of algorithms to extract data descriptors for images and video on the Web
- Specification and implementation of a suite of modules to support content-based image and video retrieval, supporting relevance feedback to accommodate user contexts;
- Definition of storage structures to support reuse and composition of multimedia data on the Web, starting from the descriptors proposed
- Evaluation of the effectiveness of the descriptors for content-based retrieval on the Web

Deliverables:

- A suite of algorithms to extract image and video descriptors, to allow content-based retrieval on the Web, supporting relevance feedback and multiple user contexts.
- Journal and conference papers describing the algorithms and their evaluation.
- M.Sc. dissertations and D.Sc. theses on these topics.

G2. Design and development of mechanisms to annotate multimedia data, and to index and retrieve such annotations, having in mind educational uses on the Web

Activities:

- Analysis of non-textual annotation paradigms for multimedia data, to assess their use on the Web
- Specification and implementation of modules for semi-automatic image and video annotation
- Design and development of data structures for multimedia data mining on the Web, combining the descriptors developed in G1 and annotations created in the previous activity
- Design and development of a multimedia data annotation framework, to be used in educational environments, for distributed and collaborative content annotation
- Evaluation of the performance of the annotation modules and the framework, involving all universities in the Institute

Deliverables:

A Web-based tool to annotate multimedia data, applied to educational environments, available through the Institute's Web site

Journal and conference papers, describing the annotation mechanisms and the use of these mechanisms to mine and compose multimedia data on the Web

M.Sc. dissertations and D.Sc. theses based on the annotation mechanisms proposed

Timeframe

		Y1		Y2		Y3	Y4	Y5
Goal	Activity	S1	S2	S3	S4			
Definition of data descriptors and of storage structures to support reuse and composition of multimedia data on the Web	Design of algorithms to extract data descriptors for images and video on the Web	X	X					
	Specification and implementation of a suite of modules to support content-based image and video retrieval, supporting relevance feedback to accommodate user contexts		X	X	X			
	Definition of storage structures to support reuse and composition of multimedia data on the Web, starting from the descriptors proposed		X	X	X			
	Evaluation of the effectiveness of the descriptors for content-based retrieval on the Web	X	X	X	X	X		
Design and development of mechanisms to annotate multimedia data, and to index and retrieve such annotations, having in mind educational uses on the Web.	Analysis of non-textual annotation paradigms for multimedia data, to assess their use on the Web	X	X					
	Specification and implementation of modules for semi-automatic image and video annotation		X	X		X		
	Design and development of data structures for multimedia data mining on the Web, combining the descriptors developed in G1 and annotations created in the previous activity			X	X	X		
	Design and development of a multimedia data annotation framework, to be used in educational environments, for distributed and collaborative content annotation					X	X	X
	Evaluation of the performance of the annotation modules and the framework in an educational context, involving all universities in the Institute							X

Team

Name	**Affiliation**	**Degree**	**CNPq level**
Claudia Maria Bauzer Medeiros	IC-UNICAMP	PhD, U. Waterloo, 1985	1B
Ricardo da Silva Torres	IC - UNICAMP	D.Sc., UNICAMP, 2004	2
Rodolfo Jardim de Azevedo	IC- UNICAMP	D.Sc, UNICAMP, 2002	2

Siome Klein Goldenstein	IC-UNICAMP	PhD. University of Pennsylvania, 2002	2

Recent publications

ALMEIDA, P.; AZEVEDO, R.. INCREASING TABLET PC ACCEPTANCE BY USING SCREEN CAPTURE. In: III Workshop on the Impact of Pen-Based Technology in Education. West Lafayete - EUA. 2008. Accepted

DALTIO, J.; MEDEIROS, C.B. AONDÊ: AN ONTOLOGY WEB SERVICE FOR INTEROPERABILITY ACROSS BIODIVERSITY APPLICATIONS. Information Systems, v. 33, 2008. (Accepted for publication)

SANTANCHE, A.; MEDEIROS, C. B.. PASTORELLO JR., G. Z.; USER-AUTHOR CENTERED MULTIMEDIA BUILDING BLOCKS. Multimedia Systems, Berlin, v. 12, p. 403-421, 2007.

TORRES, R. da S. ; FALCAO, A. X. ; GONCALVES, M. A. ; Papa, J. P. ; ZHANG, B. ; FAN, W. ; Fox, E. A. A GENETIC PROGRAMMING FRAMEWORK FOR CONTENT-BASED IMAGE RETRIEVAL. Pattern Recognition, 2008.

TORRES, R. da S. ; MEDEIROS, Claudia B ; GONCALVES, Marcos Andre ; FOX, E. A.. A DIGITAL LIBRARY FRAMEWORK FOR BIODIVERSITY INFORMATION SYSTEMS. International Journal on Digital Libraries, 6:3-17, 2006

VOGLER, Christian ; GOLDENSTEIN, S. ; STOLFI, J. ; PAVLOVIC, V. ; METAXAS, Dimitris . Outlier Rejection in High-Dimensional Deformable Models. Image and Vision Computing, v. 25, p. 274-284, 2007.

Relevant theses and dissertations

Completed

A. Santanchè: The Fluid Web and Digital Content Components: from a document-centered to a content-centered view. PhD Thesis, UNICAMP, 2006 (Orientador C. B. Medeiros)

R. da S. Torres, Ambiente de Gerenciamento de Imagens e Dados Espaciais ara Desenvolvimento de Aplicações em Biodiversidade, Phd Thesis, UNICAMP, 2004. (Orientador C. B. Medeiros)

F. A. Andaló. Descritores de Forma baseados em Tensor Scale. MSc degree, UNICAMP, 2007. (Orientador: Ricardo da Silva Torres).
C. D. Ferreira. Recuperação de Imagens por Conteúdo com Realimentação de relevância baseada em Programação Genética. MSc degree, UNICAMP, 2007. (Orientador: Ricardo da Silva Torres).

Ongoing

C. Macario. Um framework para anotacoes de conteudo digital na Web, para aplicacoes agricolas. Inicio: 2006 PhD Thesis, UNICAMP (Orientador C. B. Medeiros)

J. G. Malaverri. Servicos de metadados para sistemas de biodiversidade. Início: 2007. MSc, UNICAMP. (Orientador C. B. Medeiros).

P. Almeida. Suporte ao Ensino através de TabletPC. Inicio: 2007. Msc, UNICAMP (Orientador Rodolfo Azevedo).

N. P. Kozievitch. Gerenciamento de Dados Complexos em Bibliotecas Digitais. PhD Thesis, UNICAMP. (Orientador Ricardo da Silva Torres).

R. Caceffo. Ferramenta colaborativa para aprendizado, utilizando computação pen-based e móvel. Inicio: 2008. MSc, UNICAMP (Co-orientador Rodolfo Azevedo)

A. Rocha. Informacões de alto e baixo nível para inferência em imagens. Início 2008. PhD. UNICAMP (Orientador S. Goldenstein)

F. Andaló. Simulação e Aprendizado de Multidões. Início 2007. PhD. UNICAMP (Orientador S. Goldenstein)

E. Morais. Visualização e Simulação de Fenomenos Naturais. Início 2008. PhD. UNICAMP (Orientador S. Goldenstein)

O. A. B. Penatti. Estudo Comparativo de Descritores para Recuperação de Imagens por Conteúdo na Web. Inicio 2007. MSc UNICAMP. (Orientador Ricardo da Silva Torres).

J. A. dos Santos. Reconhecimento e Vetorização Automática de Entidades Geográficas em Imagens de Sensoriamento Remoto. MSc, Inicio 2007. UNICAMP. (Orientador Ricardo da Silva Torres).

3.3.2. Accessing the Deep Web

Unlike the "Surface Web" of static pages, the "Deep Web" comprises data stored in databases, dynamic pages, scripted pages and multimedia data, among other types of objects. Estimates suggest that the size of the Deep Web greatly exceeds that of the Surface Web – with nearly 92,000 terabytes of data on the Deep Web versus only 167 terabytes on the Surface Web, as of 2003.

In particular, Deep Web databases are typically under-represented on search engines due to the technical challenges of locating, accessing, and indexing the databases. Indeed, since Deep Web data is not available as static Web pages, traditional search engines cannot discover data stored in the databases through the traversal of hyperlinks, but rather they have to interact with (potentially) complex query interfaces.

A strategy to uncover Deep Web databases is to build *query mediators* that are specifically designed to locate Web-enabled databases and to interact with their query interfaces. A query mediator typically offers a *mediated schema*, which describes the data available through its interface and which captures aspects of the specific application domain the mediator covers. Among other problems, a query mediator should:

(1) Locate databases available on the Web that pertain to the application domain it covers (the *database identification* problem).

(2) Build the mediated schema as it identifies new databases about the application domain it covers (the *mediated schema definition* problem).

(3) Match the database schemas with the mediated schema, that is, develop a set of mappings between the mediated schema and the database schemas (the *schema matching* problem).

(4) Translate queries submitted to the mediated schema into queries over the database schemas, and translate back the answers received from the databases (the *query processing* problem).

(5) Interact with Semantic Web search engines, making the mediated schema available to the search engine in a suitable way and the accepting *semantic searches* over the mediated schema (the *semantic search processing* problem).

We will follow three basic approaches to solve the schema matching and the mediated schema definition problems, which are closely related. The first approach, that we call *syntactical a posteriori*, is to match two schemas and to construct the mediated schema based on syntactical hints, such attribute data types and naming (syntactical) similarities. This approach depends on the implicit assumption that syntactical proximity implies semantic similarity. The second approach, that we call the *semantic approach*, uses semantic clues to generate hypotheses about schema matching and about the definition of the mediated schema. It generally tries to detect how the same real world objects are represented in different databases and leverages on the information obtained to match the schemas and to define the mediated schema. The third approach, that we call *design a priori*, emphasizes that, whenever specifying databases that will be exposed on the Web, the designer should start by selecting an appropriate standard, if one exists, to guide the design of the database schemas and the mediated schema. If carefully applied, the *design a priori* approach transforms schema matching and mediated schema definition into much more tractable problems.

The research effort under this subtopic has two major expected contributions. First, it will define heuristics to solve the database identification, the mediated schema definition and the schema matching problems. Second, it will develop query mediators that incorporate such heuristics and investigate their interactions with semantic Web search engines.

Goals, Activities, and Deliverables

G1. Definition of heuristics for database identification, mediated schema definition and schema matching

Activities:

Development of heuristics to solve the database identification problem.

Development of heuristics to solve the mediated schema definition and the schema matching problems, based on the syntactic a posteriori, semantic a posteriori and design a priori approaches.

Evaluation of the heuristics.

Deliverables:

Heuristics to solve the database identification, the mediated schema definition and the schema matching problems, published on the Web site of the institute.

Journal and conference papers describing the heuristics and their evaluation.

M.Sc. dissertations and D.Sc. theses based on the heuristics and their evaluation.

G2. Development of query mediators

Activities:

Development of query processing heuristics, driven by schema mappings.

Development of semantic search processing heuristics, based on publishing the mediated schema as a searchable ontology.

Definition of a mediator framework having as "hot spots" the database identification, the mediated schema definition and the schema matching heuristics, as well as the query processing and the semantic search processing heuristics.

Instantiation of the mediator framework with the heuristics with the best potential.

Evaluation of the performance of the final mediator using real Deep Web databases.

Deliverables:

Heuristics for query processing, driven by the schema mappings, and published on the Web site of the institute.

Heuristics for semantic search processing, based on publishing the mediated schema as a searchable ontology, and published on the Web site of the institute.

A mediator framework, available through the Web site of the institute.

A mediator for Deep Web databases, available through the Web site of the institute.

Journal papers, as well as conference papers, describing the query processing and the semantic search processing heuristics, the mediator framework and the final mediator implemented.

M.Sc. dissertations and D.Sc. theses based on describing the query processing and the semantic search processing heuristics, the mediator framework and the final mediator implemented.

Timeframe

		Y1		Y2		Y3	Y4	Y5
Goal	Activity	S1	S2	S3	S4			
Heuristics for database identification, mediated schema definition and schema matching	Definition of heuristics to solve the database identification problem.	X	X					
	Definition of heuristics to solve the mediated schema definition and the schema matching problems, based on the syntactic, semantic and a posteriori approaches.		X	X				
	Evaluation of the heuristics	X	X	X				
Query mediators	Development of query processing heuristics, driven by schema mappings.			X	X			

	Development of semantic search processing heuristics, based on publishing the mediated schema as a searchable ontology.					X		
	Definition of a mediator framework having as “hot spots” the database identification, the mediated schema definition and the schema matching heuristics, as well as the query processing and the semantic search processing heuristics.					X		
	Instantiation of the mediator framework with the heuristics with the best potential.						X	X
	Evaluation of the performance of the final mediator using real Deep Web databases.							X

Team

Name	**Affiliation**	**Degree**	**CNPq level**
Antonio Luz Furtado	PUC-Rio	PhD, U. Toronto, 1974	—
Karin Koogan Breitman	PUC-Rio	D.Sc., PUC-Rio, 2000	—
Marco Antonio Casanova	PUC-Rio	PhD, Harvard U., 1979	1C
Michael G. Hinchey	LERO	PhD, U. Cambridge	—
Rubens Nascimento Melo	PUC-Rio	D.Sc., ITA, 1976	

Recent publications

BRAUNER, Daniela F. ; GAZOLA, A. ; CASANOVA, M. A. ; BREITMAN, Karin K . ADAPTATIVE MATCHING OF DATABASE WEB SERVICES EXPORT SCHEMAS. In: International Conference on Enterprise Information Systems., 2008, Barcelona. Proceedings of ICEIS 2008 Tenth International Conference on Enterprise Information Systems., 2008.

BREITMAN, Karin K, BRAUNER, D.F., CASANOVA, M. A., MILIDIU, Ruy, GAZOLA, A., PERAZOLO, M. Instance-Based Ontology Mapping In: 5th IEEE Workshop on Engineering of Autonomic and Autonomous Systems, 2008, Belfast.

BREITMAN, Karin K ; CASANOVA, M. A. ; TRUSZKOWSKI, Walter . Semantic Web: Concepts, Technologies and Applications. Londres: Springer, 2007. v. 1. 337 p.

CASANOVA, M. A. ; BREITMAN, Karin Koogan ; BRAUNER, Daniela F. ; Marins, A. . Database Conceptual Schema Matching. Computer (Long Beach), v. 40, p. 102-104, 2007.

BRAUNER, Daniela F.; CASANOVA, M. A.; MILIDIU, Ruy. Towards Gazetteer Integration Through an Instance-based Thesauri Mapping Approach In: Advances in Geoinformatics ed.Heidelberg : Springer, 2007, p. 235-245.

Relevant theses and dissertations

Completed

Daniela Francisco Brauner. **Alinhamento de esquemas baseado em instâncias.** 2008. Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática.

Alexandre Gazola. **Uma infra-estrutura de software para alinhamento de catálogos heterogêneos**. 2008. Dissertação (Informática) - Pontifícia Universidade Católica do Rio de Janeiro.

André Luiz Almeida Marins. **Modelos Conceituais para Proveniência**. 2008. Dissertação (Informática) - Pontifícia Universidade Católica do Rio de Janeiro.

Luiz André Paes Leme. **Uma arquitetura de software para catalogação automática de dados geográficos**. 2006. Dissertação (Informática) - Pontifícia Universidade Católica do Rio de Janeiro.

Daniela Francisco Brauner. **Uma Arquitetura para Catálogos de Objetos baseados em Ontologias**. 2005. Dissertação (Informática) - Pontifícia Universidade Católica do Rio de Janeiro.

Ongoing

Luiz André Paes Leme. **Funções de semelhança para alinhamento de esquemas.** 2009. Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática.

Bernardo Pereira Nunes. **Classificação automática de instâncias em bancos de dados**. 2009. Dissertação (Informática) - Pontifícia Universidade Católica do Rio de Janeiro.

3.3.3. Developing the Semantic Web

The Web, as implemented today, may be defined as the Syntactic Web, where data presentation is carried out by computers, and the interpretation and identification of relevant data is delegated to human beings. Of course, the interpretation process is very demanding and requires considerable effort to evaluate, classify, and select relevant data. Because the volume of available data is growing at an exponential rate, it is becoming virtually impossible for human beings to manage the complexity and volume of the available data. This phenomenon, often referred to as information overload, poses a serious threat to the very usefulness of today's Web. The question is: Why can't computers do this job for us?

One of the reasons resides in the fact that Web pages do not contain data about themselves, that is, about their contents and the subjects to which they refer.

Web search engines help to identify relevant Web pages. They can be compared using various parameters, be it their coverage (the number of hits returned given a query, particularly looking at the number of hits only achieved by that search engine); the relevance of the pages returned; the time taken; or the quality of returns. As one would expect, different engines do well on different metrics.

A related concept to the use of a particular Web resource in a process is its *significance*.

Significance can be decomposed into two types of metric: *relevance* and *quality*. Relevance is connected to the idea of querying: how many queries does a page handle? The different ways of answering that question have led to the development of a number of important algorithms, but basically the idea is that a page handles a query when it either contains information relevant to the query, or points the reader to a resource that contains such information. One approach is to look at the hyperlink structures that provide the context for Web pages and to try to deduce measures of relevance from those structures.

Together with relevance metrics, quality metrics can then rank the results of searches. The most famous quality measure is PageRank. The value of these measures is reflected in the success of Google in terms of longevity, market value and share of the search engine market. Further, other measures of quality exploit the idea of random walks, sometimes explicitly extending the ideas underlying PageRank [Berners-Lee et al. 2006].

Despite these benefits, search engines suffer from the following limitations:

Search results might contain a large number of entries, but they might have low recall precision.

Search results are sensitive to the vocabulary used. Indeed, users frequently formulate their search in a vocabulary different from that which the relevant Web pages adopt.

Search results appear as a list of references to individual Web pages. However, it is often the case that, among the Web pages listed in the search result, there are many entries that belong to the same Web site. Conversely, if the relevant information is scattered in more than one entry, it is difficult to determine the complete set of relevant entries.

The conclusion is that the Web has evolved as a medium for information exchange among people, rather than machines. As a consequence, the semantic content, that is, the meaning of the data in a Web page, is coded in a way that is accessible to human beings alone.

In 2001, Berners-Lee, Hendler, and Lassila published a revolutionary article in *Scientific American*, entitled “The Semantic Web: A New Form of Web Content That Is Meaningful to Computers Will Unleash a Revolution of New Possibilities”. In this article, the authors describe future scenarios in which the Semantic Web will have a fundamental role in the day-to-day life of individuals.

The Semantic Web depends on developing technologies to better describe and access content, and yet retain the freedom to organize the content. Among such technologies, we may single out ontologies and Web services.

Indeed, in the context of the Semantic Web, ontologies will be responsible for making the semantics of pages and applications explicit, thus allowing electronic agents to process and integrate resources automatically. With the widespread, distributed use of ontologies, different parties will inevitably develop ontologies with overlapping content. The ability to integrate different ontologies meaningfully is thus critical to assure coordinated action in multi agent systems. In this topic, we propose to investigate strategies and tools that allow for fully automatic ontology alignment for the Semantic Web.

Moreover, problems regarded to Web information quality involve many different users' profiles. Similarly as with Internet users these problems occur in complex corporate environments that keep their proper structures as one or more intranets. The awareness of problems inherent to this low information quality is the first obstacle to looking successfully for alternative solutions. Impacts more tangible such as: dissatisfactions of users and organizations; increase of costs; inefficient decision making processes; and reduction of the ability to execute the strategy are sufficiently bad already. Other minor impacts worsen these problems. Although, not generally adopted mechanisms to improve quality level of Web information are available and have been effectively applied in some projects.

The possibility to evaluate and rank Web documents based on their quality criteria is fundamental for Web Science. It leads to new opportunities and more efficient strategies when, due to overload problems, information consumers decide by filtering and personalizing search results according to their quality level perspectives. Therefore, considering that information quality depends on contexts, they should be adopted, together with metadata, as a base to quality evaluation indicators.

Web services technology supports application-to-application communication over the Internet. These services provide standardized interfaces that can be composed together in order to define more complex services. Service-oriented architectures based on Web services, however, suffer from two major problems: they depend on the availability of the individual Web services and their correct behavior, and they lack the flexibility to replace a failed Web service with a redundant alternative. If one of these services is not available or fails during execution, a dispatcher component needs to be able to dynamically switch to alternative Web services that provide equivalent functionality in order to fulfill the consumer request. The replacement of a service composition with an alternative composition increases the functional complexity involved and also increases the response time of an invocation.

Recent research activities concentrated on improving the service infrastructure by removing single points of failure by providing dynamic selection mechanisms based on consumer preferences [Casati & Shan 2001, Li et al. 2006]. Static composition describes the process of designing a workflow based on well-known existing services. BPEL has become one of the accepted standard languages for specifying the behavior of business processes based on the interaction of Web services [Milanovic & Malek 2004].

Although it is considered a standard, the static composition supported by BPEL is considered rigid, since it cannot cope with a frequently changing environment. Motivated by the requirement of more dynamic service-oriented architectures, we propose an extension of BPEL in order to allow the selection of services according to dynamic conditions of the system infrastructure. Consequently, the best currently available service can be selected at runtime, based on criteria that can change dynamically. The dynamic selection rule requires the management of conversations with the dynamically selected services. Other researchers, such as Li et al. (2006) and Casati & Shan (2001) present models of dynamic service composition, but based on static selection criteria. The contributions presented in our proposal are twofold. First, we propose that the criteria of dynamic composition should also be dynamically updated. Having dynamic criteria for composing services, it is possible to prioritize different levels of resilience, according to the user contract. Second, we also propose the dynamic selection of service's orchestration scenarios (not only service's selection). These scenarios should represent different strategies of fault tolerance that can be selected dynamically, depending on the current state of the system infrastructure.

Goals, Activities, and Deliverables

G1. Definition of a method to develop reference domain ontologies based on the composition and alignment of existing (upper level) ontology fragments.

Activities:

Elaborate a survey on alignment techniques with a focus on ontology

Definition of a method to develop reference domain ontologies

Implement a software tool to support the method

Validate the method and the tool with existing ontologies

Deliverables:

A method to develop reference domain ontologies

A software tool to support the method, available through the Web site of the institute

Journal papers, as well as conference papers, describing the method and the tool

M.Sc. dissertations and D.Sc. theses based on describing the method and the tool

G2. Definition of a method for the description, discovery and automated composition of semantic Web services

Activities:

Elaborate a survey on languages to describe, discover and automatically compose semantic Web services

Definition of a framework to describe, discover and automatically compose semantic Web services

Instantiation of the framework with specific strategies to describe, discover and automatically compose semantic Web services

Validate the tool with existing Web services

Deliverables:

A framework to describe, discover and automatically compose semantic Web services

A tool to discover and automatically compose semantic Web services

Journal papers, as well as conference papers, describing the framework and the tool

M.Sc. dissertations and D.Sc. theses based on describing the framework and the tool

G3. Design techniques to improve the resilience of service-oriented architectures based on Web services

Activities:

A survey of the existing approaches of dynamic service composition and orchestration. Implementation technologies such as aspect-oriented programming and computational reflection should also be considered;

A survey of languages for specifying service orchestration;

Define a notation for specifying the criteria of dynamic selection of services (dynamic composition);

Define a notation for specifying the criteria of dynamic selection of execution scenarios (dynamic orchestration);

Evaluate the proposed approach in real Web applications;

Elaboration of scientific papers.

Deliverables:

Dissertations and Theses

Papers in international and national conferences

Techniques to improve the development of dynamic and resilient Web applications

G4. Define a framework to support the development of dynamic and resilient Web applications

Activities:

Design and implementation of a tool to support the specification of dynamic criteria of service composition and orchestration;

Select an existing middleware to be extended with the proposed dynamic approach;

Extend the selected middleware with dynamic service composition and orchestration;

Evaluate the proposed framework in a real case study;

Elaboration of scientific papers.

