

Brazilian Institute for Web Science Research¹

Nelson Maculan¹, Edmundo A. de Souza e Silva¹, Geraldo B. Xexéo¹,
Claudia B. Medeiros²,
Simone D.J. Barbosa³, Karin K. Breitman³, Marco A. Casanova³, Carlos José P. de Lucena³

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

² Institute of Computing – University of Campinas, Campinas, Brazil

³ Department of Informatics – PUC-Rio – Rio de Janeiro, Brazil

Summary

This paper 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.

1. Introduction

This paper introduces the **Brazilian Institute for Web Science Research**, conceived as a center of excellence for state-of-the-art, high quality international standard research in Web Science [1, 2]. The Institute is also framed in the context of the Grand Research Challenges in Computer Science in Brazil [4].

The mission of the institute is fourfold:

1. To advance scientific research in a number of topics within Web Science, especially in those topics in which the Institute researchers have already produced high-quality publications, which are available to both the national and international scientific communities;
2. To educate qualified professionals and researchers towards innovation and to 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 internships at the Institute.

The Institute congregates 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

¹ The authors of this paper represent just a small subset of the Brazilian Institute for Web Science Research team, whose contributions they gratefully acknowledge.

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 d’Informatique de Paris VI, France.

The total budget of the Institute is approximately 5M US dollars, over the next five years. CNPq has already awarded² approximately 1.5M US dollars over the next 3 years. The rest of the budget will come from the participating universities and Brazilian state government agencies for the advancement of science.

Web Science was chosen as the focus of the Institute exactly because it provides a common goal to such a large number of researchers, with a wide range of interests. The research proposal submitted to CNPq was written in a bottom-up process: a team of senior researchers selected an initial list of topics from [1, 2]; this list was sent to a number of researchers in Brazil; the latter replied with short research proposals and provided instances of their recent research production in the proposed themes; the proposal team combined the input received into a (hopefully) coherent text; the text was sent back to the researchers involved for feedback. The proposal team centrally decided on the budget proposal, based on commonly agreed guidelines and on the constraints imposed by the call-for-proposals from CNPq. The overall process was run on a very tight schedule and took about 45 days. This short paper is a brief summary of the final proposal, which is about 200 pages long [3].

The rest of this paper is organized as follows. Section 2 outlines the research program. Section 3 describes the mechanisms that will be used to transfer the results to the society. Finally, Section 4 summarizes the expected scientific contributions.

2. Research Program Outline

The research program is organized along five dimensions, depicted in Figure 1 (largely inspired by a similar figure from <http://webscience.org/>). The “People & Society” dimension 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 dimension is intimately related with the fourth Grand Research Challenge identified by the Brazilian Computer Society (“Participative and universal access to knowledge for the Brazilian citizen”) [4].

² See the full list of institutes approved by CNPq at http://www.cnpq.br/programas/inct/_apresentacao/index.html (in Portuguese)

