

**Memoria Anual Núcleo Milenio**  
*Centro de Investigación de la Web*  
*Center for Web Research*  
2002

Nombre: Centro de Investigación de la Web

Número de identificación: P01-029-F

Dirección: Blanco Encalada 2120, 3 piso

Teléfono: +56 2 689 2736

Fax: +56 2 689 5531

Correo electrónico: [ciw@dcc.uchile.cl](mailto:ciw@dcc.uchile.cl)

Dirección web del Núcleo: <http://www.ciw.cl>

Nombre y firma del Investigador Responsable del Núcleo:

Ricardo Baeza Yates

Institución Albergante: Departamento de Ciencias de la Computación,  
Universidad de Chile

**Annual Report Millenium Nucleus**  
*Centro de Investigación de la Web*  
*Center for Web Research*  
2002

Name: Web Research Center  
Identification Number: P01-029-F  
Address: Blanco Encalada 2120, 3<sup>rd</sup> floor  
Telephone: +56 2 689 2736  
Fax: +56 2 689 5531  
Email: [cwr@dcc.uchile.cl](mailto:cwr@dcc.uchile.cl)  
Web Address: <http://www.cwr.cl>

Name and Signature Project Leader:

Ricardo Baeza Yates

Hosting Institution: Departamento de Ciencias de la Computación,  
Universidad de Chile

## Contents

<b>PRESENTATION</b> .....	<b>4</b>
<b>ABSTRACT</b> .....	<b>4</b>
<b>INTRODUCTION</b> .....	<b>5</b>
<b>TECHNOLOGICAL AND SCIENTIFIC RESEARCH</b> .....	<b>6</b>
A) DATABASES AND INFORMATION RETRIEVAL .....	6
B) DISTRIBUTED SYSTEMS AND NETWORKS .....	8
<b>EDUCATION AND STUDENTS</b> .....	<b>10</b>
<b>COLLABORATION NETWORKS</b> .....	<b>10</b>
<b>OUTREACH</b> .....	<b>11</b>
<b>ADMINISTRATION</b> .....	<b>12</b>
Administrative Staff.....	12
<b>FINANCES</b> .....	¡ERROR! MARCADOR NO DEFINIDO.
<b>ANNEXES</b> .....	¡ERROR! MARCADOR NO DEFINIDO.

## Presentation

The first year of the project has been dedicated to consolidation of the work group, creating the teams, choosing the students involved, and to the purchase and installation of the hardware. The research work of this project, which is very broad in extent, has been divided into two main areas: a) Databases and Information Retrieval, and b) Distributed Systems and Networks. Through the year 2002, many conferences and invited talks were organized, designed to attract motivated students to the subjects of the project. Some invited professors came to present state of the art research subjects and technologies to our group, and to establish cooperation relationships with us.

## Abstract

During the first year, the nucleus published 10 articles in ISI journals, 3 articles in a book and 7 articles in international conferences. It funded 4 grants for PhD students, 2 for MSc students and 7 for undergraduates. Two short courses (Multimedia Information Retrieval and ProActive Programming) and 8 research talks were given (by local researchers and 4 invited professors), one international conference (HIS 2002) was sponsored and one national meeting (Information Society) was organized. We participated in the Millenium regional meeting at San Luis de Potosí, Mexico, creating a possible cooperation with the Universidad de las Américas (Puebla, Mexico).

In general, the initial goals of the project during the first year have been reached as expected, and we expect an even more productive second year.

## Introduction

The Web Research Center was created to research all the technologies and scientific subjects associated to the Web. This makes the Center research subjects very broad and extensive, covering very diverse aspects of the computer science and networking technologies. Initially, we have focused the efforts in two main areas, to make easier the team work and to avoid a dispersion of the results:

- a) Databases and Information Retrieval: This area covers multimedia retrieval, spatial information and semi-structured data. At the heart of them is combinatorial pattern matching, a research area that studies from a combinatorial point of view how to search for given patterns in regular and discrete structures such as sequences or graphs. Another research topic is how to add semantic information to the contents, such as meta-data and the Semantic Web.
- b) Distributed Systems and Networks: This area covers web application programming technologies (Web Agents, Web Services, Distributed Programming), new media communication protocols (Multimedia over IP) and technologies to improve performance of web tools (parallel search, Web crawling technologies).

In the near future, other areas could be included, depending on the research interests and the collaborations developed by the project researchers.

## Technological and Scientific Research

### A) Databases and Information Retrieval

#### Main Goals:

Create new technologies and algorithms to store and search the Web efficiently. We cover two sub-areas: multimedia analysis and searching techniques, and the semantic web and web mining.