Deliverables:

M.Sc. dissertations

D.Sc. Theses

Papers in international and national conferences

A framework for developing dynamic and resilient Web applications

G5. Design and development of a specification and a model to evaluate, and personalize search results according to users' quality level perspectives

Activities:

Elaborate a survey with a focus on information quality on the Web

Design a specification of a Web Metadata Based-Model for Information Quality Prediction

Develop a Web Metadata Based-Model for Information Quality Prediction

Embody the Model as a final step of a Web search engine running in collaboration with a browser

Data preparation and development of a Web based application

Experimentation planning and preparation

Validation and comparison of the achieved results of the proposed model and approach

Experimentation planning and preparation adopting other huge data sets

Analyze the applicability of the solution in other scenarios

Elaboration of scientific papers

Deliverables:

A Web Metadata Based-Model for Information Quality Prediction

A Web based application that implements the proposed techniques

Journal papers, as well as conference papers, describing the model and the application

M.Sc. dissertations and D.Sc. theses based on describing the model and the application

Organization of one research Lab to develop applications and work with overload information aspects

Timeframe

		Y1		Y2		Y3	Y4	Y5
Goal	Activity	S1	S2	S3	S4			
A method to develop reference domain ontologies	Elaborate a survey on alignment techniques with a focus on ontology	x						
	Definition of a method to develop reference domain ontologies		x					
	Implement a software tool to support the			x	x			

	method							
	Validate the method and the tool with existing ontologies			x	x			
A method for the description, discovery and automated comp. of semantic Web services	Elaborate a survey on languages to describe, discover and automatically compose semantic Web services					x		
	Definition of a framework to describe, discover and automatically compose semantic Web services					x	x	
	Instantiation of the framework with specific strategies to describe, discover and automatically compose semantic Web services						x	x
	Validate the tool with existing Web services							x
Design techniques to improve the resilience of service-oriented architectures based on Web services	A survey of the existing approaches of dynamic service composition and orchestration. Implementation technologies such as aspect-oriented programming and computational reflection should also be considered	X						
	A survey of languages for specifying service orchestration	X						
	Define a notation for specifying the criteria of dynamic selection of services (dynamic composition)	X	X					
	Define a notation for specifying the criteria of dynamic selection of execution scenarios (dynamic orchestration)		X	X				
	Evaluate the proposed approach in real Web applications			X	X			
	Elaboration of scientific papers		X	X	X			
Define a framework to support the development of dynamic and resilient Web applications	Design and implementation of a tool to support the specification of dynamic criteria of service composition and orchestration				X	X		
	Select an existing middleware to be extended with the proposed dynamic				X			
	Extend the selected middleware with dynamic service composition and					X	X	
	Evaluate the proposed framework in a real case study						X	X
	Elaboration of scientific papers					X	X	X
Design and development of a specification and a model to	Elaborate a survey with a focus on information quality on the Web	X						
	Design a specification of a Web Metadata Based-Model for Information Quality Prediction		X					

evaluate and personalize search results according to users' quality level perspectives	Develop a Web Metadata Based-Model for Information Quality Prediction			X	X			
	Embody the Model as a final step of a Web search engine running in collaboration with a browser				X	X		
	Data preparation and development of a Web based application				X	X		
	Experimentation planning and preparation					X	X	
	Validation and comparison of the achieved results of the proposed model and approach						X	X
	Experimentation planning and preparation adopting other huge data sets						X	X
	Analyze the applicability of the solution in other scenarios						X	X
	Elaboration of scientific papers		X	X	X	X	X	X

Team

Name	**Affiliation**	**Degree**	**CNPq level**
Antonio Luz Furtado	PUC-Rio	PhD, U. Toronto, 1974	—
Cecília Mary Fischer Rubira	IC-UNICAMP	PhD, Newcastle University, UK,1994	1D
Flávia Coimbra Delicato	UFRN	D.Sc. UFRJ, 2005	—
Geraldo Bonorino Xexéo	PESC/COPPE	D.Sc., COPPE, 1994	
Jano Moreira de Souza	PESC/COPPE	Ph.D., University of East Anglia, 1986	1D
Karin Koogan Breitman	PUC-Rio	D.Sc., PUC-Rio, 2000	—
Marco Antonio Casanova	PUC-Rio	PhD, Harvard U., 1979	1C
Michael G. Hinchey	LERO	PhD, U. Cambridge, 1995	—
Paulo de Figueiredo Pires	UFRN	D.Sc. UFRJ, 2002	—
Ricardo de Oliveira Anido	IC-UNICAMP	PhD, Imperial College, 1989	—
Rubens Nascimento Melo	PUC-Rio	D.Sc., ITA, 1976	—

Recent publications

BARBOSA, S. J., BREITMAN, Karin K, CASANOVA, M. A., FURTADO, A. L. Similarity and Analogy over Application Domains In: XXII Simposio Brasileiro de Banco de Dados, 2007, João Pessoa.

BARROS, R., XEXÉO, G. B., PINHEIRO, W. A., SOUZA, J. M., A Web Metadata Based-Model for Information Quality Prediction. Chapter XIX of the Handbook of Research on Web Information Systems Quality. Ed. Information Science Reference (an imprint of IGI Global), Hershey, USA. Editores: Coral Calero, Maria Ángeles Moraga, Mario Piattini. 2008.

BARROS, R., XEXÉO, G. B., PINHEIRO, W. A., SOUZA, J. M., User and Context Aware Quality Filters Based on Web Metadata Retrieval. Chapter VIII of the Personalized Information Retrieval and Access: Concept, Methods and Practices. Ed. Idea Book, Hershey, USA. Editores: Ajantha Dahanayake, Nong Chen. 2008.

BREITMAN, Karin K ; CASANOVA, M. A. ; TRUSZKOWSKI, Walter . *Semantic Web: Concepts, Technologies and Applications*. Londres: Springer, 2007. v. 1. 337 p.

BREITMAN, Karin K, BARBOSA, S. J., CASANOVA, M. A., FURTADO, A. L. "Using Analogy to Promote Conceptual Modeling Reuse". In: 2007 ISoLA Workshop On Leveraging Applications of Formal Methods, Verification and Validation - Formal Methods in Avionics, Space and Transport, 2007, Poitiers.

BREITMAN, Karin K, BRAUNER, D.F., CASANOVA, M. A., MILIDIU, Ruy, GAZOLA, A., PERAZOLO, M. "Instance-Based Ontology Mapping". In: 5th IEEE Workshop on Engineering of Autonomic and Autonomous Systems, 2008, Belfast.

BREITMAN, Karin Koogan, FELICISSIMO, C.H., CASANOVA, M. A. "CATO - A Lightweight Ontology Alignment Tool". In: The 17th Conference on Advanced Information Systems Engineering (CAiSE'05), 2005, Porto.

CAMPOS, Linair Maria ; PIRES, P. F. ; CAMPOS, Maria L M . OntoGuide-4WS: Uma Abordagem Sistemática para Descrição de Serviços Web. In: Simpósio Brasileiro de Sistemas Multimídia e Web, 2004, São Carlos. Anais Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia), 2004. v. 2. p. 20-27.

CASANOVA, M. A., BARBOSA, S. J., BREITMAN, Karin K, FURTADO, A. L. "Generalization and Blending in the generation of Entity-Relationship schemas by Analogy". In: International Conference on Enterprise Information Systems., 2008, Barcelona.

Castor Filho, F., Brito, P.H.S. and Rubira, C.M.F.Specification of Exception Flow in Software Architectures. Journal of Systems and Software, Elsevier, North-Holland, ISSN 0164-1212, Vol.79, No.10, October 2006, pp. 1359-1482 (periódico internacional, qualis A).

Castor Filho, F., Guerra, P.A.C., Pagano, V.A. and Rubira, C.M.F.. A Systematic Approach for Structuring Exception Handling in Robust Component-based Software. Journal of the Brazilian Computer Society, Vol. 10, No.3, April 2005, pp. 5-19 , SBC (Sociedade Brasileira de Computação), ISSN 0104-6500. (periódico internacional, qualis B)

CAVALCANTI, M. C. ; BAIÃO, F. ; ROSSIE, S. ; BISCH, P.M. ; PIRES, P. F. ; TARGINO, R. ; CAMPOS, Maria L M ; MATTOSO, Marta L Q . Managing Structural Genomic Workflows using Web Services. Data & Knowledge Engineering, NORTH-HOLLAND, v. 53, n. 1, p. 45-74, 2005.

COSTA, L.A.G. ; PIRES, P. F. ; MATTOSO, Marta L Q . Automatic Composition of Web Services with Contingency Plans. In: 2004 IEEE International Conference on Web Services

(ICWS 2004), 2004, San Diego. Proceedings of 2004 IEEE International Conference on Web Services (ICWS), 2004. p. 454-461.

COSTA, L.A.G. ; PIRES, P. F. ; MATTOSO, Marta L Q . WebComposer: a tool for the composition and execution of Web Service-based Workflows. In: WebMedia & LA-Web 2004 Joint Conference, 2004, São Carlos. Proceedings of 10th Brazilian Symposium on Multimedia and the Web 2nd Latin American Web Congress (Joint WebMedia & LA Web), 2004. p. 30-37.

CRUZ, S.M.S. ; CAMPOS, Linair Maria ; PIRES, P. F. ; CAMPOS, Maria L M . Monitoring E-Business Web Services Usage through a Log Based Architecture. In: 2004 IEEE International Conference on Web Services (ICWS 2004), 2004, San Diego. Proceeding of 2004 IEEE International Conference on Web Services (ICWS), 2004. p. 61-69.

de Lemos, R., Guerra, P.A.C. and Rubira, C.M.F. A Fault-tolerant Architectural Approach for Dependable Systems. IEEE Software, Vol. 23, No. 2, March/April 2006, pp. 80-87, IEEE Computer Society Press, ISSN 0740-7459 (periódico internacional, qualis A)

FREIRE, Cássio José Santos ; PIRES, P. F. ; DELICATO, Flávia Coimbra ; CAMPOS, Maria L M ; PIRMEZ, Luci ; OLIVEIRA, M. V. M. . Regente: Um Arcabouço para Gerenciamento Eficiente de Orquestrações de Serviços Web. In: Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos, 2008, Rio de Janeiro. Anais do 26 Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos. Porto Alegre, RS : SBC, 2008. v. 1. p. 149-162.

LAGES, A. ; DELICATO, Flávia Coimbra ; PIRES, P. F. ; PIRMEZ, Luci . SATYA: A Reputation-based Approach for Service Discovery and Selection in Service Oriented Architectures. In: ACM international workshop on Web information and data management, 2007, LIsboa. Proceedings of the nineth ACM international workshop on Web information and data management (WIDM 07), 2007.

MARINS, A., CASANOVA, M. A., FURTADO, A. L., BREITMAN, Karin K “Modeling Provenance for Semantic Desktop Applications”. In: XXXIV Seminário Integrado de Software e Hardware, 2007, Rio de Janeiro.

MENDES JÚNIOR, J.R. ; PIRES, P. F. ; DELICATO, Flávia Coimbra ; Batista, T.V. . Um Ambiente para Especificação e Execução ad-hoc de Processos de Negócio Baseados em Serviços Web. In: WebMedia 2007 - XIII Brazilian Symposium on Multimedia and the Web, 2007, Gramado. Proceedings of the XIII Brazilian Symposium on Multimedia and the Web. Porto Alegre : SBC, 2007. p. 238-245.

PIRES, P. F. ; MATTOSO, Marta L Q ; BENEVIDES, M.R.F. . Building Reliable Web Services Compositions. In: Revised Papers from the NODe 2002 Web and Database-Related Workshops on Web, Web-Services, and Database Systems, 2003. Lecture Notes on Computer Science. London, UK : Springer Verlag. v. 2593. p. 59-72.

PIRES, P. F. ; AZEVEDO JÚNIOR, V. ; MATTOSO, Marta L Q ; BENEVIDES, M.R.F. . WebTransact-EM: Um Modelo para a Execução Dinâmica de Serviços Web Semanticamente Equivalentes. In: Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia), 2003, Salvador. Anais 2003 Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia), 2003. p. 53-68.

PIRES, P. F. ; MATTOSO, Marta L Q ; BENEVIDES, M.R.F. . Coordenação de Transações em ambientes de Serviços Web. In: Simpósio Brasileiro de Sistemas Multimídia e Web, 2003, Salvador. Anais Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia), 2003. p. 69-84.

PIRES, P. F. ; BENEVIDES, M.R.F. ; MATTOSO, Marta L Q . Mediating Heterogeneous Web Services. In: IEEE Symposium on Applications and the Internet, 2003, Orlando, FL. 2003 IEEE Symposium on Applications and the Internet (SAINT), 2003. p. 344-347.

Rubira, C.M.F., de Lemos, R., Ferreira, G.R.M. & Castor Filho, F. Exception Handling in the Development of Dependable Component-based Systems. Software: Practice & Experience, Volume 35, No.3, March 2005, pp. 195-236, John Wiley & Sons, Ltd. ISSN 0038-0644.(periódico internacional, qualis A)

Relevant theses and dissertations

Completed

André Luiz Almeida Marins**.** 'Modelos Conceituais para Proveniência**".** Dissertação de Mestrado**.** Programa de Pós-Graduação em Informática da PUC-Rio.

AZEVEDO JR, V.. WebTransact-EM: Um Modelo para a Execução Dinâmica de Serviços Web Semanticamente Equivalentes. 2003. 116 f. Dissertação (Mestrado em Engenharia de Sistemas e Computação) - Universidade Federal do Rio de Janeiro.

CAMPOS, L. M.. Uma Abordagem Sistemática de Descrição de Serviços Web para fins de Descoberta Dinâmica. 2004. Dissertação (Mestrado em Informática) - Universidade Federal do Rio de Janeiro.

FREIRE, C.J.S. Regente: Um arcabouço para o Gerenciamento Eficiente de Orquestrações de Serviços Web. 2007. Dissertação (Mestrado em Informática) - Universidade Federal do Rio de Janeiro.

MENDES JR, J.R., WebFlowAH: Um Ambiente para Especificação Ad-Hoc e Execução de Composição Serviços Web. 2008. Dissertação (Mestrado em Informática) - Universidade Federal do Rio Grande do Norte.

PIRES, P.F, WebTransact: A Framework for Building Reliable Web Services Compositions, 2002. Tese (Doutorado em Engenharia de Sistemas e Conputação) - Universidade Federal do Rio de Janeiro.

Ongoing

LOPES, F.A.S. Overlay Middleware baseado em Tecnologias da Web Semântica para Ambientes Ubíquos. Tese de Doutorado. Programa de Pós-Graduação em Sistemas e Computação -Universidade Federal do Rio Grande do Norte.

A Web Metadata Based-Model for Information Quality Prediction. Ricardo Oliveira Barros, Ph. D. Candidate, PESC/COPPE – UFRJ. Expected completion: March, 2009.

3.3.4. Knowledge Discovery on the Web

The exponential growth of the total volume of contents accessible through the Web enables a new generation of innovative applications based on Machine Learning. In addition, the low-cost computational infrastructure based on the Parallel External Memory (PEM) paradigm facilitates the scalability of learning algorithms. The companies Google and Netflix are two emblematic examples of the success in combining such technological opportunities.

In particular, important Computational Linguistics tasks, such as: Part-of-Speech Tagging, Noun Phrase Identification, Text Chunking, Sentence Chunking, Clause Chunking and also Semantic Roles Labeling are currently solved using Machine Learning. These tasks, in ascending order of complexity, are intermediate steps in the discovery and capture of the semantics in textual data bases.

Our research in this topic focuses in two central issues:

1. development of learning algorithms for classification and regressive problems
2. PEM implementation of learning algorithms

Recently, we developed the ETL (Entropy Guided Transformation Learning) algorithm. Two additional tasks are in our research agenda for extending ETL. First, to handle the regression problem using the same approach, generating the model already called R-ETL. Second, to implement ETL in a distributed environment using the Map-Reduce metaphor made available by Hadoop.

Moreover, we are also investigating algorithms for building Ensemble Models. In this line, we have already developed Boosting at Start (BAS), a scheme that generalizes AdaBoost. Using BAS as a tool for Semi-supervised Learning is the next planned step.

For regressive problems, we are investigating predictive techniques that explore the latent structure in data. PLS and PLSA are two of them, the former being deterministic and the latter probabilistic. These modeling and learning schema are appropriate for the construction of Recommender Systems.

Goals, Activities and Deliverables

G1. Develop *learning algorithms* for classification and regression problems.

Activities:

Develop R-ETL algorithm for regressive and auto-regressive problems

Develop a semi-supervised learning scheme based on the BAS boosting algorithm

Develop a predictive model with latent variables for predicting high-frequency series

Deliverables:

- A regressive algorithm based on ETL, a semi-supervised learning algorithm based on BAS, a predictive model for high-frequency series based on latent variables
- Submission to international and national journals and conferences describing algorithms and their evaluations
- Master dissertations and doctoral theses in the related themes

G2. Develop *machine learning frameworks* with EPM implementation for large volumes of data

Activities:

- Develop a "Hadoop-ized" version of the ETL algorithm
- Develop a "Hadoop-ized" version of the PLS algorithm
- Develop a framework for extracting semantic content from textual documents
- Develop a framework for high-frequency series forecasting
- Develop a framework for building Recommender Systems
- Develop applications for semantic extraction based on the elaborated frameworks

Deliverables:

- Machine learning algorithms with proven performance in tasks of semantic content extraction
- Examples of semantic relations Extractors in textual documents
- Examples of Recommender Systems based in collaborative filtering
- Submission to international and national journals and conferences describing the algorithms, the frameworks and the example applications developed
- Master dissertations and doctoral theses investigating and describing the algorithms, the frameworks, and the example applications developed

TimeFrame

		Y1		Y2		Y3	Y4	Y5
Goal	Activity	S1	S2	S3	S4			
Develop *learning algorithms* for classification and regression problems	Develop R-ETL	X	X					
	Develop a semi-supervised learning scheme based on the BAS boosting algorithm	X	X					
	Develop a predictive model with latent variables for high-frequency series forecasting	X	X					
	Models validation, refinement and testing			X	X			
Develop *machine learning frameworks* with EPM implementation for large	Develop a "Hadoop-ized" version of the ETL algorithm		X	X	X			
	Develop a "Hadoop-ized" version of the PLS algorithm	X	X					
	Develop a framework for extracting semantic					X	X	

volumes of data	content from textual documents							
	Develop a framework for predicting high-frequency series					X	X	
	Develop a framework for building Recommender Systems					X	X	
	Develop applications for semantic extraction based on the elaborated frameworks						X	X

Team

Name	Affiliation	Degree	CNPq level
Ruy Luiz Milidiú	PUC-Rio	PhD, UC Berkeley, 1985	1C
Geraldo Bonorino Xexéo	PESC/COPPE	D.Sc., COPPE, 1994	
Luis Alfredo Vidal de Carvalho	PESC/COPPE	D.Sc., COPPE, 1989	1C
Josefino Cabral Melo Lima	DCC-IM/UFRJ	Docteur, Paris VI, 1992	-

Recent publications

MILIDIU, R.L., DOS SANTOS, C.N., DUARTE, J.C. (2008) Phrase Chunking using Entropy Guided Transformation Learning. In Proceedings of ACL-08: HLT – The 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Columbus, OH, June 15-20, 2008.

DOS SANTOS, C.N., MILIDIU, R.L., RENTERIA, R. P. (2008) Portuguese Part-of-Speech Tagging using Entropy Guided Transformation Learning. PROPOR 2008 – International Conference on Computational Processing of Portuguese, Aveiro, Portugal, September 8-10, 2008.

MILIDIU, R.L., DUARTE, J.C., DOS SANTOS, C.N. (2007) Evolutionary TBL Template Generation. In Journal of the Brazilian Computer Society , 2007.

MILIDIU, R.L., DUARTE, J.C., DOS SANTOS, C.N. (2007) TBL Template Selection: An Evolutionary Approach. In Proceedings of CAEPIA -12th Conference of the Spanish Association for Artificial Intelligence, 2007.

MILIDIU, R.L., DUARTE, J.C., CAVALCANTE, R.P. (2007) Machine Learning Algorithms for Portuguese Named Entity Recognition. In Inteligencia Artificial, Revista Iberoamericana de IA, v 11, n 36, pp 67-75. Asociacion Espanhola para la Inteligencia Artificial, Valencia, 2007.

Relevant theses and dissertations

Ongoing

Cícero Nogueira dos Santos, “ETL: Entropy Guided Transformation Learning”. Ph. D. Candidate. Expected completion: April, 2009.

Júlio César Duarte, “Semi-supervised Learning Through Boosting At Start”, Ph. D. Candidate. Expected completion: April, 2010.

Eraldo Luís Rezende Fernandes, “Semantic Role Labeling through ETL Ensemble Models”. Status: first semester Ph.D. student.

Leandro Alvim, “Machine Learning Forecasting of High Frequency Time Series”. Status: second semester Ph.D. student.

Carlos Crestana, “Semantic Role Labeling through PEM algorithms”. Expected completion: July, 2009.

3.4. Web Infrastructure

3.4.1. Future Internet Architectures

Goals, Activities, and Deliverables

The Internet has been a great success in changing the way we conduct our lives, both at work and play. Since the formulation of the current IP technology almost 30 years ago, incremental changes have been made to extend its applicability to the demands which result from an ever increasing population of users, information resources and applications. The user population is now numbered in billions, and the major contribution to traffic comes from Web applications of increasing diversity.

Much of this success is due to the great flexibility of IP technology, which provides a uniform end to end transport mechanism, independent of the intimate details of lower-level transport mechanisms. Thus a large and increasing array of transport technologies has been successively employed to extend both the geographical coverage and the available capacity of Internet connectivity. In recent years these have included both fixed fiber optic links, which have permitted the vertiginous growth in network capacity in order to permit the scalability of the Internet in accordance with increased demands, and, increasingly, wireless links, which free Internet users from dependence on fixed points of access.

In spite of this success, the basic IP technology contains the seeds of its own limitations, which are becoming more evident every day. The idea that “one size fits all” which requires identical treatment for all information flows within the Internet at the IP packet level is neither desirable, nor particularly cost-effective, especially when certain classes of application, such as those involving interactive media, or remote access to scientific instruments require quality of service (QoS) guarantees which are unnecessary for most other applications. New approaches have recently begun to be used in so-called hybrid networks, where certain classes of large bandwidth applications use dedicated end-to-end virtual circuits, totally avoiding the IP routers responsible for intermediate handling of IP packets. Serious limitations have been recognized in both the scheme used for identifying endpoints of Internet connectivity, with current IPv4 addressing having long been known to be inadequate to scale to the present size of the Internet without breaking fundamental assumptions, such as universally unique identifiers, and the Domain Name System (DNS) used to translate

user-friendly names to endpoint identifiers. The increasing number of mobile users which challenge not only data presentation with their limited platforms, but also require a better balance between transparency (given by Mobile IP) and mobility awareness (required for better performance of applications). Last, but by no means least, are the problems of security, which are hydra-like in their manifestations, providing the opportunity for all manner of insidious activities, ranging from such nuisances as undesired e-mail ("spam") to full-blown information crimes involving identity theft and worse.

It is commonly believed that the solution to many, if not most, of these problems will involve a fundamental redesign of the current Internet architecture, based on IP, and the major objective of this research topic is the formulation and evaluation of alternative architectures to replace it.

This clearly relates to Web Science as it deals with alterations to the supporting network infrastructure on which the Web depends, and has the potential to alter many of the current characteristics of current usage of the Web.

The study of alternative architectures for the future Internet is important for several reasons, not the least of which is the common belief, mentioned above, that this is a truly effective way to resolve many of the outstanding problems which currently afflict Web use. Another important reason is the potential benefit to be obtained by finding a general solution to such problems. Some of the disadvantages of continued persistence in the use of the current architecture include:

- imminent exhaustion of the currently available space of IPv4 endpoint identifiers, causing a "balkanization" of the Internet, without widely available true global connectivity;
- increased costs of IP routers, due to the non-scalable nature of the internal routing tables, and of the performance requirements to process IP packet headers at line speed on very high-speed links, thus restraining network growth;
- immense investments in palliative measures to counter such security problems as are currently caused by spam, denial of service and outright information crimes;.
- difficulty of combining access transparency and application performance for mobile users.

The adoption of an alternative architecture can alter this situation.

It is highly relevant to note that the pursuit of alternative solutions has already begun in parts of the developed world. In the USA, the Global Environment for Network Innovations (GENI) program was launched by the National Science Foundation in 2005, and has already carried out much of the necessary groundwork for carrying out research into alternative network architectures (see http://www.geni.net/). This year, the European Commission launched its Future Internet Research & Experimentation (FIRE) initiative, which has already selected 14 research proposals for immediate funding (see http://cordis.europa.eu/fp7/ict/fire/).

One of the great difficulties of carrying out experimental research into networking architectures is the need for access to real communications networks, in

order to modify the way they work. Naturally, the current operators of such networks will have great objections to permitting experimentation of this kind on their production networks. An ingenious solution to this problem was proposed a few years ago by Larry Peterson, who created the PlanetLab facility, which allowed large-scale experimentation of new applications on the current Internet by means of a worldwide laboratory implemented as an overlay network on top of the Internet. The overlay network, with its own architecture, routing and applications, was implemented by a large number of PlanetLab nodes, implemented on cheap PCs (see http://www.planet-lab.org/). More recently, the PlanetLab approach has been adapted to studying network architecture in the VINI project (see http://www.vini-veritas.net/), which is one of the approaches used in GENI.