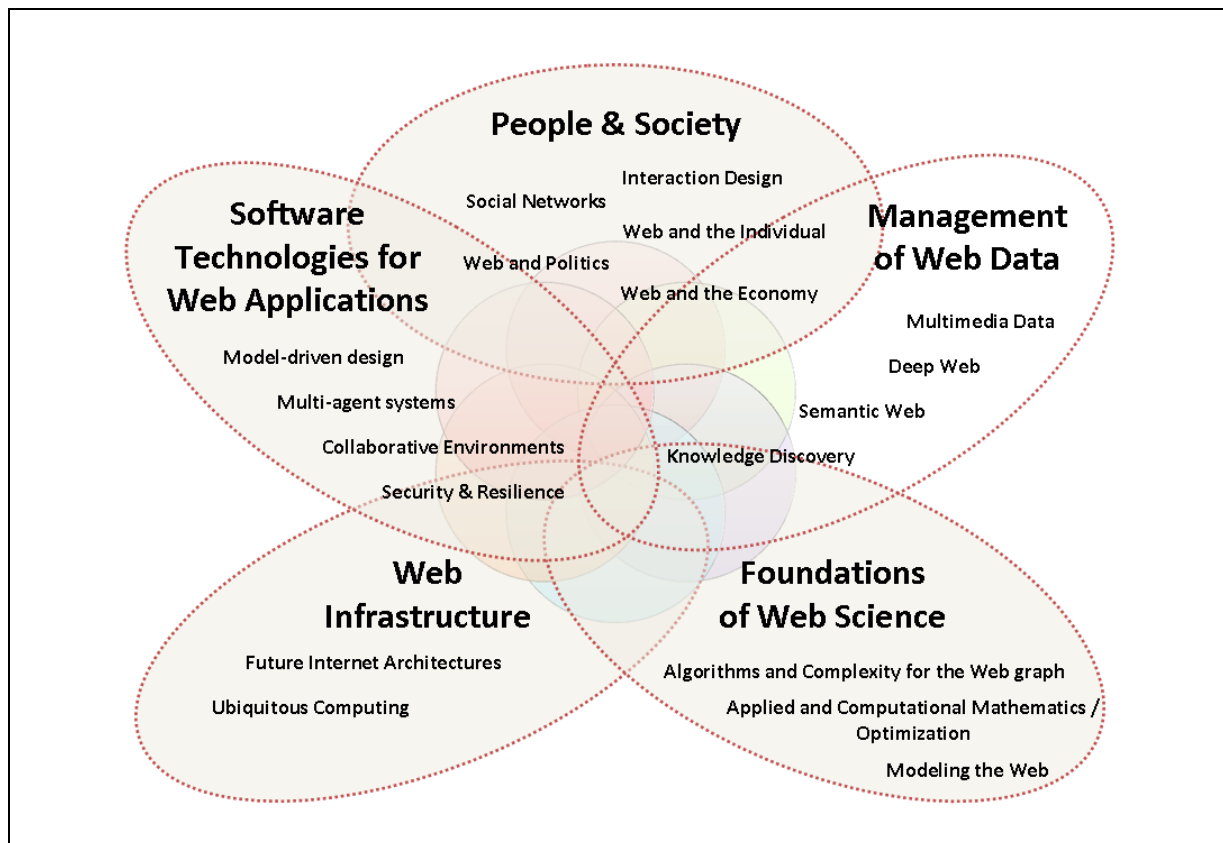


Figure 1. Web Science research program.

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

“Management of Web Data” 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. While present Web knowledge management tools mainly deal with text, new technologies are needed to enhance semantics, and to access multimedia data. This dimension contributes to efforts towards the first Grand Research Challenge (“Management of information over massive volumes of distributed multimedia data”).

The “Web Infrastructure” dimension will primarily deal 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. It will borrow from, and contribute to research in the fifth Grand Research Challenge (“Technological development with quality: dependable, scalable and ubiquitous systems”).

Finally, research on the “Foundations of Web Science” will contribute to optimize the performance of systems that execute on the Web, to create mathematical models of the Web graph, and models that address the complex interactions between Web applications and their users.

3. Transfer to Society

The institute will adopt several channels to transfer research results and technologies to the industry and service sectors, including the Genesis Institute for Innovation and Entrepreneur Action at PUC-Rio, an incubator and technological park, and RNP – the Brazilian Research and Education Network. It is expected that the former will leverage the interaction with the Institute to decide future lines of action to foster new companies that adopt new Web technologies. RNP will be the main channel to disseminate the Institute’s results to the research community.

The Institute will be the focal point of the involved academic institutions to create new course offerings in Web Science at various levels – undergraduate, graduate and continuing education and professional development. The courses will be offered within the framework traditionally adopted by the participating academic institutions, as well as offered as distance education within the context of the Brazilian Open University (Universidade Aberta do Brasil - <http://uab.capes.gov.br>). The Institute will also offer a post-doctoral program at the participating academic institutions, a summer program for students and university professors, and hands-on short courses to motivate high-school students to proceed to an undergraduate course in Computer Science with heavy Web Science content.

The deliverables of the research lines, detailed in Section 2, have potential to generate patents, prototypes and technological products. The major universities involved in the project have specific offices that will help establish file patents and protect the intellectual property of the Institute, as decided by the Institute’s Steering Committee.

Finally, technological transfer mechanisms include an internship program for IT professionals, conventional and Web-based educational programs, and a Web site to publicly distribute methods and tools under appropriate licensing agreements, if so decided by the Steering Committee.

4. Expected Scientific Contributions

Research at the Institute is expected to provide a broad range of contributions to the development of Web Science. Under “People & Society”, it will contribute to a better understanding of how people interact with Web resources, the dynamics of social networks, and

the psychological impacts of the Web on individuals. In particular, the institute will address the problem of providing 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 will 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.

References

- [1] BERNERS-LEE, T., HALL, W., HENDLER, J.A., O’HARA, K., SHADBOLT, N., WEITZNER, J. *A Framework for Web Science (Foundations and Trends(R) in Web Science)*. Now Publishers Inc, 2006.
- [2] 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.
- [3] MACULAN, N.; LUCENA, C. J. P. (eds.). “Brazilian Institute for Web Science Research”. Technical Report 46/08. Department of Informatics, PUC-Rio.
- [4] MEDEIROS, C. M.B. “Grand Research Challenges in Computer Science in Brazil”. *IEEE Computer*, 41, 6, pp. 59-65, 2008.