- Multimedia Analysis and Searching Techniques: Researchers: Gonzalo Navarro, Andrea Rodriguez, Javier Ruiz del Solar. PhD Students: R. Paredes, K. Figueroa. MSc Student: F. Godoy. Undergraduate Students: J. L. Delard, R. Verschae. In general, this area covers all the multimedia search problems. The main subjects are Space Searching algorithms and Face Recognition algorithms.

#### Publications:

- Navarrete, P., and Ruiz-del-Solar, J.; Eigenspace-based Face Recognition: A comparative study of different hybrid approaches, In A. Abraham. J. Ruiz-del-Solar, M. Köppen (Eds.), *Soft-Computing Systems: Design, Management and Applications*, *Frontiers in Artificial Intelligence and Applications* 87, IOS Press, 663 - 672 (ISBN 1-58603-297-6).
- Nora Reyes, Gonzalo Navarro; Eliminación en Arboles de Aproximación Espacial Dinámicos. In *Actas del CACIC'02*, pp 821--833.
- Benjamin Bustos, Gonzalo Navarro; Probabilistic Proximity Searching Algorithms Based on Compact Partitions. In *Proc. SPIRE'02*, pp 284--297. LNCS 2476. (ISI)
- Gonzalo Navarro, Rodrigo Paredes, Edgar Chávez; t-Spanners as a Data Structure for Metric Space Searching. In *Proc. SPIRE'02*, pp 298--309. LNCS 2476. (ISI)
- Gonzalo Navarro, Nora Reyes; Fully Dynamic Spatial Approximation Trees. In *Proc. SPIRE'02*, pp 254--270. LNCS 2476. (ISI)
- Kimmo Fredriksson, Gonzalo Navarro, Esko Ukkonen; Optimal Exact and Fast Approximate Two Dimensional Pattern Matching Allowing Rotations. In *Proc. CPM'02*, pp 235--248. LNCS 2373. (ISI)
- Ruiz-del-Solar, J., and Navarrete, P.; FACERET: An Interactive Face Retrieval System based on Self-Organizing Maps. *Lecture Notes in Computer Science* 2383, Springer, 157 - 164. (ISI)
- Ruiz-del-Solar, J., and Navarrete, P.; FACERET: Sistema de Administración y Búsqueda Inteligente de Caras en Bases de Datos de Imágenes. *Anales del Instituto de Ingenieros de Chile*, Vol. 114, No. 3, 117-125.
- A. Rodríguez and F. Godoy. A Content-Based Approach to Searching and indexing Spatial Configurations, *Geographic Information Science* 2002, Colorado, Sep. 2002, pp. 260-275, LNCS vol. 2478, Springer-Verlag. (ISI)
- A. Rodríguez. A Spatial Dimension for Searching the World Wide Web. in A. Abraham et al. (eds.) *Soft Computing Systems: Design, Management and Applications*, IOS Press, 2002.

- Mary Carmen Jarur and M. Andrea Rodríguez, A non-Deterministic versus Deterministic Algorithm for Searching Spatial Configurations. in A. Abraham et al. (eds.) *Soft Computing Systems: Design, Management and Applications*, IOS Press, 2002, p. 602-611.
- Semantic Web and Web Mining: Researchers: Ricardo Baeza-Yates, Claudio Gutiérrez, Carlos Hurtado. Undergraduate Students: E. Krsulovic, M. Calderón, C. Vásquez. This area tries to advance the research on adding and mining semantic information to/from the web contents, using Meta-tags and other tools. New query languages and formats (such as XML) are also explored.
- Publications:

- R. Baeza-Yates, F. Saint-Jean, C. Castillo. Web Structure, Dynamics and Page Quality, In *Proceedings of SPIRE 2002*, LNCS 2467, Springer, Lisbon, Portugal, September 2002. **(ISI)**.
- L. Bertossi, J. Chomicki, A. Cortes, C Gutierrez, "Consistent Answers from Intergrated Data Sources" "FQAS 2002, Lecture Notes in Artificial Intelligence, LNCS 2522, Springer-Verlag, pp. 151-162 **(ISI)**.
- A. Delgado, R. Baeza-Yates, "An Analysis of Query Languages for XML", *UPGRADE*, Vol III, Issue 3, June 2002.
- E. Krsulovic, C. Gutierrez, "Building Yearbooks with RDF" in A. Abraham et al. (Eds.) *Soft Computing Systems: Design, Management & Applications*, IOS Press, 2002, pp. 593-601.
- M. Calderon, C. Gutierrez" "Metadatos Dublin Core para la Web chilena" "Actas del X Encuentro Chileno de Computacion (ECC'2002), Copiapo, Noviembre, 2002, Santiago, Chile" .

## B) Distributed Systems and Networks

### Main Goals:

Create new technologies and systems to program distributed applications for the web, communication protocols to traverse the web and to transmit multimedia content, and parallelism based on cluster technologies to improve the performance of the Database and Information Retrieval algorithms.