Both GENI and FIRE recognize the great importance of testbed networks for R&D in Internet architectures. A testbed is a platform for experimentation for large development projects. Testbeds allow for rigorous, transparent and replicable testing of scientific theories, computational tools, and other new technologies. In GENI, much of the technical discussion has been related to the construction of the GENI Facility, a national testbed network, which will be used for experimentation with proposed architectures (see http://www.geni.net/GDD/GDD-07-44.pdf). In FIRE, 11 of the 14 selected projects are testbed projects (see http://cordis.europa.eu/fp7/ict/fire/test-beds-projects_en.html).

In order to be able to carry out research into Internet architectures, this proposal includes the setting up of a testbed network for design experimentation. The testbed network uses the RNP network infrastructure to interconnect all 8 of the proposed partner institutions, but provides a virtual private network (VPN) interconnecting programmable backbone nodes, at 5 of RNP's points of presence (PoPs), and programmable edge nodes at the collaborating laboratories. The edge nodes will be connected to terminal equipment using the local networks of these laboratories. The edge nodes may use both fixed and wireless access technologies to permit the study of both forms of access to Web applications.

The testbed infrastructure will be based on commodity off the shelf (COTS) equipment, using regular PCs, incremented with programmable wireless boards in the case of wireless edge nodes. Such a configuration has long been the basis for the PlanetLab infrastructure and, more recently, the VINI infrastructure. Additionally, the entire testbed network will require centralized management facilities, for configuration purposes. It will also be instrumented in order to evaluate communication performance, using the approach adopted in the international perfSonar project (http://www.perfsonar.org) in which RNP is a major stakeholder, and UNIFACS is a technical participant.

It is worthwhile noting that RNP, together with CPqD, has operated an optical network testbed in Brazil since 2004, as an integral part of Project GIGA, as well as participating in the PlanetLab initiative since 2005. RNP will continue to make such networking facilities available for use by appropriate research projects, such as this one.

G1. Design and deploy a testbed for R&D in network architectures

Activities:

Evaluation of alternatives for the network architecture testbed;

Acquisition and deployment of the testbed;

Performance testing and optimization;

Evaluation and upgrade of the testbed infrastructure;

Elaboration of scientific papers.

Deliverables:

Platform for network architecture research

2 conference papers

2 international journal papers

G2. Explore new design approaches aiming at Web Applications

Activities:

Elaborate a survey of new design approaches aiming at Web Applications

Evaluate the network requirements of the major Web Applications identified by the other research activities

Elaborate an ontology for the identified Web applications and their requirements

Select the most promising design approaches for experimentation

Implementation and evaluation of the selected design approaches

Analyze the achieved results

Elaborate scientific papers

Elaborate a proposal for an operational service using the most promising approaches

Deliverables:

Proofs of concept that implement new design approaches,

1 national conference paper,

1 international conference paper,

1 international journal paper

G3. Design and implementation of a measurement infrastructure for the testbed

Activities:

Evaluate the available techniques and tools for implementing a measurement infrastructure for the testbed

Customization of the selected environment to be deployed in the testbed

Deployment of the selected environment in the testbed

Tuning and adaptation as necessary to support the pilot projects

Evaluation of the use of the measurement infrastructure

Elaboration of a proposal for an infrastructure to support new Web applications in an operational environment

Elaborate scientific papers

Deliverables:

1 conference paper,

1 international journal paper

G4. Design, deployment and use of a wireless testbed for research on the impact of mobility on web applications

Activities

Evaluate platforms that support virtualization of the wireless interface for implementing a testbed and if there is need of further development

Development of a customized platform

Deployment of the platform in the testbed

Performance testing

Planning and experiments to measure impact of mobility on web applications

Performing experiments

Elaborate scientific papers

Deliverables:

Platform for wireless experiments

1 conference paper

1 international journal paper

Timeframe

		Y1		Y2		Y3	Y4	Y5
Goal	Activity	S1	S2	S3	S4			
Design and deploy a testbed for R&D in network architectures	Evaluation of alternatives for the network architecture testbed	X						
	Acquisition and deployment of the testbed	X	X					
	Performance testing and optimization		X					
	Evaluation and upgrade of the testbed infrastructure					X	X	X
	Elaboration of scientific papers		X	X		X	X	X
Explore new design approaches aiming at Web Applications	Elaborate a survey of new design approaches aiming at Web Applications	X						
	Evaluate the network requirements of the major Web Applications identified by the other research activities	X	X					
	Elaborate an ontology for the identified Web applications and their requirements		X	X				

	Select the most promising design approaches for experimentation			X	X			
	Implementation and evaluation of the selected design approaches					X	X	
	Analyze the achieved results						X	X
	Elaborate scientific papers		X	X	X	X	X	X
	Elaborate a proposal for an operational service using the most promising approaches							X
Design and implementation of a measurement infrastructure for the testbed	Evaluate the available techniques and tools for implementing a measurement infrastructure for the testbed	X						
	Customization of the selected environment to be deployed in the testbed		X	X				
	Deployment of the selected environment in the testbed			X	X			
	Tuning and adaptation as necessary to support the pilot projects					X	X	
	Evaluation of the use of the measurement infrastructure						X	X
	Elaboration of a proposal for an infrastructure to support new Web applications in an operational environment							X
	Elaborate scientific papers				X	X	X	X
Design, deployment and use of a wireless testbed for research on the impact of mobility on web applications	Evaluate platforms for implementing a testbed	X						
	Development of a customized platform	X	X					
	Deployment of the platform in the testbed		X	X				
	Performance testing			X				
	Planning and experiments to measure impact of mobility on web applications			X	X	X	X	
	Performing experiments						X	X
	Elaborate scientific papers		X	X	X	X	X	X

Team

Name	**Affiliation**	**Degree**	**CNPq level**
Antônio Jorge Gomes Abelém	UFPA	Dr, PUC-Rio, 2003	
Iara Machado	RNP	M., UFF, 1998	
José Augusto Suruagy Monteiro	UNIFACS	PhD, UCLA, 1990	
Kelvin Lopes Dias	UFPA	Dr., UFPE, 2001	
Luíz Claúdio Schara Magalhães	UFF	PhD, UIUC, 2002	

Michael Anthony Stanton	RNP	PhD, U. Cambridge, 1971	
Tereza Cristina Melo de Brito Carvalho	USP	Dra, USP, 1996	
William Ferreira Giozza	UNIFACS	Dr. Eng., Paris 6, 1982	

Recent publications

ABELEM, A. ; Albuquerque, Célio V. N. ; Saade, Débora Christina Muchaluat ; AGUIAR, E. S. ; DUARTE, J. L. ; FONSECA, J. E. M. ; MAGALHAES, LUIZ . Redes Mesh: Mobilidade, Qualidade de Serviço e Comunicação em Grupo. Livro de Minicursos do Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos. : , 2007, v. , p. -.

ALBUQUERQUE, Célio V. N. ; SAADE, Débora Christina Muchaluat ; TAROUCO, L. ; MAGALHAES, LUIZ ; GUERRANTE, A. ; Carrano, Ricardo . Desvendando o Padrão IEEE 802.11s. Livro de Minicursos do Simpósio Brasileiro de Redes de Computadores. Rio de Janeiro: , 2008, v. , p. -.

ASSIS, K.D.R; GIOZZA, W.F. and WALDMAN, H. - *WDM Optical Networks: A Complete Design*. IEEE ComSoc/SBrT Joint Issue of the Journal of the Brazilian Telecommunication Society, dezembro, 2005.

CARRANO, Ricardo ; SAADE, Débora Christina Muchaluat ; CAMPISTA, ; MORAES, I.M. ; ALBUQUERQUE, Célio V. N. ; MAGALHAES, LUIZ ; RUBINSTEIN,M.G. ; COSTA,L.H.M.K. ; DUARTE,O.C.M.B. . Multihop MAC: IEEE 802.11s WIreless Mesh Networks. In: Dharma Agrawal. (Org.). Encyclopedia on Ad Hoc and Ubiquitous Computing. : , 2008, v. , p. -.

FIGUEIREDO, Gustavo B; FONSECA, Nelson L. S; MONTEIRO, José Augusto S. A Minimum Interference Routing Algorithm with Reduced Computational Complexity. Computer Networks, v. 50, n. 11, p. 1710-1732. 2006.

KOGA, Ivo K., SAMPAIO, Leobino, MONTEIRO, José A S., FLAVOR: A dynamic and open framework for the development of network measurement access and visualization tools. In: Anais do 25º Simpósio Brasileiro de Redes de Computadores, 2007. Belém, PA. v. 2, p. 665-678.

MARANHÃO NETO, J. C.; SOARES, A. C. B.; GIOZZA, W. F. e CUNHA, P. R. F. – *Wavelength Conversion Architectures in OBS Networks*. IEEE/IFIP Network Operations and Management Symposium (NOMS 2008), April 7-11, Salvador, BA, Brasil, 2008.

MORAES, Priscilla S., SAMPAIO, Leobino N., MONTEIRO, José A S., PORTNOI, Marcos, MonONTO - A domain ontology for network monitoring and recommendation for advanced Internet applications users. In: 6th IEEE Workshop on End-to-End Monitoring Techniques and Services (E2EMon 2008), 2008. Salvador, BA.

SAMPAIO, Leobino, KOGA, Ivo K., SOUZA, Herbert, KOGA, Ivan K., RHODEN, Guilherme E., VETTER, Fausto, LEIRIA, Gerson, MONTEIRO, José A. S., MELO, Edison. piPEs-BR: Uma arquitetura para a medição de desempenho em

redes IP. In: Anais do 24º Simpósio Brasileiro de Redes de Computadores, 2006. Curitiba, PR. v. 1, p. 523-538.

SAMPAIO, Leobino, KOGA, Ivo, COSTA, Rafael, MONTEIRO, Herbert, MONTEIRO, José A Suruagy, VETTER, Fausto, FERNANDES, Guilherme, VETTER, Murilo, Implementing and Deploying Network Monitoring Service Oriented Architectures - Brazilian National Education and Research Network Measurement Experiments (Invited Paper). In: 5th Latin American Network Operations and Management Symposium (LANOMS), 2007. Petrópolis, RJ, p. 28-37.

SOARES, André Castelo Branco, MARANHÃO NETO, José Carneiro, GIOZZA, W. F., CUNHA, Paulo Roberto Freire - *First Load Priority: A Wavelength Converter Placement Scheme for Optical Networks*. Revista Brasileira de Redes de Computadores e Sistemas Distribuídos RB-RESD. , v.1, p.1 - 13, jul. 2008.

SOARES, A.C.B.; GIOZZA, W. F. e CUNHA, P. R. F. – *Classification Strategy to Mitigate Unfairness in All-Optical Networks*. 15th IEEE International Conference on Networks (ICON2007), November 19-21, Adelaide, Austrália, 2007.

SOARES, A. C. B.; GIOZZA, W. F. e CUNHA, P. R. F. – *A New Approach for Performance Evaluation of Survivability Services in All-Opical Networks*. Proc. of the 14th International Conference on Telecommunications (ICT 2007), May 14-17, Penang, Malaysia, 2007.

SOUZA, Herbert M., COSTA, Rafael, SURUAGY, José A., SAMPAIO, Leobino, CUNHA, Paulo F., Uma infra-estrutura para publicação e descoberta de serviços de monitoramento de rede. In: Anais do 25º Simpósio Brasileiro de Redes de Computadores, 2007. Belém, PA. v. 2, p. 679-692.

Relevant theses and dissertations

Completed

Ivo Kenji Koga. Um Arcabouço para aplicações de acesso a serviços de monitoramento multidomínio. 2007. Dissertação (Mestrado em Sistemas e Computação) - Universidade Salvador, Fundação de Amparo à Pesquisa do Estado da Bahia. Orientador: José Augusto Suruagy Monteiro.

Herbert Monteiro Souza. Um Serviço de Publicação e Descoberta de Serviços para Infra-Estrutura de Monitoramento de Rede. 2007. Dissertação (Mestrado em Sistemas e Computação) - Universidade Salvador. Orientador: José Augusto Suruagy Monteiro.

Priscilla Santos Moraes. MonONTO - Uma proposta de ontologia de monitoramento de redes e de aplicações avançadas da Internet. 2007. Dissertação (Mestrado em Sistemas e Computação) - Universidade Salvador. Orientador: José Augusto Suruagy Monteiro.

Mércia Eliane Bittencourt Figueredo. Estudo e Implantação de Qualidade de Serviço na Rede Nacional de Ensino e Pesquisa. 2006. Dissertação (Mestrado em Sistemas e Computação) - Universidade Salvador. Orientador: José Augusto Suruagy Monteiro.

Douglas Vidal Teixeira. Aperfeiçoando a Operação de Redes em Malha. 2004. Dissertação (Mestrado em Engenharia de Telecomunicacoes) - Universidade Federal Fluminense, . Orientador: Luiz Claudio Schara Magalhaes.

Arthur Cabral Fernandes Guerrante Gomes. Medições em Redes IEEE 802.11. 2007. Trabalho de Conclusão de Curso. (Graduação em Engenharia de Telecomunicacoes) - Universidade Federal Fluminense. Orientador: Luiz Claudio Schara Magalhaes.

Marcelo Machado de Pinheiro. Qualidade de serviço em redes sem fio: estudo do impacto da distância das estações ao ponto de acesso em rede infra-estruturada IEEE 802.11e. Dissertação (Sistemas e Computação) - Universidade Salvador. dez., 2007.

Ongoing

Leobino Nascimento Sampaio. Ambiente inteligente de serviços de monitoramento de redes IP. Tese (Doutorado em Ciência da Computação) - Universidade Federal de Pernambuco. Co-orientador: José Augusto Suruagy Monteiro.

Rafael Lima Costa. Compondo serviços web de acordo com o padrão Business Process Execution Language for Web. Dissertação (Mestrado em Sistemas e Computação) - Universidade Salvador. Orientador: José Augusto Suruagy Monteiro.

Luiz Gonzaga de Almeida Neto. Detecção de Intrusões. Dissertação (Mestrado em Sistemas e Computação) - Universidade Salvador. Orientador: José Augusto Suruagy Monteiro.

Arthur Cabral Fernandes Guerrante Gomes. Otimização de Redes em Malha IEEE802.11. Início: 2008. Dissertação (Mestrado em Engenharia de Telecomunicacoes) - Universidade Federal Fluminense, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. (Orientador).

André Castelo Branco Soares. Uma metodologia integrada para provisão dinâmica de circuitos ópticos com garantia de grau e qualidade de serviço. Tese (Ciência da Computação) - Universidade Federal de Pernambuco (defesa marcada para dezembro de 2008).

3.4.2. Ubiquitous and Location-aware Web

The Web is being increasingly used for querying, searching and sharing location-specific information produced by stationary, as well as mobile users. Nowadays, most such uses require a map-based Web application such as Google-maps, in which users can post photographs or textual descriptions associated with a particular geographic spot. However, so far, the user must be geo-navigating on the map to create and upload her contribution. This may be particularly fiddly for mobile users, since portable devices have smaller screens and less comfortable means of user input. Instead, it should be possible to automatically upload any kind of user-generated content (e.g. a voice recording, a video-clip, or a picture taken with the user's device) and post it to the current position of the user without the need of using a map application. Moreover, the user should be able to tag the geographic position (and its contributions) with semantic information, such as her current impressions, emotions, plans, etc. This Ubiquitous location-aware Web (ULW) would enable new forms of sharing, communication and collaboration.

The use of the Web for sharing location-related information, especially by mobile users, opens a wide range of possibilities for mobile social networking, location-aware searching, resource sharing, workforce coordination and entertainment. In fact, we envision huge potentials of the Ubiquitous and Location-aware Web in several application areas with economic and social impact such as: public security, transportation, logistics, tourism, health care, m-learning, entertainment, e-government, and others. However, in order to explore these new opportunities for user communication and collaboration, new services, middleware infrastructures, mobile applications and possibly even new mobile devices have to be developed. Hence, the goal of this research line is to develop and experiment with innovative middleware services and applications for the Ubiquitous and location-aware Web.

Goals, Activities, and Deliverables

G1. Develop prototype location-aware Web services

Activities:

- Survey of State-of-the-Art commercial and academic location-aware Web applications
- Identify the potential of new forms of communication and sharing
- Definition of main requirements
- Implementation of some prototypes
- Field Experimentation with users
- Evaluation of the prototypes
- Elaboration of scientific papers

Deliverables:

- Freely available ULW prototypes with documentation
- Technical Reports of the prototype services (design decisions) and on the user studies
- Submission of scientific papers (various conference and journal papers)

G2. Develop frameworks and middleware services that ease the development of such ULW applications

Activities:

- Summarize the experience gained with the ULW prototypes
- Make a survey of existing frameworks and middleware platforms
- Definition of main requirements and design framework and services
- Development of several frameworks and services
- Integration of the tools
- Re-design of the prototypes using the implemented tools
- Validation of the proposed method
- Elaboration of one scientific papers

Deliverables:

Freely available frameworks and middleware services with documentation

Technical reports with partial results

Submission of scientific papers (various conference and journal papers)

Development of graduate-level course material on ULWs

Timeframe

		Y1		Y2		Y3	Y4	Y5
Goal	Activity	S1	S2	S3	S4			
Prototype location-aware Web services	Survey of State-of-the-Art	X	X					
	Potentials of new forms of communication and sharing	X	X					
	Definition of main requirements		X	X				
	Implementation of some prototypes		X	X	X			
	Field Experimentation with users			X	X	X		
	Evaluation of the prototypes			X	X	X		
	Elaboration of scientific papers		X	X	X	X	X	X
Development of frameworks and middleware services	Summarize the experience gained with the ULW prototypes				X			
	Survey of existing frameworks and middleware platforms			X	X			
	Definition of main requirements and design framework and service				X			
	Development of frameworks and services				X	X	X	
	Integration of the tools					X	X	
	Re-design of the prototypes using the tools							X
	Validation of the proposed method							X
	Elaboration of one scientific paper		X	X	X	X	X	X

Team

Name	**Affiliation**	**Degree**	**CNPq level**
Markus Endler	PUC-Rio	Dr. Rer. nat. TU Berlin, 1992	2
Noemi Rodriguez	PUC-Rio	D.Sc. PUC-Rio, 1993	2
Renato Cerqueira	PUC-Rio	D.Sc., PUC-Rio, 2000	—

Recent publications

SEGHROUCHNI, A.E.F., BREITMAN, K.K., ENDLER, M., SABOURET, N., CHARIF, Y., BRIOT J.-P.. Ambient Intelligence Applications: Introducing the Campus Framework *Proc. of the 13th IEEE International Conference on Engineering of Complex*

Computer Systems (ICECCS 2008), ISBN 0-7695-3139-3, pp. 165-174, Belfast, April 2008.

SKYRME, A., RODRIGUEZ, N., IERUSALIMSCHY, R.. Exploring Lua for Concurrent Programming. Journal of Universal Computer Science, 2008 (to appear).

VITERBO, J. MAZUEL, L,.CHARIF, Y. ENDLER, M. SABOURET, N. BREITMAN, K.K., SEGHROUCHNI, A.E.F.,BRIOT J.-P., Ambient Intelligence: Management of Distributed and Heterogeneous Context Knowledge, Chapter in *Context-Aware Computing and Self-Managing Systems, Waltenegus Dargie (Editor)*, CRC/Francis & Taylor, Series: Chapman & Hall/CRC Studies in Informatics Series, ISBN:9781420077711, December 2008 (to appear).

VITERBO, J., ENDLER, M., BRIOT J.-P, Ubiquitous Service Regulation based on Dynamic Rules. *Proc. of the 13th IEEE International Conference on Engineering of Complex Computer Systems (ICECCS 2008)*, ISBN 0-7695-3139-3, pp. 175-182, Belfast, March 2008

Relevant theses and dissertations

Completed

MILANÉS, A.; RODRIGUEZ, N. (Supervisor): Programming Language Support for the Heterogeneous Migration of Computations, Doctoral Thesis, July 2008.

RUBINSZTEJN, H.K.; ENDLER, M. (Supervisor) Support for Context-aware Content Adaptation for Mobile Devices in Publish/Subscribe Systems, Doctoral Thesis, PUC-Rio, September 14th, 2007.

Ongoing

ROCHA, R.C.A.; ENDLER, M (Supervisor): A Middleware for Ubiquitous Context Awareness, Doctoral thesis, December 2009 (to appear)

SILVESTRE, B.; RODRIGUEZ, N. (Supervisor): Hybrid Concurrency Models and their Application, Doctoral Thesis, March 2009 (to appear).

VITERBO, J.; ENDLER, M (Supervisor): A Framework for Cooperative Reasoning for Ubiquitous Envrionments, Doctoral thesis, March 2009 (to appear)

3.5. Foundations of Web Science

3.5.1. Algorithms and Complexity for the Web graph

The Web graph G=(V,E) is the graph where V is the set of the pages and E is the set of hyperlinks of the World-Wide Web. The Web graph has about a billion vertices today, several billion edges, and appears to grow exponentially with time. There are many mathematical, sociological, and commercial reasons for studying the properties and the evolution of this graph.

The members of the Algorithms and Graphs topic are theorists who usually produce new algorithms for combinatorial problems, such as visibility, coloring

problems, graph dynamics, extremal graphs, and graph invariants. They also work with negative results for these problems as NP-complete and Max SNP-hard classifications or non-approximation results. These results are very important because they lead to the algorithmic techniques such as approximation, exact, or randomized algorithms. In spite of being essentially theoretical the research developed by our group usually produces the necessary conditions for the technology of implementation. These implementations lie in important branches of applications such as mobile communication, routing, scheduling problems, connected nets and others.

Our goal in this project consists of reviewing a set of algorithms that operate on the Web graph, addressing key problems from the Web graph, automatic community search, and classification. We intend to study a number of combinatorial measurements and properties of the Web graph, such as chromatic number, chromatic index, general and parameterized connectivity, as well all the evolution of these parameters through the time. Always observing that one of the most important goals for us is to find out when a such parameter is tractable (determined in polynomial time) or probably intractable (not determined in polynomial time).

In this project, we intend to deal with theory and models of graphs describing different aspects of real world problems, which are present in information and communication technology related to the Web graph. For this purpose we concentrate our investigations in three subjects: Path problems in the Web graph, Searching in the Web graph, and Coloring the Web graph.

Path problems in the Web graph

We study the existence and search of paths/trails between two, not necessarily distinct, vertices s and t in an edge-colored graph from an algorithmic perspective. The existence of such paths deals with a crucial need of the Web graph.

We are interested in a shortest path between s and t, because we want to transmit fast a piece of information from s to t. We are also interested to route a sign to a big number of vertices of the Web graph with low cost, therefore we want to find a long path containing as many vertices as possible, or optimistically a longest path between s and t.

Several versions of the s-t path/trail problem admit polynomial solutions including the shortest path/trail case. We give polynomial algorithms for finding a longest color alternating path/trail between s and t for some particular graphs and characterize edge-colored graphs without closed trails. We prove that deciding whether there exist k pairwise vertex/edge disjoint s-t paths/trails in a c-edge-colored graph Gc is a NP-complete problem even for k=2 and c lower bounded by n square, where n denotes the number of vertices in Gc. We have studied the Hamiltonian cycle and path problems in special classes of graphs, in particular for some n-cube subgraphs. Among the Web graph problems we are interested in, we can mention more specifically:

to determine long paths in the Web graph,

to determine shortest paths between two given vertices in the Web graph.

Searching in the Web graph

Methods for ranking World Wide Web resources according to their position in the link structure of the Web are nowadays receiving considerable attention, because they provide the first effective means for search engines to cope with the explosive growth and diversification of the Web graph. Visualization methods supporting the simultaneous exploration of a link structure and a ranking of the Web graph vertices have been developed. This has been achieved applying the results of the ranking algorithm and employing graph drawing techniques. These techniques are useful for the analysis of query results, maintenance of search engines, and evaluation of the Web graph models.

The last couple of decades have witnessed a huge growth in the World Wide Web. The Web has become a channel for data information sharing and dissemination. One of the basic questions about the Web is its size, how many vertices are on the Web graph. In the last ten years both the scientific literature and the popular press dealt with methodologies and estimates for the sizes of the Web and public search engines. The Web graph is so large that even estimating its size is a hard problem. To better understand the Web graph, people study random Web graph models -- a stochastic process that generates random graphs that with high probability resemble the Web graph. Random Web graph models are useful theoretical tools to provide insights to the evolution of the Web graph, to predict future behaviors, and to generate synthetic data sets for testing and simulation, briefly present several results in analyzing Web graph models and applying the models to real world applications.

The basic structure used to modeling the Web graph is the model of Random Graph, due to Paul Erdös and Alfréd Rényi, in 1959. The Random Graph G(n,p) represents the graph with n vertices with each edge generated with probability p. The most notable phenomenon in Erdös's random model is the existence of the phase transitions in evolution of the random graph. By phase transitions we assume the singularity events of certain properties of the graph defined in the edge generation probability increasing process. We can identify the exact transitional probabilities with certain properties of the graph, such as connectivity. Noticing that the Erdös random graph model usually does not explain the Web graph observations, new families of random graph models have been proposed in order to get the real measurements and to perform new algorithms for the modeling of the Web graph. We select the following problems

to uniformly sample Web pages; to model the Web graph;

to find duplicate hosts;

to find top gainers and losers in data streams;

Coloring the Web graph

The World-Wide Web has produced a sharing and dissemination of information on an unprecedented scale. Hundreds of millions (soon to be billions) of individuals are creating, annotating, and exploiting hyperlinked contents in a distributed fashion. These individuals come from a variety of backgrounds and have different motivations. The hyperlinks of the Web give it additional structure; the network of these links is a

rich source of latent information. Recent estimates suggest that there are over a billion vertices in the Web graph; this quantity is growing every month. The average vertex has roughly seven directed edges to other vertices, so that the graph contains several billion vertices in all.

It is known that the Web graph contains many densely connected directed multipartite induced subgraphs because cyber-communities often have such a densely connected behavior. The source vertices in such a subgraph are the "hubs" or directory vertices on the topic; the sink vertices are the "authorities" or content vertices on the topic. Algorithms to find small complete bipartite subgraphs have been widely produced in the literature. However, their color classes are relatively small, in the order of tens of vertices. In order to more completely capture these cyber-communities, it would be interesting to detect much larger bipartite subgraphs, like hundreds or thousands of vertices. They do not need to be complete, but they should be dense, i.e., they should contain at least a constant fraction of the corresponding complete bipartite subgraphs. Two natural additional problems:

to find large directed multipartite induced subgraphs in a Google Earth fashion.

to implement efficiently such algorithm with only a small part of the Web graph.

Goals, Activities, and Deliverables

G1. Define algorithms in graphs, and / or negative results for combinatorial problems.

Activities:

Weekly seminaries at UFRJ, with national and international guests.

Work meetings at universities in Brazil and abroad.

International workshop organization.

Hosting visitors from other universities.

Deliverables:

Publication of papers in periodicals.

Conference papers describing the Web graph algorithms.

M.Sc. dissertations and D.Sc. theses based on the algorithms and graphs.

Timeframe

		Y1	Y2	Y3	Y4	Y5
Goal	Activity					
Define algorithms in graphs, and / or negative results for combinatorial problems.	Weekly seminaries at UFRJ, with national and international guests.	X	X	X	X	X
	Work meetings at universities in Brazil and abroad.	X	X	X	X	X
	International workshop organization.			X		
	Hosting visitors from other universities.	X		X		X

Team

Name	Affiliation	Degree	CNPq Level
Carlos Alberto de Jesus Martinhon	UFF	D. Sc. COPPE Sistemas e Computação UFRJ (1998)	
Carlos Eduardo Fisch de Brito	UFC	Ph. D. Computer Science pela University of California - Los Angeles (2003)	
Célia Picinin de Mello	UNICAMP	Ds.C. UFRJ, 1982	1C
Claudia Linhares Sales	UFC	Docteur Informatique - Recherche Operationnelle - Université de Grenoble I (Scientifique Et Medicale - Joseph Fourier) (1996).	2
Claudio Leonardo Lucchesi	UNICAMP	Ph.D. Computer Science - University of Waterloo (1976)	1A
Claudson Ferreira Borstein	UFRJ	Ph. D. Computer Science pela Carnegie Mellon University (1998)	2
Edson Norberto Cáceres	UFMS	D. Sc. COPPE Sistemas e Computação UFRJ (1992)	2
Fábio Protti	UFRJ	D. Sc. COPPE Sistemas e Computação UFRJ (1998)	1D
Francisco Heron de Carvalho Junior	UFC	D. Sc. Ciências da Computação pela Universidade Federal de Pernambuco (2003	
Jayme Luiz Szwarcfiter	UFRJ	Ph. D. Ciência da Computação, pela University of Newcastle Upon Tyne, Inglaterra (1975)	1A
Loana Tito Nogueira	UFF	D. Sc. COPPE Sistemas e Computação UFRJ (2003)	2
Luerbio Faria	UERJ	D. Sc. COPPE Sistemas e Computação UFRJ	2

		(1998)	
Luis Antonio Brasil Kowada	UFF	D. Sc. COPPE Sistemas e Computação UFRJ (2006)	
Manoel Bezerra Campêlo Neto	UFC	D. Sc. COPPE Sistemas e Computação UFRJ (1999)	2
Márcia Rosana Cerioli	UFRJ	D. Sc. COPPE Sistemas e Computação UFRJ (1999)	2
Mitre Costa Dourado	UFRJ	D. Sc. COPPE Sistemas e Computação UFRJ (2005)	2
Noemi Costa dos Santos	UERJ	D. Sc. COPPE Sistemas e Computação UFRJ (2006)	
Paulo Eustáquio Duarte Pinto	UERJ	D. Sc. COPPE Sistemas e Computação UFRJ (2006)	
Rafael Castro de Andrade	UFC	Docteur Université de Paris XIII (Paris-Nord) (2002)	2
Ricardo Cordeiro Corrêa	UFC	Docteur Informática - Institut National Polytechnique de Grenoble (1997)	2
Simone Dantas de Souza	UFF	D. Sc. COPPE Sistemas e Computação UFRJ (2002)	2
Sulamita Klein	UFRJ	D. Sc. COPPE Sistemas e Computação UFRJ (1994)	1C

Relevant theses and dissertations

Completed

Pinto, P. E. D. Árvores Pares: Um Esquema para Detecção de Erros em Árvores Tipo Huffman. PhD Thesis, COPPE-Sistemas UFRJ (2006). (Orientador: Jayme Luiz Szwarcfiter)

dos Santos, N. C. Reprentações para Pares Modulares de um Grafo. PhD Thesis, COPPE-Sistemas UFRJ (2006). (Orientador: Sulamita Klein e Jayme Luiz Szwarcfiter)

Maria Patricia Dobson. Sobre Grafos Cubridores de los Grafos de Comparabilidad. PhD Thesis, Universidad Nacional de La Plata, (2006). (Orientador: Jayme Luiz Szwarcfiter)

Marina Esther Groshaus. Bicliques, Cliques, Neighborhoods y la Propiedad de Helly. PhD Thesis, Universidad de Buenos Aires, (2006). (Orientador: Jayme Luiz Szwarcfiter)

Ongoing

Oliveira, F. O. Intersection Graphs PhD Thesis, COPPE-Sistemas UFRJ. (Orientador: Márcia Rosana Cerioli e Jayme Luiz Szwarcfiter)

Artigas, D. Convexity in Graphs. PhD Thesis, COPPE-Sistemas UFRJ. (Orientador: Jayme Luiz Szwarcfiter e Mtre da Costa Dourado)

Rodrigues, R. F. Scheduling of Tasks PhD Thesis, COPPE-Sistemas UFRJ. (Orientador: Jayme Luiz Szwarcfiter)

Figueiredo, A. Dimension in Partially Ordered Sets, PhD Thesis, COPPE-Sistemas UFRJ. (Orientador: Sulamita Klein e Jayme Luiz Szwarcfiter)

Miranda, A. A. A. Pfaffian Graphs. PhD Thesis, UNICAMP. (Orientador: Cláudio Leonardo Lucchesi)

da Silva, C. N. Integer Flow in Graphs. PhD Thesis, UNICAMP. (Orientador: Cláudio Leonardo Lucchesi)

Bravo, R. S. F. Dynamic Algorithms for the Graph Partition Problem, PhD Thesis, COPPE-Sistemas UFRJ. (Orientador: Sulamita Klein e Loana Tito Nogueira)

Coelho, H. V. Oriented Coloring, PhD Thesis, COPPE-Sistemas UFRJ. (Orientador: Sulamita Klein e Luerbio Faria)

Bueno, L. R. On the Lovász conjecture for transitive graphs, PhD Thesis, COPPE-Sistemas UFRJ. (Orientador: Celina Figueiredo e Luerbio Faria)

Nepomuceno, N. V. Wireless Dynamic Routing and Optimization, PhD Thesis, Instituto de Computação, Universidade Federal do Ceará. (Manoel Bezerra Campêlo Neto)

Rodrigues, C. D. Graph Cooring Problems, PhD Thesis, Instituto de Computação, Universidade Federal do Ceará. (Manoel Bezerra Campêlo Neto)

Martins, E. M. Coloring Helly Property, PhD Thesis, COPPE-Sistemas UFRJ. (Orientador: Jayme Luiz Szwarcfiter e Mitre da Costa Dourado)

de Araújo, G. A. Distributed and Parallel Plataforms and Scientific Applications Components Based, PhD Thesis, Instituto de Computação, Universidade Federal do Ceará. (Orientador: Ricardo Cordeiro Corrêa)

Pará, T. S. Graph Theory and Mathematical Education, PhD Thesis, COPPE-Sistemas UFRJ. (Orientador: Luis Alfredo Vidal de Carvalho e Simone Dantas de Souza)

3.5.2. Applied and Computational Mathematics/Optimization

Introduction

Optimization, also named mathematical programming, is part of the computational mathematics. It constitutes a scientific area that is essential to solving organization, planning and scheduling of activities, since it seeks to optimize an objective (goal) function subject or not to a set of constraints. These decision problems are studied by mathematical programming models. The optimum can be a maximum or minimum, depending on the goal in question. The attainment of a global optimum is not always possible; in these cases the search for effective equilibria solutions for a given system may present values for the goal function between lower and upper bounds (primal methods, heuristics, metaheuristics and dual methods). Undoubtedly, the main resolution technique in mathematical programming is linear optimization (also called, linear programming) [1].

With the advent of the Internet, a new area emerged within theoretical computer science[5]. This new field, algorithmic foundations of the Internet, was conceived to tackle the problems arisen from the need to deal with very large networks. It appeared during the first half of the nineties, coinciding with the popularization of the worldwide web. At present, the field consists mainly of topics from graph theory, algorithms, game theory, distributed computing, statistical physics, economics, information retrieval and probability theory, but topics have continuously been aggregated as new aspects of the field come to attention.

Therefore, optimization techniques are indispensable for an effective organization of the Web, thus constituting an essential part of Web Science. In order to contribute to the analysis of the Web and to the synthesis and enhancement of its mechanisms, the optimization methods must improve their algorithms as well as be able to make a quasi-perfect use of the most recent computational technologies: huge networks of heterogeneous computers. Speed enhancement in database information search and set covering problems are formulated as combinatorial optimization problems. Such models are also employed in the routing of information flow in order to minimize answer time for one or multiple simultaneous search operations requested by users. All the strategies pertaining to the area of optimization in graphs are used in the dimensioning of computer networks.

Finally, the conception of a Web that is scalable and effective must be based on the systematic use of global optimization algorithms [2].

Subtopic 1: Optimization Methods

Two types of model can be considered for the description of decision problems: deterministic and stochastic. In the deterministic optimization models, problem data is fixed *a priori*, and finding one optimal solution or a set of optimal solutions to a given goal function constitutes the purpose of the models. On the other hand, stochastic models either have all or part of the data given through probability density functions.

[5] A. Lòpez-Ortiz, Algorithmic Foundations of the internet, *ACM SIGACT News*, June/2005, Vol. 32, No. 2.

Alternatively, for each stochastic model, there is a sample of the data that is treated in a deterministic fashion. As an additional strategy, it is possible to search for confidence intervals associated to probabilities of finding optimal values for the variables of the problem.

The methods to solve those models can also be deterministic or stochastic. Many problems formulated by stochastic models can be solved by techniques based on probability theory or queueing theory, both deterministic methods. On the other hand, the Monte Carlo methods and simulated annealing are stochastic methods that are commonly used in the search for solutions to deterministic problems involving integrals and combinatorial optimization.

In the context of this project, new versions of interior-point algorithms will be developed to provide solutions to problems of linearly constrained convex optimization. The group has experience on the subject, as shown by the reported research results [15] [16] [17].

A Jacobian approximation is used in the solution of a system of equations and inequalities associated to optimality conditions (KKT) of a linear or quadratic convex programming problem. A solution of that approximation makes use of an iteration of either the Newton method or one of its variants, many of which constitute large systems of linear equations. In this context, hundreds or even thousands of problems are solved. These problems are of type $Qx=b$, where Q is a known square matrix, x is the variables vector and b is a given vector. Matrix Q may be dense or sparse. The development of new matrices, named pre-conditioners, will be an important aspect of novel computational implementation proposals for interior-point methods [15] [16]. Problems of computer network dimensioning may be studied via models of stochastic optimization with some integer variables [3] [4]. New algorithms and modeling ideas for these problems will also be considered.

The study of problems through non-linear optimization models manages to encompass several decision-making techniques. The development of algorithms for the solution of differentiable non-linear programming models always leads to the search of a point that satisfies Lagrange-KKT optimality conditions. If the model is not convex (minimization) but differentiable, it is possible to find points that correspond to local minima and not to the *minimum minimorum*. There are global optimization techniques that are suitable in this case [2] [5] [6]. Methods for the solution of non-linear optimization have been published in the last few years by most of the components of the group of researchers associated to this topic (see [17] [18] [19] [20] [21]). By generalizing the optimal complementarity conditions, it is possible to obtain problems modeled by variational inequalities [7] [13] [19]. For these problems, equilibria solutions are the most appropriate.

Many problems convey combinatorial characteristics, for example, when part of their variables pertain to the set {0, 1}, which frequently represents Boolean values: 1 when the variable is true and 0 otherwise. In other cases, the value of a variable may represent another kind of information, for example, the set of discrete values {1, 3, 12, 34, 72} could be associated to the values of pipe diameters that are commercially available. The rounding of continuous solutions can result in a value for the objective

function that is too far from the optimum (see Chapter 10 of [1]). Problems related to the search of connected subgraphs of a graph may be formulated by means of combinatorial optimization models (integer or mixed programming). These models can be used, for example, to obtain spanning trees, elementary paths and Hamiltonean cycles. In network flow problems as well as in network dimensioning, mixed integer programming models are used. Undoubtedly, we can divide the algorithms that tackle combinatorial optimization problems into: exact, metaheuristic and mixed. The exact algorithms usually start from a solution for the relaxed model, which is associated to a linear programming model and develop, from this first relaxation, an implicit enumeration method (branch-and-bound). In order for this to happen, it is important to provide, in the case of minimization, lower bounds (associated to the relaxed model) and upper bounds (associated to the viable solutions to the problem). To obtain lower bounds we can apply cuts, additional constraints aiming to eliminate non-integer Lagrangean relaxations of the original model, and SDP programming (see [8] [9] [10] [11] [12] [22] [23] [24] [36]).

In combinatorial optimization, metaheuristics has developed considerably and, for some problems, it contains the only techniques that yield good results. In other situations, stochastic techniques and local optimization are combined drawing inspiration from nature. Some of these models, for example genetic algorithms [14] and evolutionary computation, act according to the evolution of life, while others base themselves on the collective behaviour of particles swarms or on immune systems (see [14] [26] [27] [28] [29] [30] [31] [32] [35] [41] [42]. The combination of exact methods with metaheuristics can advance the state-of-the-art of large, complex combinatorial problems such as those concerning the Web graph.

Subtopic 2: Intensive Utilization of New Architectures and Computer Networks in Optimization

Nearly all optimization problems, being linear, non-linear, combinatorial or stochastic, make use of linear optimization techniques as sub-routines. In order to ensure that the implementation of new optimization algorithms be efficiently executed, it is necessary that the linear optimization methods employed be able to profit from the use of both new computer architectures and computer networks. The universities that support this project have advanced high performance computing systems, which can be expanded to cope with the computational demand of the work to be developed (see [15] [16] [32] [33] [34] [37] [38] [39] [40] [43] [44] [45] [46]). The parallelization of the optimization mechanisms will be carried out in three granularity levels: threads, tightly coupled processes and loosely coupled processes.

Graphics Processing Units — GPUs — and multicore processors are parallel hardware already available into the newest computer systems of our daily life. Those two alternatives represent two different granularity levels in the parallel computing domain: while GPUs exploit potentially finer grains of parallelism (such as vector processing), multicore processors allow one to accelerate a computation by triggering multiple threads or processes into the multiple CPUs typically arranged as shared memory multiprocessor system. Both models are not yet fully exploited and there are still many opportunities for devising new parallel optimization algorithms. Although parallel computation at this granularity level is geographically local, its study is still

justified in this context by the need of parallel and distributed systems to combine the strengths of each technology, thus, obtaining results that advance the state-of-the-art of the size of solvable instances.

The next level corresponds to the tightly coupled processes, where the need for an intense exchange of messages exists. Such processes can exchange information via shared memory, real or distributed (DSM, *distributed shared memory*), or via message exchanging, explicitly. Such systems can be implemented both in customized multiprocessor systems or computer clusters, i.e., a set of computers interconnected by a dedicated network, typically. At this parallelism granularity, the cost of communication among processes is about one or two orders of magnitude greater than the same type of cost found when GPUs are considered. Such a difference is totally reflected in the way parallel algorithms are conceived at this granularity level.

The interest in obtaining performance gains that allows for tackling large problem instances is a real motivation for the use of parallel and distributed processing environments. Computational grid is the parallel computing model which counts, typically, with the largest number of processors and, at the same time, the highest communication cost among all models considered. As an example of exploitation of this model, stochastic methods that allow for initiating search from any point in the space of states of solutions are considered. A parallelization strategy counting on some or none inter-process communication is a natural step: such a strategy is known as *bag-of-tasks*. The choice of starting points can be obtained through other methods, what does not limit stochastic methods to this simple scheme.

The greatest part of computational effort in interior-point algorithms focuses on the solution of a sequence of symmetric linear systems. The solution to these systems uses direct factoring methods or iterative ones. In an early phase, this project intends to evaluate a parallelized version of the multi-frontal method applied to the solution of systems generated by interior-point methods. In distributed memory multi-computer architectures, the parallelized method divides computation amongst processors via MPI (message-passing interface) in an asynchronous manner. Furthermore, it is advantageous to explore shared memory parallelism, inherent to linear algebra operations, when the nodes of these architectures comprise multicore processors. Additionally, iterative methods can be employed as implementation strategy for interior-point algorithms. In this case, parallelization of direct methods is also applicable, as the computation of pre-conditioners usually involves the factoring of a matrix that derives from the original system. Other possibilities of parallel schemes suitable for shared memory are the vector-matrix multiplications, which constitute a substantial portion of these methods.

Goals, Activities, and Deliverables

G1. Implementation of a continuous optimization library with and without constraints.

Activities:

Survey of existing methods and improvements proposal.

Study of the implementations of hyperbolic penalty methods.

Identification of parallelization opportunities.

Library development and publishing as freeware.

Results:

Tutorial about the methods and implementations studied.

Parallelized and enhanced methods.

Freeware consisting of the library developed, together with documentation.

Papers submitted to journals and conferences describing the new methods and their implementations.

M.Sc. dissertations and D.Sc. thesis investigating the new methods and their implementations.

G2. Development of new methods for implicit enumeration in combinatorial optimization, column generation and cutting planes.

Activities:

Survey of the existing methods and their applicability to problems involving the Web graph.

Study of the methods implementations and their scalability.

Identification of parallelization opportunities.

Experiments with the implementations in connection with the Web graph.

Results:

Tutorial about the methods and implementations studied.

Parallelized and enhanced methods.

Papers submitted to journals and conferences describing the new methods and their implementations.

M.Sc. dissertations and D.Sc. thesis investigating the new methods and their implementations.

G3. Implementation of metaheuristics aiming to search for lower bounds (in maximization) and upper bounds (in minimization).

Activities:

Survey on the state-of-the-art in metaheuristics and their connection to Web mathematics.

Study of the most recent implementations of metaheuristics.

Identification of parallelization opportunities for the enhanced metaheuristics.

Experiments and adjustments of the metaheuristics.

Results:

Tutorial about the metaheuristics and their application to Web mathematics.

Parallelized and enhanced metaheuristics.

Papers submitted to journals and conferences describing the new metaheuristics, their implementations and applications to the Web.

M.Sc. dissertations and D.Sc. thesis investigating the new metaheuristics, their implementations and applications to the Web.

Timeframe

		Y1		Y2		Y3	Y4	Y5
Goal	Activity	S1	S2	S3	S4			
Implementation of a continuous optimization library with and without constraints	Survey of existing methods and improvements proposal.	X	X					
	Study of the implementations of hyperbolic penalties methods.		X	X				
	Identification of parallelization opportunities.			X	X			
	Library development and publishing as freeware					X	X	X
Development of new methods for implicit enumeration in combinatorial optimization, column generation and cutting planes	Survey of the existing methods and their applicability to problems involving the Web graph.	X	X					
	Study of the methods implementations and their scalability.		X	X				
	Identification of parallelization opportunities			X	X			
	Experiments with the implementations in connection with the Web graph and adjustment of the methods.					X	X	X
Implementation of metaheuristics aiming to search for lower bounds (in maximization) and upper bounds (in minimization)	Survey on the state-of-the-art in metaheuristics and their connection to Web mathematics.	X	X					
	Study of the most recent implementations of metaheuristics.		X	X				
	Identification of parallelization opportunities			X	X			
	Experiments and adjustments of the metaheuristics					X	X	X

Team

Name	**Affiliation**	**Degree**	**CNPq Level**
Nelson Maculan	UFRJ	Doutor	1A

Filho			
Celso da Cruz Carneiro Ribeiro	UFF	Doutor	1A
Abílio Pereira de Lucena Filho	UFRJ	Doutor	1D
Antonio Alberto Fernandes de Oliveira	UFRJ	Doutor	1D
Cid Carvalho de Souza	UNICAMP	Doutor	1D
José Ferreira de Rezende	UFRJ	Doutor	1D
Luiz Satoru Ochi	UFF	Doutor	1D
Susana Scheimberg de Makler	UFRJ	Doutor	1D
Marcia Helena Costa Fampa	UFRJ	Doutor	2
Mario Roberto Folhadela Benevides	UFRJ	Doutor	2
Simone de Lima Martins	UFF	Doutor	2
Geraldo Galdino de Paula Júnior	UENF	Doutor	--
Luiz Mariano Paes de Carvalho Filho	UERJ	Doutor	--

Collaborators

Name	**Affiliation**	**Degree**
Adilson Elias Xavier	UFRJ	Doutor
Anne-Marie Delaunay Maculan	UFRJ	Doutor
Cláudio Thomás Bornstein	UFRJ	Doutor
Felipe Maia Galvão França	UFF	Doutor
Fernanda Maria P. Raupp	PUC-RJ	Doutor
Geraldo Gil Veiga	UFRJ	Doutor
Luidi Gelabert Simonetti	UNICAMP	Doutor
Maria Helena Cautiero Horta Jardim	UFRJ	Doutor
Paulo Roberto Oliveira	UFRJ	Doutor
Priscila Machado Vieira Lima	UFRJ	Doutor
Rosa Maria V. de Figueiredo	UERJ	Doutor
Valeria Saldanha Motta	IME	M.Sc., Doutoranda
Yuri Abitbol de Menezes Frota	UNICAMP	Doutor

Recent publications and references

[1] N. Maculan & M.H.C. Fampa, Otimização Linear, editora UnB, ISBN: 85-230-0927-2, 310 p, 2006.

[2] L. Liberti & N. Maculan (eds), Global optimization: from theory to implementation, Springer, ISBN: 0-387-28260-2, 428 p, March 2006.

[3] R. Andrade, A. Lisser, N. Maculan & G. Plateau, B&B frameworks for the capacity expansion of high speed telecommunication networks under uncertainty, Annals of Operations Research, vol. 140 (1), pp. 49-65, (2005).

[4]R. Andrade, A. Lisser, N. Maculan & G. Plateau, Enhancing a branch and bound algorithm for two-stage stochastic integer network design based models, Management Science, Vol. 52 (9), pp. 1450-1455, (2006).