- Web Crawling Techniques: Researcher: Ricardo Baeza-Yates, PhD Student: Carlos Castillo. The main subjects are how to crawl the whole web, gathering pages to index them to provide efficient search engines. Many tradeoffs arise when a search system is designed. We have proposed new technologies to improve the efficiency and accuracy of the searching technologies.

#### Publications:

- C. Castillo and R. Baeza-Yates. A New Model for Web Crawling (extended poster). In *WWW11*, Honolulu, Hawaii, May 2002.
- R. Baeza-Yates, C. Castillo, "Balancing Volume, Quality and Freshness in Web Crawling", In proceedings of the Second International Conference on Hybrid Intelligent Systems (Invited session on Web Computing), Dec. 2002 (IOS Press).
- R. Baeza-Yates, J. Piquer, 'Agents, Crawlers and Web Retrieval', Cooperative Information Agents CIA 2002, Madrid, Sep. 2002, pp. 1-9, Lecture Notes in Artificial Intelligence LNAI vol. 2446, Springer-Verlag. **(ISI)**.

- Mobile Agents and Distributed Programming: Researcher: José Piquer, PhD student: Eric Tanter. We have developed a Reflection programming platform for Java (Reflex) to enable an easier agent development platform.

#### Publications:

- Eric Tanter, Marc Segura, Jacques Noye and Jose Piquer, 'Altering Java Semantics via Bytecode Manipulation', 1st ACM SIGPLAN/SIGSOFT Conference on Generative Programming and Component Engineering (GPCE 2002), LNCS 2487, pp. 283--298, October 2002, Pittsburg, PA, USA. **(ISI)**.
- Eric Tanter, Michel Vernailen and Jose Piquer, 'Towards Transparent Adaptation of Migration Policies', 8th ECOOP Workshop on Mobile Object Systems, June 2002, Malaga, Spain.
- J. C. Vidal, A. Slodzian and J. M. Piquer, 'Prediction in an Open and Dynamically Distributed Multi-Agent System', MENDEL 2002 - 8th International Conference on Soft Computing, June 2002, Brno, Czech Republic.



- Paralelism and Simulation: Researcher: Mauricio Marin. MSc Student: J. Canuman. 7 Undergraduate Students. A 10 PC cluster has been installed and is working at the Computer Science Department of the University of Chile. A master thesis doing work upon this cluster is in progress and is related to the efficient paralelization of index data structures for text databases. Research work on the development of parallel algorithms for text databases and distributed scheduling of simulation servers on the Internet.

Publications:

- M. Marin, "Parallel text query processing using Composite Inverted Lists", In proceedings of the Second International Conference on Hybrid Intelligent Systems (Invited session on Web Computing), Dec. 2002 (IOS Press).
- M. Marin and C. Bonacic, "Distributed Allocation of Discrete-Event Simulation Servers", Proceedings of the XXVIII Conferencia Latinoamericana de Informática (CLEI2002), Montevideo, Uruguay, Nov. 2002.
- M. Marin, C. Bonacic y S. Casas. "Analysis of two indexing structures for text databases", Actas del VIII Congreso Argentino de Ciencias de la Computación (CACIC2002). Buenos Aires, Argentina, Oct. 15 - 19, 2002.
- M. Marin y C. Bonacic. "Two-level scheduling of simulation servers", Actas del VIII Congreso Argentino de Ciencias de la Computación (CACIC2002). Buenos Aires, Argentina, Octubre 15 - 19, 2002.

## Education and Students

The main goals of the project are to detect the best available students and promote the research field among them. During the year 2002, one visiting professor, 3 PhD students and 1 MSc at the Computer Science Dept of the University of Chile were financed by the project. Also, 1 MSc at Universidad de Concepción and 7 undergraduate grants were allocated, including students from the Electric Engineering Dept (3), and the Universidad de Magallanes (3).

During 2002, the following students and young researchers were financed by the Center:

- Visiting profesor: Carlos Hurtado
- PhD grants: Eric Tanter, Rodrigo Paredes, Karina Figueroa, Carlos Castillo
- MsC grants: Ernesto Krsulovic, Francisco Godoy.
- Research Assistant: Luz Jaramillo
- Undergraduate Students Grants: Marcela Calderón, Juan Delard, Pablo Navarrete, Rodrigo Verschae, Daniel González, Robinson Lienlaf, Maribel Araya.

This list includes three foreign students (France, Mexico and Colombia).

## Collaboration Networks

During the year, two coordination meetings (with all the researchers involved) were organized, one on the inauguration event (april 19<sup>th</sup>) and one during the HIS 2002 Conference (4<sup>th</sup> december).