[5] C. Lavor & N. Maculan, Interval analysis applied to global minimization of potential energy function, Advances in Convex Analysis and Global Optimization, N. Hadljisavvas & P. Pardalos (editors), series: Nonconvex Optimization and its Applications, vol. 54, Kluwer Academic Publishers, pp. 333-344, (2001).

[6] C. Lavor, L. Liberti, N. Maculan & M.A.C. Nascimento, Solving Hartree-Fock systems with global optimization methods, Europhysics Letters (EPL), vol. 77 (5), 50006 (p1-p6), (2007).

[7] S. Scheimberg, N. Maculan & M.H.C.H. Jardim, a ser publicado em Acta Mathematica Vietnamica, 2008.

[8] A. Salles da Cunha, A. Lucena, N. Maculan & M.G.C. Resende, A Relax-and-Cut Algorithm for the Prize-Collecting Steiner Problem in Graphs, to be published in Discrete Applied Mathematics.

[9] R. Andrade, A. Lucena & N. Maculan, Using Lagrangian dual information to generate degree constrained minimum spanning trees, Discrete Applied Mathematics, vol. 154, pp. 703-717, (2006).

[10] E.M. Macambira, N. Maculan & Cid C. de Souza, A note on characterizing canonical cuts using geometry, ITOR (International Transaction in Operations Research), vol. 12, pp. 581-593, (2005).

[11] M. Fampa & N. Maculan, Using a conic formulation for finding Steiner minimal trees, Numerical Algorithms, vol. 35, n. 2-4, pp. 315-330, (2004).

[12] C.T. Bornstein, L. Ferreira Alcoforado & N. Maculan, A graph-oriented approach for the minimization of the number of late jobs for the parallel machines scheduling, EJOR (European Journal of Operational Research), vol. 165, pp.649-656, (2005).

[13] N. Maculan, C. Prata Santiago, E.M. Macambira & M.H.C. Jardim, An O(n) algorithm for projecting a vector on the intersection of a hyperplane and a box in Rn, Journal of Optimization Theory and Applications (JOTA), vol. 117, n. 3, pp. 553-574, (2003).

[14] L.S. Ochi, N. Maculan & R.M.V. de Figueiredo, A New Self-Organizing Strategy based on Elastic Networks for solving the Euclidean Traveling Salesman Problem, Lectures Notes in Computer Science, v. 1416 (Editors: A. Pasqua del Pobil, J. Mira & M. Ali), Springer Verlag, pp. 479-487, ISBN: 3-540-64574-8, (1998).

[15] M.G.C. Resende & G.G. Veiga, An implementation of the dual affine scaling algorithm for minimum-cost flow on bipartite uncapacitated networks, SIAM J. Optimization, vol. 3, n. 3, pp. 516-537, (1993).

[16] I. Adler, M.C.G. Resende, G.G. Veiga & N. Karmarkar, An implementation of Karmarkar's algorithm for linear programming, Mathematical Programming, vol. 44, pp. 297-335, (1989).

[17] S. Scheimberg, R. Burachik, Luis Mauricio Graña Drummond . On some properties and an application of the barrier method, Mathematical Programming, vol. 111, pp. 1-22, (2007).

[18] S. Scheimberg, C. Sagastizábal, R. Burachik, An inexact method of partial inverses and a parallel bundle method, Optimization Methods and Software, vol. 21, pp. 385-400, (2006).

[19] S. Scheimberg, P.S. Santos, A Projection Algorithm for General Variational Inequalities. Applied Mathematics and Computation, vol. 181, pp. 649-661, (2006).

[20] S. Scheimberg, M. Campelo, A simplex approach for finding local solutions of a linear bilevel program by equilibrium points, Annals of Operations Research, pp. 01-15, (2005).

[21] S. Scheimberg, M. Campelo, A study of local solutions in linear bilevel programming. Journal of Optimization Theory and Applications, vol. 125, n. 1, pp. 63-84, (2005).

[22] Souza, Cid Carvalho de, Cavalvante, V. F., Lucena, A., A relax-and-cut algorithm to the set partitioning problem, Computers and Operations Research, v. 35, pp. 1963-1981, (2008).

[23] Souza, Cid Carvalho de ; Wosley, L. A. ; Weismantel, R. ; Martin, A. ; Ferreira, C. E., The Node Capacitated Graph Partitioning Problem: A Computational Study, Mathematical Programming, vol. 81, pp. 229-256, (1998).

[24] Souza, M.H.C., Cid Carvalho de; Weismantel,, R. ; Wosley, L. A. ; Martin, A. ; Ferreira, C. E., The Node Capacitated Graph Partitioning Problem: Formulations and Valid Inequalities, Mathematical Programming, vol. 74, pp. 247-266, (1996).

[25] Fampa, M.H.C. ; Barroso, L. A. ; Candal, D. ; Simonetti, L., Bilevel optimization applied to strategic pricing in competitive electricity markets, *Computational Optimization and Applications*, v. 39, pp. 121-142, (2008).

[26] Ochi, L. S.; Santos, H. G.; Souza, M. J. F., A Tabu Search heuristic with efficient diversification strategies for the class/teacher timetabling problem., *ACM Journal of Experimental Algorithmics*, ACM, v. 10, Section 2, art.-2.9, pp.1-15, (2006).

[27] Silva, G. C.; Andrade, M. R. Q.; Ochi, L. S.; Carvalho, A. P.; Martins, S. L., New Heuristics for the Maximum Diversity Problem, *Journal of Heuristics*, v. 13, pp. 315-336, (2007).

[28] Martins, S. L.; Pardalos, P. M.; Resende, M. G. C.; Ribeiro, C. C. C., Greedy randomized adaptive search procedures for the Steiner problem in graphs, *Dimacs Series In Discrete Mathematics And Theoretical Computer Science*, v. 43, pp. 133-146, (1998).

[29] Ribeiro, C. C. C.; Aloise, D.; Noronha, T. F. ; Rocha, C. T. M. ; Urrutia, S., An efficient implementation of a VNS/ILS heuristic for a real-life car sequencing problem, *European Journal of Operational Research*, v. 191, pp. 596-611, (2008).

[30] Ribeiro, C. C. C.; Aloise, D. ; Noronha, T. F. ; Rocha, C. T. M. ; Urrutia, S., A hybrid heuristic for a multi-objective real-life car sequencing problem with painting and assembly line constraints, *European Journal of Operational Research*, v. 191, p. 981-992, (2008).

[31] Ribeiro, C. C. C.; Martins, S. L.; Rosseti, I. C. M., Metaheuristics for optimization problems in computer communications, *Computer Communications*, v. 30, pp. 656-669, (2007).

[32] Ribeiro, C. C. C.; Rosseti, I. C. M., Efficient parallel cooperative implementations of GRASP heuristics, *Parallel Computing*, v. 33, pp. 21-35, (2007).

[33] Carvalho, D. B. F.; Lucena, C. J. P.; Ribeiro, C. C. C.. Uma proposta de framework para desenvolvimento de aplicações paralelas com mobilidade. *Revista de Informática Teórica e Aplicada*, v. XIV, pp. 29-42, (2007).

[34] Coll, P.; Ribeiro, C. C. C.; Souza, C. C., Multiprocessor scheduling under precedence constraints, *Discrete Applied Mathematics*, v. 154, n. 5, pp. 770-801, (2006).

[35] Alvim, A. C. F.; Ribeiro, C. C. C.; Glover, F.; Aloise, D. J., A hybrid improvement heuristic for the bin packing problem, Journal of Heuristics, v. 10, n. 2, pp. 205-229, (2004).

[36] Fampa, M. H. C.; Anstreicher, K. M., An improved algorithm for computing Steiner minimal trees in Euclidean d-space, *Discrete Optimization*, v. 5, pp. 530-540, (2008).

[37] Meurant, G.; Giraud, L.; Carvalho, Luiz M., Local preconditioners for two-level non-overlapping domain decomposition, *Numerical Linear Algebra with Applications*, v. 8, n. 4, pp. 207-227, (2001).

[38] Giraud, L.; Le Tallec, P.; Carvalho, Luiz M., Algebraic Two-Level Preconditioners for the Schur Complement Method, *SIAM Journal on Scientific Computing*, v. 22, n. 6, pp. 1987-2005, (2001).

[39] Carvalho, Luiz M.; Giraud, L., Parallel subdomain-based preconditioner for the Schur complement, *Lecture Notes in Computer Science*, v. 1685, pp. 1032-1039, (1999).

[40] N. Maculan and G.G. de Paula Jr., A Linear-Time Median-Finding for Projecting a Vector on the Simplex of R^n, *Operations Research Letters*, v.8, n.4, pp.219-222, (1989).

[41] Lima, P. M. V. ; Morveli-Espinoza, M. M. M. ; Pereira, G. C. ; Ferreira, T. O. ; França, F. M. G., Logical Reasoning via Satisfiability mapped into Energy Functions, International Journal of Pattern Recognition and Artificial Intelligence, v. 22, pp. 1031-1043, (2008).

[42] Lima, P. M. V.; Morveli-Espinoza, M. M. M.; França, F. M. G., Logic as Energy: a SAT-based Approach, Lecture Notes in Computer Science, v. 4729, pp. 458-467, (2007).

[43] França, F. M. G.; Alves, V. C.; Granja, E. P., Processo de Síntese e Aparelho para Temporização Assíncrona de Circuitos e Sistemas Digitais Multi-Fásicos, (2007). Patente: Privilégio de Inovação. nº PI9703819-9, "Processo de Síntese e Aparelho para Temp" ; 01 jul. 1997 (Depósito); 03 ago. 1999 (Exame); 03 jul. 2007 (Concessão).

[44] Arantes Júnior, G. M.; França, F. M. G.; Martinhon, C. A. J., Gerando Orientações Acíclicas com Algoritmos Randômicos Distribuídos, Pesquisa Operacional, Rio de Janeiro, v. 25, n. 3, pp. 301-312, (2005).

[45] Lima, P. M. V.; Pereira, G. C.; Morveli-Espinoza, M. M.; França, Felipe M. G., Mapping and Combining Combinatorial Problems into Energy Landscapes via Pseudo-Boolean Constraints, Lecture Notes in Computer Science, Berlin, v. 3704, pp. 308-317, (2005).

[46] Diacovo, R.; França, F. M. G.; Lima, P. M. V., Generation Expansion Planning as Particle Swarm Optimization with Gridified SATyrus, In: Third EELA Conference, 2007, Catania, Proceedings of the Third EELA Conference, pp. 195-204, (2007).

Relevant theses and dissertations

Luidi Gelabert Simonetti. Otimização combinatória: problemas de árvores geradoras em grafos. 2008. Tese (Doutorado em Engenharia de Sistemas e Computação) - Universidade Federal do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq. *Orientadores:* Nelson Maculan e Abílio Lucena.

Eduardo Filizola Soares. Planejamento de linhas aéreas: escala de tripulantes utilizando métodos heurísticos. 2007. Tese (Doutorado em Engenharia de Sistemas e Computação) - Universidade Federal do Rio de Janeiro. Orientador: Nelson Maculan e Celso Ribeiro.

Pedro Paulo Rodrigues Teixeira Filho. Novas Relaxações Lineares para o Problema de Scheduling de Processos em Multiprocessadores com Restrição de Atraso. 2006. Dissertação (Mestrado em Engenharia de Sistemas e Computação) - Universidade Federal do Rio de Janeiro, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES. Orientador: Nelson Maculan.

Alexandre Salles da Cunha. Árvores ótimas em grafos: modelos, algoritmos e aplicações. 2006. Tese (Doutorado em Engenharia de Sistemas e Computação) - Universidade Federal do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq. Co-Orientador: Abílio Lucena e Nelson Maculan.

Ana Lúcia Gouveia Pimentel. Uma abordagem heurística para a solução de problemas de recobrimento de conjuntos de grande porte, com aplicação à alocação de tripulações para companhias aéreas. 2005. 91 f. Tese (Doutorado em Engenharia de Sistemas e Computação) - Universidade Federal do Rio de Janeiro, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES. *Orientadores:* Abílio Lucena e Nelson Maculan.

Denise Candal Reis Fernandes. Estratégia ótima de oferta em mercados competitivos de energia. 2005. 95 f. Tese (Doutorado em Engenharia de Sistemas e Computação) - Universidade Federal do Rio de Janeiro, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES. *Orientadores:* Marcia Fampa e Nelson Maculan.

Rosa Maria V de Figueiredo. O poliedro das orientações acíclicas de um grafo sob restrições e suas aplicações. 2005. 112 f. Tese (Doutorado em Engenharia de Sistemas e Computação) - Universidade Federal do Rio de Janeiro. *Orientadores:* Nelson Maculan e Valmir Carneiro.

Elivelton Ferreira Bueno. Geração de Colunas em Problemas de Otimização Combinatória. 2005. 93 f. Dissertação (Mestrado em Engenharia de Sistemas e Computação) - Universidade Federal do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq. *Orientador:* Nelson Maculan.

Elder Magalhães Macambira. Modelos e Algoritmos de Programação Inteira no Projeto de Redes de Telecomunicações. 2003. 184 f. Tese (Doutorado em Engenharia de Sistemas e Computação) - Universidade Federal do Rio de Janeiro, Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro – FAPERJ. *Orientadores:* Nelson Maculan e Cid C. de Souza.

Luidi Gelabert Simonetti. Geração de colunas para o problema de empacotamento de árvores de Steiner. 2003. 60 f. Dissertação (Mestrado em Engenharia de Sistemas e Computação) - Universidade Federal do Rio de Janeiro, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CNPq. *Orientadores:* Nelson Maculan e Abílio Lucena.

Aletéia Patrícia Favacho de Araújo. Paralelização autonômica de metaheurísticas em ambientes de grid. 2008. Tese (Doutorado em Informática) - Pontifícia Universidade Católica do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq. *Orientador:* Celso da Cruz Carneiro Ribeiro.

Thomas Dan Otto. Abordagens computacionais aplicadas aos genomas de microbactérias e Trypanossomatídeos: da montagem à anotação. 2008. Tese (Doutorado em Biologia Celular e Molecular) - Fundação Oswaldo Cruz, Fundação Ataulpho de Paiva. *Co-Orientador:* Celso da Cruz Carneiro Ribeiro.

Thiago Ferreira de Noronha. Algoritmos de otimização aplicados a roteamento e atribuição de comprimentos de ondas. 2008. Tese (Doutorado em Informática) - Pontifícia Universidade Católica do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq. *Orientador:* Celso da Cruz Carneiro Ribeiro.

Alexandre Rocha Duarte. Atribuição de árbitros em competições esportivas: Algoritmos e aplicações. 2008. Tese (Doutorado em Informática) - Pontifícia Universidade Católica do Rio de Janeiro, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES. *Orientador:* Celso da Cruz Carneiro Ribeiro.

Andrea Cynthia dos Santos. Modelos e algoritmos para o problema da árvore geradora de custo mínimo com restrição de diâmetro. 2006. Tese (Doutorado em Informática) - Pontifícia Universidade Católica do Rio de Janeiro, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES. *Orientadores:* Celso da Cruz Carneiro Ribeiro e Abílio Lucena.

Carlos Eduardo Costa Vieira. Heurísticas para o problema das p-medianas conectadas. 2006. Tese (Doutorado em Informática) - Pontifícia Universidade Católica do Rio de Janeiro, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES. *Orientador:* Celso da Cruz Carneiro Ribeiro.

Sebastián Urrutia. Otimização em esportes: programação de tabelas e problemas de classificação. 2005. Tese (Doutorado em Informática) - Pontifícia Universidade Católica do Rio de Janeiro, Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro – FAPERJ. *Orientador:* Celso da Cruz Carneiro Ribeiro.

Dalessandro Soares Vianna. Heurísticas híbridas para o problema da filogenia. 2004. 92 f. Tese (Doutorado em Informática) - Pontifícia Universidade Católica do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq. *Orientador:* Celso da Cruz Carneiro Ribeiro.

Rafael Francisco dos Santos. Uso de cortes canônicos no método de ramificação local para problemas inteiros 0--1 mistos. 2006. Dissertação (Mestrado em Ciência da Computação) - Universidade Estadual de Campinas, Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP. *Orientador:* Cid Carvalho de Souza.

Romulo A. Pereira (co-orientado com Arnaldo Moura). Escalonamento de Atividades de Desenvolvimento de Poços de Petróleo: GRASP e Outras Técnicas. 2005. Dissertação (Mestrado em Ciência da Computação) - Universidade Estadual de Campinas. *Co-Orientador:* Cid Carvalho de Souza.

Warley Gramacho da Silva. Algoritmos para o Cálculo de Estruturas de Proteínas. 2008. Dissertação (Mestrado em Computação) - Universidade Federal Fluminense, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES. *Orientador:* Luiz Satoru Ochi.

Haroldo Gambini Santos. Formulações e algoritmos para o Problema de Programação de horários em escolas. 2007. Tese (Doutorado em Computação) - Universidade Federal Fluminense, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES. *Orientador:* Luiz Satoru Ochi.

Cristiane Maria Santos Ferreira. Algoritmos para o Problema da Árvore Geradora Mínima Generalizado. 2007. Dissertação (Mestrado em Computação) - Universidade Federal Fluminense, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES. *Orientador:* Luiz Satoru Ochi.

Tiago Araujo Neves. Heurísticas com Memória Adaptativa para o Problema de Roteamento e Scheduling de Sondas de Manutenção. 2007. Dissertação (Mestrado em Computação) - Universidade Federal Fluminense, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES. *Orientador:* Luiz Satoru Ochi.

André Renato Villela da Silva. Uma nova modelagem para o problema de escalonamento de tarefas com restrições de recursos. 2006. Dissertação (Mestrado em Computação) - Universidade Federal Fluminense, Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro – FAPERJ. *Orientador:* Luiz Satoru Ochi.

Yuri Abitbol de Menezes Frota. Problemas de coloração em grafos. 2008. Tese (Doutorado em Engenharia de Sistemas e Computação) - Universidade Federal do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq. *Orientador:* Marcia Helena Costa Fampa.

Ana Lúcia de Sousa. Um algoritmo utilizando programação semidefinida para o problema de dois níveis linear. 2006. Tese (Doutorado em Engenharia de Sistemas e Computação) - Universidade Federal do Rio de Janeiro. *Orientador:* Marcia Helena Costa Fampa.

Luiz Augusto Nóbrega Barroso. Estratégias de Ofertas Ótimas sob incerteza e cálculo de equilíbrios de Nash de agentes geradores em mercados de curto prazo de energia elétrica: uma abordagem por programação linear inteira. 2006. Tese (Doutorado em

Engenharia de Sistemas e Computação) - Universidade Federal do Rio de Janeiro. *Orientador:* Marcia Helena Costa Fampa.

Paulo Sérgio Marques dos Santos. Métodos de Projeção para o Problema de Desigualdades Variacionais Gerais. 2007. Tese (Doutorado em Engenharia de Sistemas e Computação) - Universidade Federal do Rio de Janeiro, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES. *Orientador:* Susana Scheimberg de Makler.

André Luiz Diniz Souto Lima. Uma estratégia de decomposição lagrangeana para a otimização da programação diária da operação de sistemas hidrotérmicos com modelagem detalhada da rede elétrica - Aplicação ao sistema brasileiro. 2007. Tese (Doutorado em Engenharia de Sistemas e Computação) - Universidade Federal do Rio de Janeiro. *Co-Orientador:* Susana Scheimberg de Makler.

Flávia Morgana de Oliveira Jacinto. Existência de soluções e dualidade para um problema de equilíbrio generalizado. 2007. Tese (Doutorado em Engenharia de Sistemas e Computação) - Universidade Federal do Rio de Janeiro, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES. *Orientador:* Susana Scheimberg de Makler.

Luciana Roque Brito. Novas Propostas para o Problema de Recobrimento de Rotas. 2005. Tese (Doutorado em Engenharia de Sistemas e Computação) - Universidade Federal do Rio de Janeiro. *Orientadores:* Susana Scheimberg de Makler e Luiz Satoru Ochi.

3.5.3. Modeling the Web and Applications

The area of computer and communication systems modeling and analysis has been an essential part of the Internet development since its birth. Queueing models have been used to show the efficiency of packet over circuit switching architectures and helped establishing the foundations of the design of modern computer networks. The theory of Markov chains was the basis for studying a number of contention resolution algorithms that have been proposed over the years (both for wire and wireless links); for understanding the efficiency of data protocols for distinct layers; for studying the impact of a myriad of scheduling algorithms on performance; for modeling traffic flows in the network and studying their impact on network resources; for understanding the flow and error control problems, etc. The theory was also the key to modern web page ranking algorithms which are essential for discovering information on the web.

Likewise, other theories were also fundamental to the design of the Internet as we know it today, including Graph and Optimization theories (applied, for instance, to routing and network design problems) and, more recently, Game theory that helped understanding the consequences of new applications that do not obey the rigid control of network protocols. New emerging applications often establish virtual application-centered overlay networks with the objective of optimizing the application's performance. This is done regardless of the possible negative impact that a selfish, self-centered application behavior may have on the overall network performance. Thus, conflicts of interest are now commonplace in the modern Web, and predicting their impact on the performance and evolution of the Web is a challenging problem.

Pairing with modeling, the field of Internet measurement arose as an important discipline to understand performance properties, to tune mathematical models, to provide information required to develop products, and to obtain insights about social implications of Web use. Modeling and measurement are tightly coupled, since the later may provide input parameters needed by the models or validate modeling assumptions or results obtained from models. Measuring the Internet has blossom as a discipline that requires sophisticated mathematical tools to collect and analyze the required data.

Modeling has paralleled the expansion of the Internet. Modeling has been mainly used over the years to understand the impact of protocols on the efficiency of applications or the transfer of data. However, there is a need to understand the complex interactions among the algorithms that control the flow of data, the applications and the users. This is essential to the system's robustness.

The Internet has grown considerably from its birth in 1969 at UCLA, with just a couple of connected nodes and used by a handful of computer scientists and students, to hundreds of millions of devises and users widespread all over the globe. It is a globally pervasive system that has changed our way of life. Modeling and understanding this vast complex system is a challenge that requires a combination of different expertise and the development of new techniques perhaps combining theories from different fields.

The evolution of Web technologies enabled a number of applications. Users of these new applications have their behavior highly influenced by the new technologies. As a consequence, new demands are generated which in turns exert pressure on the systems leading to the development of new technologies. The global Web can be seen as evolving like a spiral where the advent of new technology spawns new applications which influences the behavior of users which in turns causes the dawn of new technologies. How this process evolves with time? Are there fundamental laws that govern the expansion and interactivity of this complex "organism"? Recall we are faced with a large system with a huge number of interrelated and interacting parts.

One could argue that the Web has certain similarities to a self-organizing system. By self-organization we mean a process that increases in complexity without the guidance or management of a central authority; or yet, a process in which global level patterns emerge from interactions among lower-level components and the rules that govern the components interactions are executed using only local information. Why these similarities exist? First note that the Internet has grown based on the freedom of users. Focusing on a recent architecture to clarify the ideas, consider a P2P system. This is an example of an architecture where the users are organized without the control of a central authority, they freely cooperate to obtain individual benefits, may use local information to download data, and the system can sustain growth in a scalable and robust manner. This application has been possible by the current level of technological development, has influenced the way we behave, and impacted business. For example, the traffic from P2P applications is fast growing and is taking a considerable amount of the network resources. The changes on user behavior are causing a negative impact on business such as service providers, which could not predict the traffic patterns generated by these applications. What will be the most probable scenario in the near future? Will business enforce collectively some kind of

access restriction to P2P data? Or somehow the system will adapt to new business models?

Robustness and scalability are extremely important issues considering the Web. Can an application architecture sustain a continuous growth of its user population without affecting its performance and availability? What are the performance impacts that self-regarding applications have on others that share common resources? As briefly mentioned above, Web users and several Web applications are self-regarding in which decisions are made towards maximizing their individual goals. Studying the dynamic process by which users and applications adapt to the varying network conditions is a challenging and extremely difficult task.

One basic premise of P2P architectures is the willingness of a fraction of the participants of the system to service others. Users are selfish and an issue is to model the user behavior and incentive mechanisms that lead to and increase in robustness for the overall community.

Pursuing further on the P2P example, it is not difficult to foresee the application of real time multimedia data sharing on education. For instance, consider a distance learning application, where students could have access to video material from our best educators, virtual laboratories and real time lectures. Can we foresee a self-organizing community of students that, using the proper applications, individually have a learning rate higher than the normal lecturer-classroom way of teaching? Can we understand, from models of such systems, the impact on education and make the right government policies for the future?

In this scenario, we also propose a strategy for supporting research on Software Engineering (SE) Education, based on experimentation, aiming at contributing for the organization of a body of knowledge in SE Education in Brazil. The strategy involves the execution of secondary (systematic review) and primary (survey) studies. The objective is to structure and focus the SE community, pursuing a collaborative and large scale research, where various research groups will contribute to their respective research topic. This strategy may serve as a structuring framework to organize the education research in computer science in general.

Within the scope of modeling the Web, modeling the web graph emerges as one of its most interesting objects of study. In this graph, vertices represent web pages and directed edges represent the hyperlink between pages. The web graph is a fascinating object of study, as it defines the structure of arguably the most prevalent application in the Internet, the Web. The structure of the web graph has implications that span several different areas, such as technological, commercial and sociological. As one would expect, the web graph has been studied practically since the boom of the Internet, in the late 90s. As part of these studies, several models to represent different aspects of the web graph, such as structure at a given point in time or its evolution over time, have been proposed in the literature.

However, with the advance of Web technology, such as the promises of Web 2.0, the web is changing fundamentally. Web pages that were mostly static are now becoming mostly dynamic, to the point of being personalized to individual users. For example, two different users visiting the same URL, which uniquely identified a web

page, can now be exposed to different information, with different hyperlinks. In this context, even defining a web graph becomes an issue. In the extreme, every individual user may induce a web graph. Modeling this highly dynamic object is certainly a challenge and a necessary step towards understanding its implications on a myriad of important aspects. For example, will a highly dynamic web hinder the robustness of the system to failures of servers or the system scalability? In a different perspective, will a highly dynamic web induce a greater rate of structural changes or faster technological improvements?

In this proposal we aim at understanding the dynamics of the whole by focusing on self organizing Web sub-systems like P2P. Perhaps general laws can be revealed by focusing on large P2P sub-systems and other applications. Furthermore, understanding and modeling the structure of the future web graph is of fundamental importance to address robustness issues of the overall system and to answer the important questions raised above. One target application will be distance learning since our group has years of experience on developing and deploying tools that are currently in use in the Computer Science course of the CEDERJ public universities consortium of the state of Rio de Janeiro and the Open University of Brazil (UAB). In fact we have access to a distributed Laboratory where the human subjects (the students) can be monitored to provide data to study and model the way video-class information is accessed. Another application area is to explicitly model collaboration and reuse processes adopted in open software projects. In order to mine data and discover process fragments, mining process techniques can be used to build social networks that represent the open software community collaboration model. The National Research Network (RNP) will be a partner in this project and would provide a large laboratory to perform experiments and to collect measurements.

Goals, Activities, and Deliverables

G1. Study robustness/scalability issues of P2P systems and develop robustness models for those.

Activities:

- Define metrics to evaluate the robustness and scalability of P2P systems.
- Develop experiments with P2P systems aiming at finding the scenarios and metrics that have significant impact on the robustness of the system.
- Develop models to characterize the behavior of P2P systems aiming at studying their robustness and scalability.
- Study different phenomena in P2P architectures such as churn, selfish and opportunistic user behavior, incentive mechanisms, and their impact on robustness.
- Study P2P architectures for video streaming.

Deliverables:

- A modeling framework for evaluating robustness and scalability in P2P systems.

- International and national journal and conference papers, master dissertations and PhD thesis.

G2. Development of methods, tools and techniques for e-learning

Activities:

- Perform measurements of user behavior in a large scale distance learning system employed in the CEDERJ/UAB Computer Science course.
- Develop models aiming at characterizing users' behavior of the Computer Science CEDERJ/UAB interactive video streaming course.
- Modeling an analysis of the complex interactions between users and the application.

Deliverables:

- A user model framework to be used in robustness studies of distance learning streaming course application.
- Understanding the systems behavior aimed at providing better services to the user (student) population.

G3. Development of algorithms for P2P systems

Activities:

- Propose mechanisms to increase the scalability of P2P systems.
- Propose algorithms to improve the robustness of P2P systems.

Deliverables:

- Algorithms to improve the robustness and scalability of P2P systems.
- International and national conference and journal papers, master dissertations and Ph.D. thesis.

G4. Model the future web graph, where content is highly dynamic and personalized.

Activities:

- Capture the dynamic web. Design data-mining techniques that can gather information about the dynamic content of the web.
- Search and identify laws governing the dynamic web.
- Capture the structure of web servers required to construct dynamic web pages. Design techniques to obtain data at large-scale.
- Search and identify laws governing structure of web servers.
- Construct generative models for the dynamic and personalized web. Web users become an important entity in the model.
- Predict evolutionary patterns of the dynamic Web.

Deliverables:

- Publish results in international conferences and international journals.
- Software package to generate instances of the web based on developed models.

G5. Build social networks that represent the open software community collaboration model.

Activities:

- Study open source software project characteristics;
- Plan and perform experiments to ascertain the proposal applicability.

Deliverables:

- Model of social networks in open source software projects;
- Report analysis of the experiments results.

Timeframe

Goal	Activity	Y1	Y2	Y3	Y4	Y5
Study robustness/scalability issues of P2P systems and develop robustness models for those.	Define metrics to evaluate the robustness and scalability of P2P systems.	X	X			
	Develop experiments with P2P systems aiming at finding the scenarios and metrics that have significant impact on the robustness of the system.	X	X			
	Develop models to characterize the behavior of P2P systems aiming at studying their robustness and		X	X	X	
	Study different phenomena in P2P architectures such as churn, selfish and opportunistic user behavior, incentive mechanisms, and their impact on robustness.		X	X	X	
	Study P2P architectures for video streaming.	X	X	X		
E-learning	Perform measurements of user behavior in a large scale distance learning system employed in the CEDERJ/UAB Computer Science course.	X	X			
	Develop models aiming at characterizing users' behavior of the Computer Science CEDERJ/UAB interactive video streaming course.	X	X			
	Modeling an analysis of the complex interactions between users and the application.			X	X	
Development of algorithms for P2P systems	Propose mechanisms to increase the scalability of P2P systems.		X	X	X	
	Propose algorithms to improve the robustness of P2P systems.		X	X	X	
Model the future web graph, where	Capture the dynamic web. Design data-mining techniques that can gather information about the dynamic content of the web.	X	X	X		

	Search and identify laws governing the dynamic web.	X	X	X		
	Capture the structure of web servers required to construct dynamic web pages. Design techniques to		X	X	X	
	Search and identify laws governing structure of web servers.		X	X	X	
	Construct generative models for the dynamic and personalized web. Web users become an important		X	X	X	
	Predict evolutionary patterns of the dynamic web.			X	X	
Build social networks that represent the open software community collaboration model	Study open source software project characteristics		X	X	X	
	Plan and perform experiments to ascertain the proposal applicability		X	X	X	

Team

Name	Affiliation	Degree	CNPq Level
Edmundo A. de Souza e Silva	PESC-UFRJ	Ph.D. UCLA 1984	1A
Rosa M.M. Leão	PESC-UFRJ	D.Sc. Paul Sabatier 1994	-
Daniel R. Figueiredo	PESC-UFRJ	Ph.D. UMASS 2005	-
Valmir Barbosa	PESC-UFRJ	Ph.D. UCLA 1986	1A
Claudia Werner	PESC-UFRJ	D.Sc. COPPE/UFRJ	1D

Recent Publications

WATANABE, E. H. ; MENASCHÉ, Daniel Sadoc ; E. de Souza e Silva ; LEÃO, R. M. M; Modeling Resource Sharing Dynamics of VoIP users over a WLAN Using a Game-Theoretic Approach; IEEE INFOCOM 2008.

Pedram Pedarsani, Daniel R. Figueiredo, Matthias Grossglauser. Densification arising from sampling fixed graphs. Proc. ACM SIGMETRICS, 2008.

Elias Bareinboim and Valmir C. Barbosa, Descents and nodal load in scale-free networks, Physical Review E 77 (2008), 046111.

R. Santos, P. Santos, C. Werner e G. Travassos. Utilizando Experimentação para Apoiar a Pesquisa em Educação em Engenharia de Software no Brasil. I Fórum de Educação em Engenharia de Software, SBES 2008, Outubro, Campinas.

ROCHA, Antonio Augusto de Aragão ; LEÃO, R. M. M. ; E. de Souza e Silva, An End-to-End Technique to Estimate the Transmission Rate of an IEEE 802.11 WLAN; IEEE ICC 2007.

Alexandre O. Stauffer and Valmir C. Barbosa, Probabilistic heuristics for disseminating information in networks, IEEE/ACM Transactions on Networking 15 (2007), 425-435.

ROCHA, Antonio Augusto de Aragão ; LEÃO, R. M. M. ; E. de Souza e Silva A non-cooperative active measurement technique for estimating the average and variance of the one-way delay; IFIP - Networking 2007, Lecture Notes in Computer Science, 2007, v. 4479, p. 1084-1095.

VIELMOND, C. C. L. B. ; LEÃO, R. M. M. ; E. de Souza e Silva Um modelo HMM hierárquico para usuários interativos acessando um servidor multimídia; Simpósio Brasileiro de Redes de Computadores, 2007, v. I, p. 469-482.

ALVES, B. C. B. ; LEÃO, R. M. M. ; E. de Souza e Silva; Caracterizando variáveis de interatividade dos alunos do curso de computação do CEDERJ baseado no servidor multimídia RIO; SBC2007 - VI Workshop em Desempenho de Sistemas Computacionais e de Comunicação; p. 739-757.

M. Garetto, D. R. Figueiredo, R. Gaeta, M. Sereno A modeling framework to understand the tussle between ISPs and peer-to-peer file-sharing users. Performance Evaluation, 64 (9-12), pp 819-837, (2007).

E. de Souza e Silva ; LEÃO, R. M. M. ; SILVA, Ana Paula Couto da ; ROCHA, Antonio Augusto de Aragão ; DUARTE, F. P. ; SILVEIRA FILHO, F. J. ; JAIME, G. D. G. ; MUNTZ, R. R. . Modeling, Analysis, Measurement and Experimentation with the Tangram-II Integrated Environment. International Conference on Performance Evaluation Methodologies and Tools, 2006, v. 180, p. 1-10.

E. de Souza e Silva ; LEÃO, R. M. M. ; SANTOS, A. D. ; AZEVEDO, Jorge Allyson ; MACHADO NETTO, Bernardo Calil. Multimedia Supporting Tools for the CEDERJ Distance Learning Initiative applied to the Computer Systems Course. 22th ICDE World Conference on Distance Education, 2006, p. 1-11.

D.S. Menasché, D.R. Figueiredo, E. de Souza e Silva, ``An evolutionary game-theoretic approach to congestion control'', Performance Evaluation, vol. 62, pp. 295-312, 2005.

Zihui Ge, Daniel R. Figueiredo, Sharad Jaiswal, Jim Kurose, Don Towsley Modeling Peer-Peer File Sharing Systems. Proc. of IEEE INFOCOM, 2003.

Zihui Ge, Daniel Ratton Figueiredo, Sharad Jaiswal, Lixin Gao. On the Hierarchical Structure of the Logical Internet Graph. Proc. of SPIE ITCom, 2001.

Relevant theses and dissertations

Completed

Bruno Cesar Barbosa Alves. Caracterizando variáveis de interatividade dos alunos do curso de computação do CEDERJ, baseado no servidor multimídia RIO. Master Dissertation, COPPE-Sistemas, UFRJ (2008). (Orientadores: Edmundo de Souza e Silva e Rosa Maria Meri Leão)

Edson Hiroshi Watanabe. Sobre a disputa por recursos de WLAN por aplicações VoIP: uma proposta de modelagem usando teoria dos jogos. Master Dissertation, COPPE-Sistemas, UFRJ (2007). (Orientadores: Edmundo de Souza e Silva e Rosa Maria Meri Leão)

Carolina Cerqueira Le Brun de Vielmond. Um modelo HMM hierárquico para usuários interativos acessando um servidor multimídia. Master Dissertation, COPPE-Sistemas, UFRJ (2007). (Orientadores: Edmundo de Souza e Silva e Rosa Maria Meri Leão)

Ongoing

Luiz José Hoffman Filho. Avaliando arquiteturas P2P para serviços de streamimg de vídeo. Início: 2007. Master Dissertation, COPPE-Sistemas, UFRJ. (Orientadora: Rosa Maria Meri Leão)

Bruno César Barbosa Alves. Garantindo maior robustez aos sistemas P2P. Início: 2008. PhD Thesis, COPPE-Sistemas, UFRJ. (Orientadores: Edmundo de Souza e Silva e Rosa Maria Meri Leão)

4. Institute Organization

4.1. The Institute Headquarter

The Systems Engineering and Computer Science Graduate Programme (PESC) of COPPE/UFRJ has been evaluated with maximum grade (7) by CAPES (Ministry of Education agency) for the period 2004/ 2007. PESC was created in 1970 with the purpose of developing scientific research in the areas of Computer Science and Optimization through high-level academic qualification of both researchers and faculty. At that time, the country yielded a special effort on rapidly absorbing advanced technologies produced abroad, in order to improve industrial capabilities in Computer Science. The Programme intensely participated in the process, by hiring foreign faculty and by offering graduate (Doctorate and Master) courses, thus providing the country with competent professionals to deal with these new technologies as well as to develop innovative solutions for Brazilian needs.

In the 80's, PESC sought to outline its profile by hiring Brazilian PhDs graduated at the best universities abroad. This helped to consolidate the key areas in Computer Science, such as Computer Architecture and Operating Systems, Databases, Software Engineering, Optimization, Computer Graphics, Artificial Intelligence and Social Aspects of Computing. During the following decade, twenty-year-old PESC had become a highly reputed institution both for its academic production and for its faculty and alumni. By then, new areas had already been incorporated to its original structure, namely Algorithms and Combinatorics and Computer Networks.

Interdisciplinary and multi-institutional groups, dedicated to the study of Computational Biology, Data Mining, Medical Informatics, Parallel and Distributing Computing and Telecommunications, were created to keep the Programme aligned with the world research trends. One of the twelve programmes that compose COPPE/UFRJ, PESC occupies a privileged position regarding interdisciplinary activities as it can closely interact with other highly rated programmes (6 or 7) of various technological areas.

As we stand in the first decade of the XXI Century, thirty-year-old PESC constitutes a solid, traditional institution in the national and international scenario of Computer Science and Technology. Its main challenge consists of maintaining its pioneer position in innovative research areas of Computer Science in a dynamic, unpredictable, globalized world.

To cope with those new challenges, PESC requires increasingly excellence in academic production from its faculty. Additionally, high-potential young researchers are invited to join the Programme to work in strategic areas and members of faculty are sent abroad on sabbatical. Among its thirty-six permanent faculty members, twenty-three are qualified as distinguished researchers by the Brazilian Research Council (CNPq), being awarded fellowships of several levels. It is worth noticing that five of the twelve researchers awarded the highest fellowship level (1A) in Computer Science belong to PESC, four of which take part in this project.

During its thirty-eight years of existence, PESC awarded 1034 M.Sc. degrees and 281 D.Sc. degrees. Since the beginning of 2008 to present, 29 dissertations and 14 theses have been issued according to rigorous standards. The high quality of our alumni ensures rapid absorption by national academic and governmental institutions and the most important industrial companies. Another quality indicator of our courses is the number of students that apply from other parts of Brazil and abroad.

Undergraduate courses, for example, Computer Science, Electronic Engineering (emphasis on Computing), also benefit from the excellence of PESC faculty. Its members actively participate in these courses by teaching as well as by supervising final projects and scientific initiation works. In 2003, PESC involvement in undergraduate teaching has increased since the establishment of a partnership with the Polytechnic School of Engineering of UFRJ that created the course on Information and Computing Engineering. The first class, which started in 2004, is due to graduate at the end of this year. Currently, PESC also collaborates with undergraduate courses of the Biophysics Institute, as a consequence of the increasing interaction between the areas.

4.2. Institute Infrastructure

The institute will heavily benefit from the infrastructure already available at the participating departments. This includes multiple laboratories, public or thematic, classrooms, auditoriums, administrative areas and other facilities. All institutions will offer space for the installation of an interaction room, where tele-presence activities can be enacted.

As the headquarter institution, PESC-UFRJ will create a new area, expected to have of 80m^2, to support administration, a visiting professor or pos-doc fellow, a new development laboratory and the interaction room. Most part of the investment for this new area will be done by PESC-UFRJ.

PESC-UFRJ will also support administratively the institute, offering 2 administrative persons (20% time) to its activities, to be joined by a third person that is already predicted in the budget. Fundação Coppetec will be used to help in procurement and accounting issues.

The two larger participant institutions will be actively dedicating most of its laboratories to this project. Specifically, PUC-Rio and UFRJ offers the following thematic laboratories:

The Informatics Department at PUC-Rio hosts 13 thematic laboratories, grouped under the umbrella of the Software Technology Institute – ITS and involving about 350 researchers. The labs directly involved with the institute and their main research interest are:

LAC – Laboratory for Advanced Collaboration (http://www.lac.puc-rio.br): Mobile Computing

LEARN – Laboratory for Algorithm Engineering and Neural Networks: Fuzzy Systems; Neural Networks; Optimization

LES – Software Engineering Laboratory (http://www.les.puc-rio.br): Software Engineering

SERG – Semiotic Engineering Research Group Laboratory (http://www.serg.puc-rio.br): Human-Computer Interfaces

TecBD – Database Technology Laboratory (http://www.tecbd.inf.puc-rio.br): Database Technology

TecGraf – Computer Graphics Laboratory (http://www.tecgraf.puc-rio.br): Techno-Scientific Interactive Computer Graphics Systems

TecWeb – Web Engineering Laboratory (http://www.tecweb.inf.puc-rio.br): Web Applications Engineering

The PESC/UFRJ hosts the following laboratories:

LABOTIM: Optimization (Bloco I-2000; Nelson Maculan)

LBD: Databases (Bloco I-2000; Jano Moreira de Souza)

LSIG: Geographic Information Systems (Bloco I-2000; Jano Moreira de Souza)

LabGCCBD: Knowledge Management, Databases (Sala H-317; Jano Moreira de Souza)

LENS: Software Engineering (Bloco I-2000; Guilherme Horta Travassos)

LAND: Modeling, Analysis and Development of Computing and Communication Systems (Bloco I-2000; Edmundo A. de Souza e Silva)

LAM: Computer architecture and microelectronics (Sala H-319; Felipe Maia Galvão França)

4.3. Institute Organizational Structure

The Institute Organizational Structure is divided into three layers. The Steering Committee is located at the top layer. Five Research Managers (one for each stratum) compose the management team located in the middle layer. Several Researching Groups are located at the bottom one. A finance and procurement manager does provide support to decisions held by the Steering Committee.

The Steering Committee is composed by six members. Its main responsibilities are: to approve the annual research plan; to allocate the budget aligned with this proposal; to manage the institute operations; and to report the institute performance to CNPq. Additionally an Advisory Board, composed by a Brazilian and a foreign researchers, on an annual basis, will audit the institute performance according to the plan previously approved by this committee, reporting written recommendation to guide the institute in achieving the goals enumerated in this proposal. This committee activities are described in the Stakeholders Liaison Function and its structure is the content of Figure 3.

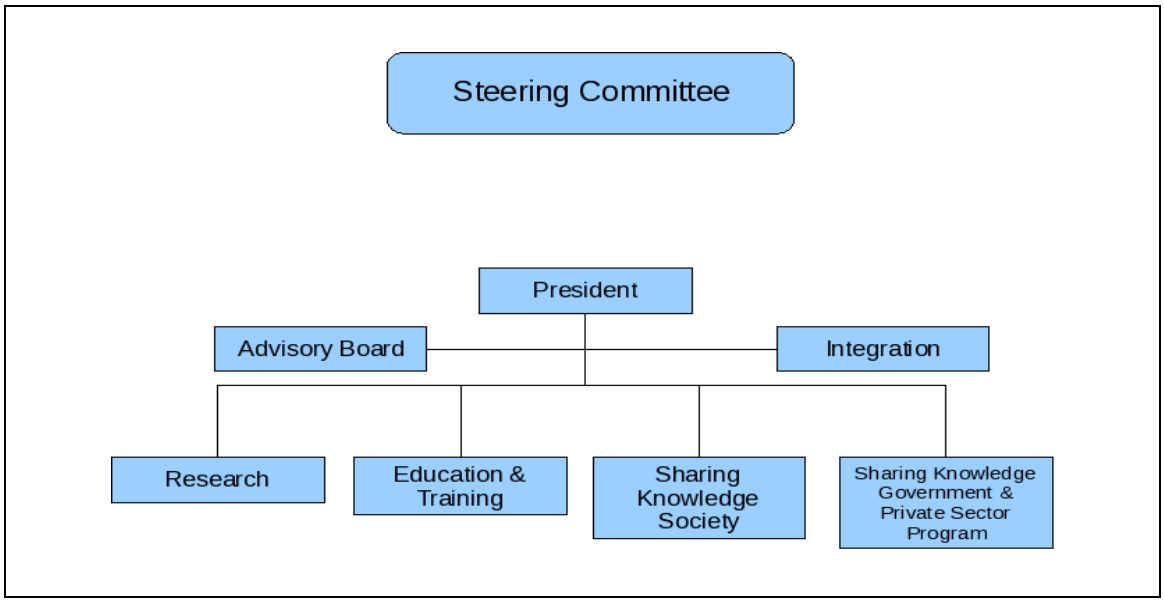

Figure 3: Steering Committee Composition

In order to fulfill the above mentioned responsibilities, the Steering Committee will be structured as follows:

- Prof. Nelson Maculan Filho (President)

  As the institute coordinator Prof. Maculan will be the steering committee president. He will focus on the liaison between the institute and the stakeholders.

  Prof. Maculan has a B.Sc. degree in Mining and Metallurgy Engineering, School of Mines, Ouro Preto, (1965), a M.Sc. degree (DEA) in Mathematics Statistics, Université de Paris VI (Pierre et Marie Curie) (1967), a Ph.D. degree in Operations Research, Federal University of Rio de Janeiro (1975) and the Diplôme d'Habilitation à Diriger des Recherches (DHR) en Sciences de la Gestion (1988), Université Paris-Dauphine (Paris IX). He is currently Full Professor of the Federal University of Rio de Janeiro. He has a vast experience in the field of computer science, with emphasis on computational mathematics, acting mainly on the following themes: combinatorial optimization, integer programming, linear programming, generation of columns and overall optimization. Prof. Maculan receives a level 1A productivity fellowship for his research from CNPq. His research areas encompass graph theory, optimization of graphs and combinatorial optimization. Former President of the Federal University of Rio de Janeiro, Director of the Alberto Luiz Coimbra Graduate School of Engineering (COPPE/UFRJ), and Head of Department of the Systems Engineering and Computer Science Graduate Program (COPPE/UFRJ), he also served as National Secretary for Higher Education (Ministry of Education) and, recently, as State Secretary for Education in Rio de Janeiro, Brazil. Prof. Maculan was awarded several honorary degrees, such as: Chevalier dans l'Ordre National du Mérite, Paris, France; Docteur Honoris Causa, Université Paris 13, France; Doctor Honoris Causa, Universidad

Nacional Mayor de San Marcos, Lima, Peru, and with the Medal of the Grand Cross – National Order of Scientific Merit – by the Presidency of the Republic, Brazil. Prof. Maculan is a full member of the Brazilian Academy of Sciences.

- Prof. Carlos José Pereira de Lucena (Integration Mission)

  The Integration Mission is responsible for conducting the internal operations and execution of the Institute's operational plan. Four missions fall within his purview: addressing research, training and education, sharing knowledge with society and sharing knowledge with government and private sector.

  Prof. Lucena has graduated in Economics and Mathematics from the Pontifical Catholic University of Rio de Janeiro (1965), he got a Master´s of Mathematics from the Department of Computer Science & Applied Analysis of the University of Waterloo (1969), a PhD in Computer Science from the Computer Science Department of the School of Engineering and Applied Sciences of the University of California at Los Angeles (1974) and was a post-doctoral research fellow at the IBM Research Laboratory in Palo Alto (1975). Since 1982 he is a full professor at the Pontifical Catholic University of Rio de Janeiro. He is also an adjunct professor at the University of Waterloo and a Research Associate at the Fraunhofer FIRST Institute in Berlin. He is a member of the editorial board of several international journals: the Communications of the ACM, the Proceedings Journal of the Brazilian Academy of Sciences, The International Journal of Formal Aspects of Computing and the International Journal of Agent-Oriented Software Engineering (IJAOSE). He has experience in the field of Computer Science, with emphasis on formal aspects of computing, software engineering and multi-agent systems. Prof. Lucena receives a level 1A research productivity fellowship from the CNPq (Research Council). In his career he worked at the Pontifical Catholic University of Rio de Janeiro as vice-rector of the University, Dean for the Center for Science and Technology, and several times Director of the Computer Science Department. Prof. Lucena was awarded the insignia for the Grand Cross of the Order of Scientific Merit of the Presidency of the Republic of Brazil, the Medal of Scientific Merit Carlos Chagas Filho from the Council of Directors of FAPERJ (State Research Council), the Alvaro Alberto Award for Science and Technology (Ministry of Science and Technology) and several times the IBM Innovation Award, among many others . Prof. Lucena is a fellow of the Guggenheim Foundation and a member of the Brazilian Academy of Sciences.

- Prof. Valmir Carneiro Barbosa (Research Mission)

  The Research Mission is primarily concerned with promoting a high quality international standard research program aligned with the Action Plan for 2007-2010 for Science, Technology and Innovation (PACTI in Portuguese).

  Prof. Barbosa holds a PhD in computer science (University of California, Los Angeles, 1986), is professor of the Federal University of Rio de Janeiro. He operates in the area of computer science, with emphasis on parallel and distributed computing and models of complex systems. He receives a level 1A research productivity fellowship from the CNPq. Was a Visiting Scholar at the

University of California, Berkeley, UCB, USA. Member of editorial board of IEEE Journal of Transactions on Computers and Journal of the Brazilian Computer Society.

- Prof. Edmundo A. de Souza e Silva (Education and Training Mission)

  The Education and Training Mission is devoted to developing and implementing a high quality international standard program at undergraduate and graduate levels. The implementation of training programs (short and long duration) for government and private sector personnel is among the objectives of this Mission, enabling the flow of knowledge from the institution into these sectors. One big challenge in this Mission is to guarantee that the education and training programs have a nationwide coverage.

  Prof. Souza e Silva has an Electrical Engineering degree from PUC-Rio (1975), a M.Sc. in Electrical Engineering from PUC-Rio (1978) , a Ph.D. in Computer Science from the University of California, Los Angeles (1984), and has residencies at the IBM Watson Research Center (1985) and the University of California, Los Angeles (1994) . He is currently Full Professor at UFRJ. He has experience in Computer Science in the areas of availability, performance, queueing networks and reliability. He has a productivity scholarship from CNPq at level 1A. He is visiting professor ate the Chinese University of Hong Kong, at the University of California, Los Angeles and at the University of Southern California, among others. He has the Medal of Scientific Merit – Commendator, and he is a member of the Brazilian Academy of Science.

- Prof. Marco Antonio Casanova (Transfer to Society Mission)

  The Transfer to Society Mission, besides the publication of articles in journals and congresses, is responsible for promoting, through its researchers and graduated students, training programs target to the general public, focusing high school audiences.

  Prof. Casanova has graduated in Electronic Engineering at the Military Institute of Engineering (1974), holds a Master degree in Computer Science from the Catholic University of Rio de Janeiro (1976), Master in Applied Mathematics from Harvard University (1978) and Ph.D. in Applied Mathematics also from Harvard University (1979). Today he is Associate Professor and Director of the Department of Information Technology of the Pontifical Catholic University of Rio de Janeiro. His main area of research is Database. Prof. Casanova receives a level 1C productivity fellowship for his research from CNPq.

- Prof. Claudia Maria Bauzer Medeiros (Transfer to Government and Private Sector Mission)

  Building a network of private and governmental initiatives enabling the participation of qualified institute personnel in the development cycle of new products and services where emerging knowledge from the institute will be embedded is part of this mission objectives. A narrow relationship with innovative enterprises embraced within the Brazilian Technology System (SIBRATEC in Portuguese) will also be a point of focus.

Full professor at the Institute of Computing, UNICAMP, since 2001, has graduated in Electrical Engineering - Catholic University of Rio de Janeiro (1976), MSc in Computer Science - Catholic University of Rio de Janeiro (1979), Ph.D. in Computer Science - University of Waterloo (1985). Member of the Advisory Committee of Computer Science and Engineering of FAPESP (2004-present), was coordinator of the Advisory Committee on Computer Science at CNPq (2001-2002), vice coordinator of the Advisory Committee on Computer Science at CAPES (1998-2000). Conducts research in databases, with emphasis in scientific databases, scientific workflows and eScience (biodiversity and agro-environmental planning). Former president of the Brazilian Computer Society (2003-2007). Receives a level 1B research fellowship from CNPq. She was awarded the Newton Faller prize (SBC 2000), the Change Agent Award (ACM and Anita Borg Institute) and twice the Zeferino Vaz Academic Excellence Award (UNICAMP). In 2007, granted a Doctor Honoris Causa by the University Antenor Orrego, Trujillo, Peru and in 2008 admitted at the Brazilian Order of Scientific Merit, as Commander.

Among the many actors in the complex environment this institute will act, we highlight four as stakeholders. The first is CNPq for sponsoring and auditing. The second is Government and Private Sector from where demand will emerge and through whom the flow of new knowledge and skilled human resources will add value to society. The third one is the Scientific Community whose knowledge will seed the research function and, in return, will benefit from the assessed deliverables (mostly in the form of published articles) from the same function. The fourth is Society whose requirements are the reason for the existence of this institute and, at the bottom line, to whom, in return, it will serve.

4.4. Institute Functional Structure

Four layers constitute the Functional Structure of the Institute. The Institutional Function is located at the top layer. Its goal is to establish the liaison between the institute and the stakeholders. This function is bounded by feedback from society, recommendations from the Advisory Board, the PACTI, the institute approved proposal and demands from government and private sector. Capital goods, human resources and income constitute the inputs for this function and the outputs are; CNPq assessment report; institute deliverables; and invoice and payments.

The management layer lies below the top one. It consists of 4 functions: Institute Management Function, Operations Function, Finance and Procurement Function; and Performance Assessment Function. The Institute Management Function main objective is to manage and assess the institute agenda. It is bounded by the PACTI, Advisory Board recommendations; the Institute Approved Proposal; and CNPq Assessment Report. The only input is the Proposed Operational Plan that comes from the Operation Function. The output is the Approved Operational Plan.

The Operations Function's objective is to execute the plan of the four missions, i.e., Research, Training and Education, Sharing Knowledge with Society and Sharing

Knowledge with Government and Private Sector. It is bounded by Government and Private Sector Demands and Feedback from Society. Inputs are Human Resources, Supplies and Capital Goods. Outputs are: Proposed Operational Plan, Performance Report and Expense Requests.

The Finance and Procurement Function's objective encompasses the execution of back-office activities like accounting, purchasing, bidding, invoicing and payments. It is bounded by Approved Operational Plan from the Institute Management function. Inputs are income from outside sponsors and expense requests from Operation Function. Outputs are: Invoice and Payments; and performance report to the Performance Assessment function.

The Performance Assessment Function's objective is to compare the operation performance report and the financial performance report against the approved operational plan. After receiving these reports from their respective functions it produces the CNPq assessment report, closing the managerial cycle of planning, approving, executing, measuring and assessing the whole operation.

The third layer is the operational layer. It encompasses the Research Mission Function, Training and Education Function, Sharing Knowledge with Society and Sharing Knowledge with Government and Private Sector Functions. We will delve into these functions at the fourth layer. From the managerial point of view each manager is responsible for the four missions within his or her stratum.

At the fourth layer are the sets of functions the constitute each mission. Each set consists of a planning, execution and assessment functions. The Planning function is bounded by government and private sector demands as well as feedback from society. Input comes from the mission performance assessment function (the mission assessment report) and output is the mission proposed operational plan to be approved in the institute management function at the second layer. The mission execution function, bounded by the approved operational plan has no input. Outputs differ for each mission.

The management team's responsibilities focus the execution of the approved plan and measuring its performance to report to the steering committee its development. Six managers constitute this team. Their activities executioners of the functions described below and, while performing them, will be concerned with reaching the goals across the four missions within the agreed time schedule, allocating human and financial resources when and where necessary, and leveraging integration inter and intra strata.

4.5. Advisory Board

The Advisory Board is constituted by two internationally recognized researchers:

- Prof. Philippe Michelon (B.Sc., D.E.A, Ph.D., H.D.R.) is a full professor in computer science and vice president for research at Université d'Avignon et des Pays de Vaucluse. He is the advisory member working in France (http://www.univ-avignon.fr/fr/presentation/organisation/conseils.html); and

- Prof. Lindolpho de Carvalho Dias (B.Sc. and Ph.D.) is a full member of the Brazilian Academy of Science (ABC in Portuguese) and also at the National Academy of Engineering (ANE in Portuguese). He was rewarded due to the merits achieved during his academic carreer with the Grand Cross Medal – National Order of Scientific Merit – by the Presidency of the Republic, Brazil. Prof. Lindolpho de Carvalho Dias curriculum can be accessed at http://www.abc.org.br/sjbic/curriculo.asp?consulta=lindolpho.

5. Expected Benefits

5.1. Scientific and Technical Contributions

5.1.1. Current Situation of the Web

An assessment of the current situation of the Web is tantamount to the very reasons for creating a Web Science.

With 1.5 billion users, the Web is now central to our daily lives and of high economic, social and cultural significance, having moreover become one of the primary mediums for scientific collaboration. **Yet we know relatively little about why the Web is so successful or how it works**. We need to understand it better, to anticipate future developments, and to identify opportunities and threats.

Current engineering-based approaches have reached a limit, while other avenues, such as complex systems science, have failed to provide the insights expected. **Although substantial data is available on Web developments and trends, we need to understand in a much clearer way what the data is telling us**. Web Science, or the 'Science of the Web', offers a means of studying the Web as a decentralized information system with its own intrinsic framework, rules and laws.

The central challenge of this new science is to understand emergence in the context of the Web. How and why do simple principles and behaviors lead to complex structures and phenomena, and how can we anticipate them? In particular, Web Science research aims at providing insights on networked information and knowledge – i.e. **how people relate to and use information in this kind of group context and how consensus (or 'social wisdom') emerges from individual behaviors**.

Web Science draws on a wide range of disciplines, from computer science and engineering, the physical and mathematical sciences, to social sciences, the arts and policy-making. Like any science, it must be developed around hypotheses that can be clearly articulated, tested and verified. **The foundations for Web Science will emerge through interaction between the many different contributing disciplines**. It is essential to establish a dialogue between these disciplines based on a common language and metaphors.

Web Science will lead to a new kind of Web. This new scientific domain aims both at understanding the Web and at focusing on its development so as to better meet human needs. The benefits will be felt in all areas of science, business and government, and by individual users in their daily lives.

5.1.2. Contributions to the Development of Web Science

The activities under the umbrella of the institute will contribute to the development of Web Science in a variety of ways, as detailed in Section 3. Following the same classification used throughout the proposal, we may summarize the contributions as follows.

Under “People & Society”, the institute will contribute to a better understanding of how people interact with Web resources, what are the dynamics of social networks, and what are the psychological impacts of the Web on individuals. In particular, the institute will address support for users with special needs. It will also contribute to understanding how the Web is used in politics, and how to leverage on economic data extracted from the Web.

The “Software Technologies for Web Applications” layer focus on understanding the Web as a decentralized information system. The institute will create new software techniques to develop Web-wide applications, involving hundreds of thousands of independent processes, and to create Web-based collaborative virtual worlds.

The efforts under “Management of Web Data” will address the crucial questions of searching Web data and organizing Web content. It will propose tools and techniques to manage multimedia Web data, to better access databases available throughout the Web, to endow Web pages with semantics, and to improve searching the Web.

Activities under “Web Infrastructure” will investigate the design and deployment of new network architectures that will support the Web of the future. They will also contribute to deploying Web applications on mobile platforms.

Finally, Web Science needs foundational work. The institute will contribute to the investigation of properties of the Web graph, to the development of new combinatorial optimization methods, involving the Web graph, and to the investigation of metaheuristics in connection with Web mathematics.

In summary, research at the institute is expected to provide a broad range of contributions to the development of Web Science. The contributions range from understanding the impact the Web has on the daily lives of individuals to meeting the challenges of the Web graph. They cover the problems of developing software for Web-wide applications, of searching, retrieving and managing data stored in hundreds of millions of Web sites, and of proposing novel architectures that overcome the limitations of the current Web infrastructure.

5.2. Transfer to Society – social, technological and economic impacts

The institute will adopt several channels to transfer research results and technologies to the industry and service sectors:

- IG - Gênesis Institute at PUC-Rio: develops entrepreneurship programs and houses start-*up* projects in the context of the PUC-Rio community.
- RNP – Rede Nacional de Ensino e Pesquisa: integrates more than 300 Brazilian teaching and research institutions, reaching over a million users.

IG will be one of the channels to communicate results to the industrial and services sectors. It is expected that IG will leverage on the interaction with the institute to decide future lines of action to foster new companies that adopt new Web technologies.

RNP is the Brazilian NREN (National Research and Education Network). As such, one of its key roles is to promote the development of network communication

technologies, including innovative services and applications, especially those that can benefit its user organizations.

One of the mechanisms that RNP uses for this promotion is the Working Group (WG) program. To promote interaction between research groups and the set of services it offers, which can be connectivity-oriented or, more generally, network-related, RNP created in 2002 a program for financing working groups (GT-RNP). This program supports collaborative projects that aim to demonstrate the feasibility of using new protocols, network services and applications, which can help RNP enlarge the range of services it offers.

Working groups are formed by researchers in private or public institutions, selected by RNP through annual calls for projects. Both companies and universities may participate in these projects, provided that the criteria established by the working group coordinator are satisfied (The list of working groups that have been supported at some point since 2002 is available at http://www.rnp.br/pd/gt).

The working group model furnishes an important channel for transferring research results to society. Experience with services such as VoIP (voice over IP) and videoconference has shown that, with well defined deliverables and close follow-up by RNP staff, it is possible, starting out from research results and experience, to create new services that are useful to the entire research and education community. Some of the projects that originated in working groups are bearing fruits beyond this community. A good example is the ICP-Edu (Public Key Infrastructure for Education) Working Group. Its coordinators are actively participating in the definition of the architecture of ICP-Brasil (the official Brazilian Public Key Infrastructure).

There are also cases in which the technology developed by working groups is being transferred to society through commercial products and/or new companies. The hardware designed to manage private keys and accelerate cryptographic operations in ICP-Edu is now the recommended hardware for ICP-Brasil. Another example is the creation of Dynavideo, a company set up to further develop and market projects initially developed by the LAVID (Digital Video Lab), at UFPB, many of them in RNP's working groups program.

RNP's responsibilities also include supporting some actions of the Science and Technology Ministry (MCT). Specifically, in 2008 MCT set up a new program for supporting technological development, called SIBRATEC, with eleven target areas as priorities, one of which is the area of digital information and communication technologies (TICs, in Portuguese). RNP is incubating a virtual organization, called CTIC, which concentrates the activity of SIBRATEC in this area. The role of CTIC is to act as an innovation center, encouraging the development of new technologies by research institution and the use of these technologies by existing and new companies. CTIC must also promote the collaboration among research and commercial institutions.

With CTIC under its responsibility, RNP must thus promote cooperation and the development of products that stem from research and development activity in research institutions, not only in the area of computer networks, but also in the wider range of TICs. Currently, CTIC is selecting a first set of projects from a set of proposals

submitted in response to a call for projects in specific areas that are considered relevant for the effective dissemination of the Brazilian Digital TV System.

Another one of RNP's roles is to provide resources for communication among its client organizations, so as to support experiments in communication and in distributed computing. Experiments that require conventional Internet resources may be carried out using the connectivity services available for all users. However, when non-conventional resources, such as, for instance, high capacity, are needed, a specific communications infrastructure may be required. The GIGA project (2003 to 2008), coordinated by RNP and CPqD, set up high capacity testbed network, interconnecting research laboratories in 20 institutions in the states of Rio de Janeiro and São Paulo. There are plans to extend this experimental network so as to extend this testbed to a larger number of institutions in the near future. In these testbed networks, it is and will continue to be possible to carry out research on new network architectures. This support is needed if new architectures for the Internet and for the Web are to be investigated and validated. The RNP support team is the following:

Name	**Affiliation**	**Degree**
Ana Lúcia de Moura	RNP	D.Sc., PUC-Rio, 2004
André Luíz Almeida Marins	RNP	M.Sc., PUC-Rio, 2008
Daniela Francisco Brauner	RNP	D.Sc., PUC-Rio, 2008

5.3. Education and Training

The institute will be the focal point of the academic institutions involved to create new course offerings at various levels:

- At the undergraduate level:

  The institute will promote the offering of Web Science as a Computer Science minor domain with about 8-10 disciplines, or a total of 450 hours. Specifically, the disciplines currently offered at the participating academic institutions will be complemented with new disciplines, covering roughly the research lines proposed:

  - Foundations of Web Science
  - Human-Computer Interaction on the Web
  - Social Impacts of the Web
  - Web Application Environments
  - Knowledge Discovery on the Web
  - The Deep Web and the Semantic Web
  - Multimedia Data on the Web
  - Distributed Computing on the Web
  - Security and Resilience of the Web

- At the graduate level:

  The institute will promote the offering of graduate disciplines that cover novel aspects of the Web Science. Along the 5 years of the project, it is expected that the participating academic institutions will offer new disciplines, covering the research lines proposed (see above) or specific topics of interested, among those listed in Section 3. Needless to say that theses and dissertations will emerge in the context of the research carried out at the institute.

- Continuing Education and Professional Development in Web Science:

  The institute will promote the offering of extension courses, in the spirit of lifelong learning, to recycle IT professionals and the general public in topics related to Web Science. The minor domain, proposed at the undergraduate level, will be repackaged as a full 450 hour extension course, with a 50 hour final project. Specific topics, such as Social Networks, will generate short courses to the general public.

The course offerings listed above will be offered within the presential framework, traditionally adopted by the participating academic institutions, as well as offered non-presentially within the context of the Brazilian Open University (http://uab.capes.gov.br).

The institute will also offer the following specific programs:

- To the academic institutions:

  The institute will offer a post-doctoral program, at the participating academic institutions, whereby professionals with doctoral degrees, especially university professors, will be able to participate in the research developed at the institute.

  The institute will also offer a summer program to students and university professors alike whereby they will be able to participate in the research developed at the institute.

- To secondary schools:

  The institute will offer hands-on short courses to motivate students to proceed to an undergraduate course in Computer Science with heavy Web Science content.

  The institute will also offer a summer program to high school to put them in contact with the research developed at the institute and thereby motivate their students.

5.4. Patents, Prototypes and Technological Products

The deliverable of the research lines, detailed in Section 3, have potential to generate patents, prototypes and technological products, as summarized in Table 3. The major universities involved in the project – UFRJ, PUC-Rio and UNICAMP – have specific offices that will help file patents and protect the intellectual property of the Institute, as decided by the Steering Committee.

As for the technological transfer mechanisms, the institute will offer:

- a residency program whereby IT professionals will be immersed in one of the laboratories of the participating institutions for a hands-on, direct contact with the Web Science concepts, tools and technology developed
- conventional and Web-based educational programs
- a Web site to distribute methods and tools to be publicly available under the appropriate licensing agreement, if so decided by Steering Committee.

Table 2. Summary of prototypes and technological products amenable to patenting.

People & Society	
Interaction Design	tools to support the development of interaction mechanisms non-speech sound interfaces
Social Networks	methods for social networks analysis and mining methods for knowledge discovery and organization from narrative social content
Software Technologies for Web Applications	
A Multi-Agent Systems Approach for Developing Autonomic Web Applications	a framework to support the development of autonomic Web applications
Model-driven Design and Implementation of Web Applications	a model-driven framework to support specification, design and implementation of Web applications, seen as part of men-machine teams
Security and Resilience of Web Applications	static analysis techniques supporting development approaches for detecting possible threats to application security
Web-based Collaborative Virtual Environments	a framework for Web-based collaborative virtual worlds
Management of Web Data	
Managing Web Multimedia Data	new multimedia data descriptors and of storage structures to support their indexing, reuse and composition on the Web
Accessing the Deep Web	heuristics for database identification, mediated schema definition and schema matching a query mediator framework
Developing the Semantic Web	a method for the description, discovery and automated composition of semantic Web services a framework to support the development of dynamic and resilient Web applications a model to evaluate, and personalize search results according to users' quality level perspectives
Knowledge Discovery on the Web	learning algorithms for classification and regression problems machine learning frameworks with EPM implementation for large volumes of data
Web Infrastructure	
Ubiquitous and Location-aware Web	a framework for the development of ubiquitous and location-aware applications

References

ADLER, M., AND MITZENMACHER, M., Towards Compressing Web Graphs, Proceedings of the Data Compression Conference (DCC '01), 2001, 203-212, IEEE Computer Society, washington, DC, US}, 2001.

AVIZIENIS, A., RANDELL, B., AND LANDWEHR, C., Basic Concepts and Taxonomy of Dependable and Secure Computing, IEEE Transactions on Dependable and Secure Computing, 1(1), January-March 2004.

AVIZIENIS, A., Toward Systematic Design of Fault-Tolerant Systems, IEEE Computer, 30(4), p. 51-58., 1997.

BARNES, J. A., 1987, "Redes Sociais e Processo Político", In: FELDMAN-BIANCO, B (org.), Antropologia das Sociedades Contemporâneas – Métodos. São Paulo: Global, 1987, pp. 159-194.

BERNERS-LEE, T., HALL, W., HENDLER, J.A., O'HARA, K., SHADBOLT, N., WEITZNER, J. (2006) A Framework for Web Science., 1, pp.1-130. Now Publisher.

BOLLOBÁS, B., KOHAYAKAWA, Y., RÖDL, V., SCHACHT, M. TARAZ, A., Essentially infinite colourings of hypergraphs Proceedings of the London Mathematical Society Advance Access published on July 19, 2007 Proc. London Math. Soc. 2007 95: 709-734; doi:10.1112/plms/pdm024

BRAUNER, D. F.; CASANOVA, M. A.; LUCENA, C. J. P. (2004) Using Ontologies as Artifacts to Enable Databases Interoperability. In: Proc. ACM Conf. on bject-Oriented Programming, systems, Languages and Applications, Vancouver, BC, Canada.

BRAUNER, D.F.; CASANOVA, M.A.; LUCENA, C.J.P. (2004) Geo-Object Catalogs to Enable Geographic Databases Interoperability. In: Proc. VI Brazilian Symposium on GeoInformatics. São José dos Campos, SP: INPE, 235-246.

BRAUNER, Daniela F, INTRATOR, C., C.FREITAS, J., CASANOVA, M. A. An Instance-based Approach for Matching Export Schemas of Geographical Database Web Services In: IX Brazilian Symposium on GeoInformatics, 2007, Campos do Jordão.

BRAUNER, Daniela F., CASANOVA, M. A., MILIDIU, Ruy. Towards Gazetteer Integration Through an Instance-based Thesauri Mapping Approach In: Advances in Geoinformatics ed.Heidelberg : Springer, 2007, p. 235-245.

BRAUNER, Daniela F., GAZOLA, A., CASANOVA, M. A., BREITMAN, Karin K. ADAPTATIVE MATCHING OF DATABASE WEB SERVICES EXPORT SCHEMAS In: Proceedings of ICEIS 2008 – Tenth International Conference on Enterprise Information Systems, 2008.

BREITMAN, Karin K, BRAUNER, D.F., CASANOVA, M. A., MILIDIU, Ruy, GAZOLA, A., PERAZOLO, M. Instance-Based Ontology Mapping In: 5th IEEE Workshop on Engineering of Autonomic and Autonomous Systems, 2008, Belfast.

BRITTENHAM, P. ET AL., IT service management architecture and autonomic computing, June 27, 2007, Disponível em http://www.research.ibm.com/journal/sj/463/brittenham.html

BROWN, A.B.; PATTERSON, D.A.; “To Err is Human”; First Workshop on Evaluating and Architecting System Dependability, Goteborg, Sweden, 2001; URL: http://roc.cs.berkeley.edu/papers/easy01.pdf

BUTZ, A., HOLLERER, T. et al. 1999. Enveloping users and computers in a collaborative 3d augmented reality. In IWAR 99, 35–44.

CAPPELLI, C., LEITE, http://ceur-ws.org/Vol-322/paper13.pdfJ.C.S.P., OLIVEIRA, A.P.A. "Exploring Business Process Transparency Concepts," *Requirements Engineering Conference, 2007. RE '07. 15th IEEE International* , vol., no., pp.389-390, 15-19 Oct. 2007

CASANOVA, M. A., BREITMAN, Karin Koogan, BRAUNER, Daniela F., Marins, A. Database Conceptual Schema Matching. *Computer* (Long Beach). , v.40, p.102 - 104, 2007.

CASATI, F. and SHAN, M.-C. “Dynamic and adaptive composition of e-services,” Information Systems, vol. 26, pp. 143–163, May 2001.

CLEMENTS, P. and NORTHROP, L. (2002). Software Product Lines: Practices and Patterns. Addison-Wesley, Boston, USA.

CZARNECKI, K. and HELSEN, S. (2006). Feature-based survey of model transformation approaches. IBM Systems Journal, 45(3):621–645.

DE SOUZA, C.S. (2005) The Semiotic Engineering of Human-Computer Interaction. The MIT Press.

DE SOUZA, C.S. (2008) Semiotic engineering in practice: redesigning the CoScripter interface. In Proceedings of AVI 2008, pp. 165-172.

DeBRESSON, C. & AMESSE, F. (1991) Network of innovators: A review and introduction to the issue. Research Policy, no 20, pp. 363-379.

FEINER, S., SHAMASH, A. 1991. Hybrid user interfaces: Breeding virtually bigger interfaces for physically smaller computers. In ACM UIST 91, ACM, 9–17.

FISCHER, L., editor. Workflow Handbook 2001, Workflow Management Coalition. Future Strategies, Lighthouse Point, Florida, 2001.

FOX, A.; “Toward Recovery-Oriented Computing”; Invited talk; VLDB 2002; URL: www.cs.ust.hk/vldb2002/VLDB2002-proceedings/papers/S25P01.pdf

FREEMAN, C. (1991) Network of innovators: A synthesis of research issues. Research Policy, no 20, pp. 499-514.

GOEBBELS, G., LALIOTI, V., GÖBEL, M. Design and Evaluation of Team Work in Distributed Collaborative Virtual Environments. In The Proc. of the ACM Symposium on Virtual Reality Software and Technology, VRST’03, Osaka, Japan, 2003.

GREENFIELD, J., SHORT, K.: “Software Factories: Assembling Applications with Patterns, Frameworks, Models and Tools”, John Wiley and Sons, 2005.

HARMON, P. Business Process Change: 2003. A Manager’s Guide to Improving , Redesigning and Automating Processes. Morgan Kaufmann. London, 2001.

HENDLER, J., SHADBOLT, N., HALL, W., BERNERS-LEE, T., WEITZNER, D. Web Science: An Interdisciplinary Approach to Understanding the Web. Communications of the ACM, 51, pp. 60-69, 2008.

HENNESSY, J. (2002). Back to the Future: Time to Return to Some Long-Standing Problems in Computer Science. Almaden Institute, IBM Almaden Research Center, San Jose, CA, April 10, 2002

HENZINGER, M., Algorithmic Challenges in Web Search Engines, Internet Mathematics, 1, 115-126, 2003.

HENZINGER, M., Combinatorial algorithms for Web search engines: three success stories, Symposium on Discrete Algorithms Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms New Orleans, Louisiana Pages: 1022 - 1026, 2007.

HOUSTIS, E. N., GALLOPOULOS, E., BRAMLEY, R., RICE, J. R. Problem-Solving Environments for Computational Science. IEEE Computational Science and Engineering, 4(3): 18-21, 1997.

HOWARD, M., AND LEBLANC. D., Writing Secure Code. Microsoft Press, 2001.

HULME. G., New software may improve application security. http://www.informationweek.com/story/ IWK20010209S0003, 2001.

HUMPHREY, W.S.; “The Software Quality Challenge”; Crosstalk The Journal of Defense Software Engineering 21(6); 2008; pags 4-9; URL: http://www.stsc.hill.af.mil/crosstalk/

IBM (2003). Autonomic Computing – Creating Self-managing Computing Systems. Disponível em http://www-3.ibm.com/autonomic/index.shtml.

JABLONSKI, S. & BUSSLER, C. Workflow Management: Modeling Concepts, Architecture, and Implementation. International Thomson Computer Press, London, UK, 1996.

JENNINGS, N. R., An agent-based approach for building complex software systems, Communications of the ACM, v.44 n.4, p.35-41, April 2001.

KEPHART, JEFFREY AND CHESS, DAVID M. (2003) The Vision of Autonomic Computing. IEEE Computer Magazine 36(1): 41-50.

KUMAR R., RAGHAVAN, P. RAJAGOPALAN, S., AND TOMKINS, A., Trawling the Web for emerging cyber-communities, Comput. Netw., 31, 11-16, 1999, http://dx.doi.org/10.1016/S1389-1286(99)00040-7, Elsevier North-Holland, Inc., New York, NY, USA.

LARUS, J.R.; BALL, T.; DAS, M.; DELINE, R.; FÄHNDRICH, M.; PINCUS, J.; RAJAMANI, S.K.; VENKATAPATHY, R.; “Righting Software”; IEEE Software 21(3); 2004, p 92-100.

LEHMAN, M.M. "Programs, life cycles, and laws of software evolution", Proceedings of the IEEE, pages 1060–1076, September 1980.

LEITE, J.C.S.P., CAPPELLI, C., Exploring i* Characteristics that Support Software Transparency. iStar 2008: 51-54

LEITE, J.C.S.P., WERNECK, V.M.B, OLIVEIRA, A.P.A., CAPPELLI, C., CERQUEIRA, A.L.A., CUNHA, H.S., GONZALEZ-BAIXAULI, B. Understanding the Strategic Actor Diagram: an Exercise of Meta Modeling.. Anais do WER07 - Workshop em Engenharia de Requisitos, Toronto, Canada, May 17-18, 2007, pp 2-12 (http://www.inf.puc-rio.br/~wer/WERpapers/artigos/artigos_WER07/Cwer07-leite01.pdf)

LESHED, G., HABER, E.M., MATTEUWS, T., LAU, T.A. "CoScripter: Automating & Sharing How-To Knowledge in the Enterprise". Proceedings of CHI'08, 2008.

LEYMANN, F. & ROLLER, D. Production Workflow: Concepts and Techniques. Prentice-Hall PTR, Upper Saddle River, New Jersey, USA, 1999.

LI, G, HAN, Y., ZHAO, Z., WANG, J., and WAGNER, R. M., "Facilitating dynamic service compositions by adaptable service connectors," International Journal of Web Services Research, vol. 3, pp. 68–84, January 2006.

LIVSHITS, V.B.; LAM, M.S.; "Finding Security Vulnerabilities in Java Applications with Static Analysis"; Proceedings of the 14th Conference on USENIX Security Symposium; 2005; available at URL: suif.stanford.edu/papers/usenixsec05.pdf

MAGALHÃES, J.A.P.; STAA, A.V.; LUCENA, C.J.P.; "Evaluating the Recovery Oriented Approach through the Systematic Development of Real Complex Applications"; Software Practice and Experience; New York: Wiley; to appear

MEDEIROS, C.B. (2008) Grand Research Challenges in Computer Science in Brazil. IEEE Computer, 41, 6, pp. 59-65.

MILANOVIC, N. and MALEK, M., "Current solutions for Web service composition," IEEE Internet Computer, vol. 8, pp. 51–59, November 2004.

MYLOPOULOS, J., CHUNG, L., & YU, E. 1999. From object-oriented to goal-oriented requirements analysis. *Commun. ACM* 42, 1 (Jan. 1999), 31-37. DOI= http://doi.acm.org/10.1145/291469.293165.

O'HARA, K., HALL, W. Web Science, Association of Learning Technologies Newsletter: Issue 12, May 2008

PATTERSON, D., BROWN, A., BROADWELL, P., CANDEA, G., CHEN, M., CUTLER, J., ENRIQUEZ, P., FOX, A., KICIMAN, E., MERZBACHER, M., OPPENHEIMER, D., SASTRY, N., TETZLAFF, W., TRAUPMAN, J., AND TREUHAFT, N., Recovery Oriented Computing (ROC): Motivation, Definition, Techniques, and Case Studies. Technical Report UCB/CSD-02-1175, Computer Science Department, University of California Berkeley; 2002; Available at: URL: http://roc.cs.berkeley.edu/papers/ROC_TR02-1175.pdf

PITAC - Presidents Information Technology Advisory Committee; Information Technology Research: Investing in our Future; Report to the President; 1999; (page 4, Priorities for Research)

POHL, K., et al. (2005). Software Product Line Engineering: Foundations, Principles and Techniques. Springer-Verlag, New York, USA.

RAPOSO, A. B., CRUZ, A. J. A., ADRIANO, C. M., MAGALHÃES, L. P. Coordination components for collaborative virtual environments. Computers And Graphics, 25(6), p. 1025-1039. 2001.

SILVA, L.F.: Uma Estratégia Orientada a Aspectos para Modelagem de Requisitos, Tese de Doutorado, Março de 2006, Departamento de Informática, PUC-Rio.

STAA, A.v. Programação Modular. Editora Campus. 2000.

STAA, A.v.: Engenharia de Software Fidedigno. In: PUC-Rio (ed.): Monografias em Ciência da Computação, n. 13/06. Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2006.

STAHL, T., VOELTER, M.: "Model-Driven Software Development: Technology, Engineering, Management", Wiley, 2006.

STEFANI, O., HOFFMANN, H., RAUSCHENBACH, J. Design of Interaction Devices for Optical Tracking in Immersive Environments. In: Human-Centred Computing: Cognitive, Social and Ergonomic Aspects, Volume 3 of Proc. HCI International. 2003.

TESAURO G. , CHESS D. M., WALSH E. W , DAS R. , SEGAL A. , WHALLEY I. , KEPHART J. O. , WHITE S. R., A Multi-Agent Systems Approach to Autonomic Computing, Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems, p.464-471, July 19-23, 2004, New York, New York

THOMAS, D.; "The Deplorable State of Class Libraries"; Journal of Object Technology 1(1); Zürich, CH: ETH Zürich; 2002; pp. 21-27

VAN DER AALST, W.M.P, DESEL, J., & OBERWEIS, A. (eds.) Business Process Management: Models, Techniques, and Empirical Studies, volume 1806 of Lecture Notes in Computer Science. Springer-Verlag, Berlin, 2000.

VAN DER AALST, W.M.P. & VAN HEE, K.M. Workflow Management: Models, Methods, and Systems. MIT press, Cambridge, MA, 2002.

WEIJTERS, A.J.M.M. & VAN DER AALST, W.M.P.. Workflow Mining: Discovering Workflow Models from Event-Based Data. In C. Dousson, F. Höppner, and R. Quiniou, editors, Proceedings of the ECAI Workshop on Knowledge Discovery and Spatial Data, pages 78-84, 2002.

WEITZNER, D. J., ABELSON, H., BERNERS-LEE, T., FEIGENBAUM, J., HENDLER, J., and SUSSMAN, G. J. 2008. Information accountability. *Commun. ACM* 51, 6 (Jun. 2008), 82-87. DOI= http://doi.acm.org/10.1145/1349026.1349043

Annex - Project Team

Name	Affiliation	Degree	CNPq level	Research Subtopics
Adriana S. Vivacqua	ADDLabs/UFF	D.Sc., COPPE/UFRJ, 2007	—	Social Networks
Anne Marie Delaunay Maculan	Eng. Prod./COPPE	Ph.D. , U. Quebec a Montreal, 1989		People and Society
Alberto Barbosa Raposo	PUC-Rio	D.Sc., UNICAMP, 2000	–	Interaction Design; Web-based Collaborative Virtual Environments
Ana Maria Nicolaci da Costa	PUC-Rio	PhD, U. of London, 1983	1C	Web and the Individual
Angelo Ernani Maia Ciarlini	UNIRIO	D.Sc., PUC-Rio, 1999	—	Web and the Economy; Web-based Collaborative Virtual Environments
Antônio Jorge Gomes Abelém	UFPA	Dr, PUC-Rio, 2003	–	Future Internet Architectures
Antonio Luz Furtado	PUC-Rio	PhD, U. Toronto, 1974	—	Web-based Collaborative Virtual Environments; Accessing the Deep Web; Developing the Semantic Web
Arndt von Staa	PUC-Rio	PhD, Waterloo, 1974	–	Security & Resilience of Web Applications
Bruno Feijó	PUC-Rio	PhD, U. of London, Imperial College, England, 1988	1C	Web-based Collaborative Virtual Environments
Carlos Alberto de Jesus Martinhon	UFF	D. Sc. COPPE Sistemas e Computação UFRJ (1998)	–	Foundations of Web Science
Carlos Eduardo Fisch de Brito	UFC	Ph. D. Computer Science pela University of California - Los Angeles (2003)	–	Foundations of Web Science
Carlos José Pereira de Lucena	PUC-Rio	PhD, UCLA, 1974	1A	Multi-agent Systems & Autonomic Web Applications; Model-driven Design and Implementation of Web Applications
Cecília Mary Fischer Rubira	IC-UNICAMP	PhD, Newcastle University, UK,1994	1D	Developing the Semantic Web
Cesar Tadeu Pozzer	UFSM	D.Sc., PUC-Rio, 2005	—	Web-based Collaborative Virtual Environments
Clarisse Sieckenius de Souza	PUC-Rio	Dr., PUC-Rio, 1987	1C	Interaction Design

Claudia Linhares Sales	UFC	Docteur Informatique - Recherche Operationnelle - Université de Grenoble I (Scientifique Et Medicale - Joseph Fourier) (1996).	2	Foundations of Web Science
Claudia Maria Bauzer Medeiros	IC-UNICAMP	PhD, U. Waterloo, 1985	1B	Interaction Design ; Managing Web Multimedia Data
Claudia Maria Lima Werner	COPPE	D.Sc., COPPE/UFRJ	1D	
Claudio Leonardo Lucchesi	UNICAMP	Ph.D. Computer Science - University of Waterloo (1976)	1A	Foundations of Web Science
Claudson Ferreira Borstein	UFRJ	Ph. D. Computer Science pela Carnegie Mellon University (1998)	2	Foundations of Web Science
Daniel Schwabe	PUC-Rio	PhD, UCLA,	1B	Interaction Design; Model-driven Design and Implementation of Web Applications
Edson Norberto Cáceres	UFMS	D. Sc. COPPE Sistemas e Computação UFRJ (1992)	2	Foundations of Web Science
Fábio Protti	UFRJ	D. Sc. COPPE Sistemas e Computação UFRJ (1998)	1D	Foundations of Web Science
Fernanda Araujo Baião Amorim	UNIRIO	D.Sc., COPPE-UFRJ, 2001	—	Social Networks
Flávia Coimbra Delicato	UFRN	D.Sc. UFRJ, 2005	–	Multi-agent Systems & Autonomic Web Applications; Model-driven Design and Implementation of Web Applications; Developing the Semantic Web
Flávia Maria Santoro	UNIRIO	D.Sc., COPPE-UFRJ, 2001	—	Social Networks
Francisco Heron de Carvalho Junior	UFC	D. Sc. Ciências da Computação pela Universidade Federal de Pernambuco (2003	–	Foundations of Web Science
Geraldo Bonorino Xexéo	PESC/COPPE	D.Sc., COPPE, 1994	—	Social Networks; Design and Implementation of Autonomic Workflows; Developing the Semantic Web
Hugo Fuks	PUC-Rio	PhD, Univ. of London, Imperial College, 1991	1D	Interaction Design; Web-based Collaborative Virtual Environments

Iara Machado	RNP	M., UFF, 1998	–	Future Internet Architectures
Jacques Wainer	IC-Unicamp	PhD, Pennsylvania State U. - 1991	1D	Web and Politics
Jair Cavalcanti Leite	UFRN	D.Sc., PUC-Rio, 1998	—	Interaction Design
Jano Moreira de Souza	PESC/COPPE	Ph.D., University of East Anglia, 1986	1D	Social Networks; Design and Implementation of Autonomic Workflows; Developing the Semantic Web
Jayme Luiz Szwarcfiter	UFRJ	Ph. D. Ciência da Computação, pela University of Newcastle Upon Tyne, Inglaterra (1975)	1A	Foundations of Web Science
João Magalhães	PUC-Rio	PhD candidate, PUC-Rio, to conclude in 2009	–	Security & Resilience of Web Applications
Jonice de Oliveira Sampaio	UERJ	D.Sc., COPPE, 2007	—	Social Networks; Design and Implementation of Autonomic Workflows
José Augusto Suruagy Monteiro	UNIFACS	PhD, UCLA, 1990	–	Future Internet Architectures
Julio Cesar Sampaio do Prado Leite	PUC-Rio	PhD, University of California at Irvine, 1988	1C	Social Networks
Karin Koogan Breitman	PUC-Rio	D.Sc., PUC-Rio, 2000	—	Accessing the Deep Web; Developing the Semantic Web
Kelvin Lopes Dias	UFPA	Dr., UFPE, 2001	–	Future Internet Architectures
Leonardo Bandeira Rezende	PUC-Rio	PhD, Stanford University, 2003	—	Web and the Economy
Loana Tito Nogueira	UFF	D. Sc. COPPE Sistemas e Computação UFRJ (2003)	2	Foundations of Web Science
Luerbio Faria	UERJ	D. Sc. COPPE Sistemas e Computação UFRJ (1998)	2	Foundations of Web Science
Luis Alfredo Vidal de Carvalho	UFRJ	D.Sc. COPPE (1989)	1C	Knowledge Discovery on the Web
Luis Antonio Brasil Kowada	UFF	D. Sc. COPPE Sistemas e Computação UFRJ (2006)	–	Foundations of Web Science
Luíz Claúdio Schara Magalhães	UFF	PhD, UIUC, 2002	–	Future Internet Architectures

Manoel Bezerra Campêlo Neto	UFC	D. Sc. COPPE Sistemas e Computação UFRJ (1999)	2	Foundations of Web Science
Marcelo Gattass	PUC-Rio	PhD, Harvard University, 1979	1A	Interaction Design; Web-based Collaborative Virtual Environments
Márcia Rosana Cerioli	UFRJ	D. Sc. COPPE Sistemas e Computação UFRJ (1999)	2	Foundations of Web Science
Márcio de Oliveira Barros	UNIRIO	D.Sc., COPPE-UFRJ, 2001	2	Web and the Economy
Marco Antonio Casanova	PUC-Rio	PhD, Harvard U., 1979	1C	Accessing the Deep Web; Developing the Semantic Web
Maria Cecilia Calani Baranauskas	UNICAMP	Dr., UNICAMP 1993	—	Interaction Design
Markus Endler	PUC-Rio	Dr. Rer. nat. TU Berlin, 1992	2	Ubiquitous and Location-aware Web
Michael Anthony Stanton	RNP	PhD, U. Cambridge, 1971	–	Future Internet Architectures
Michael G. Hinchey	LERO	PhD, U. Cambridge	—	Accessing the Deep Web; Developing the Semantic Web
Mitre Costa Dourado	UFRJ	D. Sc. COPPE Sistemas e Computação UFRJ (2005)	2	Foundations of Web Science
Noemi Costa dos Santos	UERJ	D. Sc. COPPE Sistemas e Computação UFRJ (2006)	–	Foundations of Web Science
Noemi Rodriguez	PUC-Rio	D.Sc. PUC-Rio, 1993	2	Ubiquitous and Location-aware Web
Paulo de Figueiredo Pires	UFRN	D.Sc. UFRJ, 2002	–	Multi-agent Systems & Autonomic Web Applications; Model-driven Design and Implementation of Web Applications; Developing the Semantic Web
Paulo Eustáquio Duarte Pinto	UERJ	D. Sc. COPPE Sistemas e Computação UFRJ (2006)	–	Foundations of Web Science
Rafael Castro de Andrade	UFC	Docteur Université de Paris XIII (Paris-Nord) (2002)	2	Foundations of Web Science
Renata Mendes de Araujo	UNIRIO	D.Sc., COPPE-UFRJ, 2000	—	Social Networks

Renato Cerqueira	PUC-Rio	D.Sc., PUC-Rio, 2000	–	Multi-agent Systems & Autonomic Web Applications; Model-driven Design and Implementation of Web Applications; Ubiquitous and Location-aware Web
Ricardo Cordeiro Corrêa	UFC	Docteur Informática - Institut National Polytechnique de Grenoble (1997)	–	Foundations of Web Science
Ricardo da Silva Torres	UNICAMP	Dr., UNICAMP 2004	2	Interaction Design; Managing Web Multimedia Data
Ricardo de Oliveira Anido	IC-UNICAMP	PhD, Imperial College, 1989	—	Developing the Semantic Web
Roberta de Souza Coelho	UFRN	D.Sc., PUC-Rio, 2008	–	Model-driven Design and Implementation of Web Applications; Security & Resilience of Web Applications
Rodolfo Jardim de Azevedo	IC- UNICAMP	D.Sc, UNICAMP, 2002	2	Managing Web Multimedia Data
Rubens Nascimento Melo	PUC-Rio	D.Sc., ITA, 1976	—	Accessing the Deep Web; Developing the Semantic Web
Ruy Luiz Milidiú	PUC-Rio	PhD, UC Berkeley, 1985	1C	Knowledge Discovery on the Web
Sean Wolfgand Matsui Siqueira	UNIRIO	D.Sc., PUC-Rio, 2005	—	Web and the Economy; Web-based Collaborative Virtual Environments
Simone Dantas de Souza	UFF	D. Sc. COPPE Sistemas e Computação UFRJ (2002)	2	Foundations of Web Science
Simone Diniz Junqueira Barbosa	PUC-Rio	D.Sc., PUC-Rio, 1999	2	Interaction Design; Web-based Collaborative Virtual Environments
Siome Klein Goldenstein	IC-UNICAMP	PhD. University of Pennsylvania, 2002	2	Managing Web Multimedia Data
Sulamita Klein	UFRJ	D. Sc. COPPE Sistemas e Computação UFRJ (1994)	1C	Foundations of Web Science
Tereza Cristina Melo de Brito Carvalho	USP	Dra, USP, 1996	–	Future Internet Architectures
Thais Vasconcelos Batista	UFRN	D.Sc., PUC-Rio, 2000	–	Multi-agent Systems & Autonomic Web Applications; Model-driven Design and Implementation of Web Applications

Thomas Dwyer	IFCH-Unicamp	Dr., École des Hautes Études en Sciences Sociales, 1978	1C	Web and Politics
Uirá Kulesza	UFRN	D.Sc., PUC-Rio, 2007	–	Multi-agent Systems & Autonomic Web Applications; Model-driven Design and Implementation of Web Applications
Uirá Kulesza	UFRN	D.Sc., PUC-Rio, 2007	–	Security & Resilience of Web Applications
Edmundo A. de Souza e Silva	PESC-UFRJ	Ph.D. UCLA 1984	1A	Modeling the Web and Applications
Rosa M.M. Leão	PESC-UFRJ	D.Sc. Paul Sabatier 1994	-	Modeling the Web and Applications
Daniel R. Figueiredo	PESC-UFRJ	Ph.D. UMASS 2005	-	Modeling the Web and Applications
Valmir Barbosa	PESC-UFRJ	Ph.D. UCLA 1986	1A	Modeling the Web and Applications
Claudia Werner	PESC-UFRJ	D.Sc. COPPE/UFRJ	1D	Modeling the Web and Applications
Vera Lúcia Doyle Louzada de Mattos Dodebei	UNIRIO	D.Sc., ECO-UFRJ, 1997	2	Social Networks
Nelson Maculan Filho	UFRJ	Doutor	1A	Applied and Computational Mathematics/ Optimization
Celso da Cruz Carneiro Ribeiro	UFF	Doutor	1A	Applied and Computational Mathematics/ Optimization
Abílio Pereira de Lucena Filho	UFRJ	Doutor	1D	Applied and Computational Mathematics/ Optimization
Antonio Alberto Fernandes de Oliveira	UFRJ	Doutor	1D	Applied and Computational Mathematics/ Optimization
Cid Carvalho de Souza	UNICAMP	Doutor	1D	Applied and Computational Mathematics/ Optimization
José Ferreira de Rezende	UFRJ	Doutor	1D	Applied and Computational Mathematics/ Optimization
Luiz Satoru Ochi	UFF	Doutor	1D	Applied and Computational Mathematics /Optimization
Susana Scheimberg de Makler	UFRJ	Doutor	1D	
Marcia Helena Costa Fampa	UFRJ	Doutor	2	

Mario Roberto Folhadela Benevides	UFRJ	Doutor	2	
Simone de Lima Martins	UFF	Doutor	2	
Geraldo Galdino de Paula Júnior	UENF	Doutor	--	
Luiz Mariano Paes de Carvalho Filho	UERJ	Doutor	--	
Josefino Cabral Melo Lima	DCC-IM/UFRJ	Docteur, Paris VI, 1992	-	