Besides the interaction inside the project (Universities of Chile, Concepción and Magallanes), during the year 2002 the following activities were financed:

- Prof Krishnamurthy Balanchander, AT&T Research Labs: One week visit. He gave a talk on "Improving Web Performance by Client Characterization Driver Server"
- Denis Caromel, University of Nice/INRIA Sophia Antipolis, 2 visits of one week. He gave a short one day course on ProActive (a Distributed Java Programming Platform) and a talk on "Towards Software Components for the Grid".
- Isabelle Attali, INRIA Sophia-Antipolis: One week visit. She gave a talk on "Formal Semantics and Environment for Java and JavaCard".
- Dulce Ponceleón, IBM Almaden Research Lab: One week visit. She gave a three day tutorial on "Multimedia Information Retrieval" (in spanish).

With INRIA we cooperated with funding of CONICYT (Project ProXiMos), enabling also a PhD student from the Center to visit INRIA during 2002, and to finance the airline tickets of the visiting researchers from INRIA.

During Dulce Ponceleón visit, a cooperation with IBM Almaden has been agreed.

During November, we participated on the Millenium Meeting in Mexico, and we are developing a possible cooperation with the Universidad de las Américas, Puebla (Gerardo Ayala).

As part of the networking and collaboration effort, 17 conference registrations were funded by the project, mainly for young researchers and students.

## Outreach

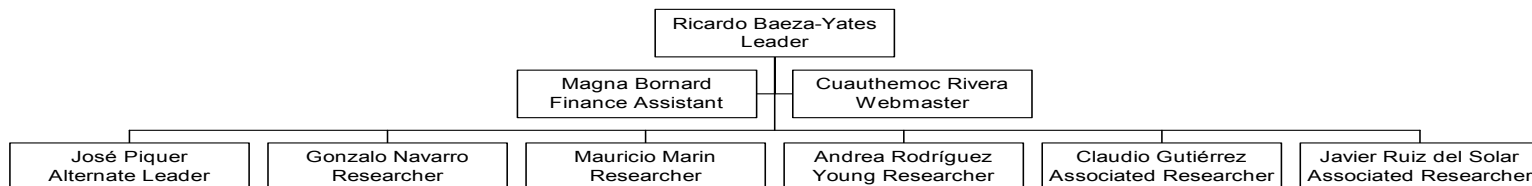
The main goals of the project's outreach is to promote web technology in the Chilean society. During this first year, we organized many technical talks to promote the knowledge and understanding of the underlying technology on the Web. We also sponsored the Hybrid Intelligent Systems Conference (HIS 2002), in Santiago, Chile on December (<http://www.cec.uchile.cl/~his02>), chairing a special session on Web Computing. A meeting to discuss Information Technologies impact on society (III meeting on Information Society) was also organized by the center (<http://www.ciencia.cl/encuentro>).

The following talks and conferences were organized by the project during 2002, at the Auditorium of the Computer Science Department, U. of Chile:

- April 19<sup>th</sup>: Inauguration/Presentation of the Web Research Center. 9:30-18:30
- May 15<sup>th</sup>, 12:00-13:00, Programas en Lógica para Obtener Respuestas Consistentes a Consultas en Bases de Datos" Dr. Leo Bertossi, Carleton University, Ottawa, Canada.
- July 5<sup>th</sup>, 16:00-17:00, Heterogeneidad Estructural en Bases de Datos OLAP, Carlos Hurtado, University of Toronto, Canada.
- July 22<sup>nd</sup>, 16:00-17:00, Improving Web experience by client characterization driven server adaptation, Balachander Krishnamurthy AT&T Labs—Research.
- August 14<sup>th</sup>, 13:00-14:00, Excavando la Web, R. Baeza-Yates, CWR.
- September 30<sup>th</sup>, 14:00-15:00, Randomly Distributed Inverted Files, E. Fox, Virginia Tech, USA.
- October 31<sup>st</sup>, 9:00-18:00, ProActive: Active Objects in Java, D. Caromel, INRIA Sophia-Antipolis.
- December 17<sup>th</sup>, 11:00-12:00, Oracle does Text Search, Omar Alonso, Oracle USA.
- December 17-18-19, 15:00-17:00, Multimedia Information Retrieval, Dulce Poncela, IBM Almaden Research La, USA.
- December 18, 10:00-11:00, Towards Software Components for the Grid, D. Caromel, INRIA Sophia-Antipolis
- December 18, 11:00-12:00, Formal semantics and environment for Java and JavaCard. I. Attali, INRIA Sophia-Antipolis

## Administration

### Staff and Researchers:



### Scientific Staff:

Ricardo Baeza Yates, Leader, 100%  
 José M. Piquer, Alternate Leader, 75%  
 Gonzalo Navarro, Senior Researcher, 75%  
 Mauricio Marín, Researcher, 50%  
 Andrea Rodríguez, Young Researcher, 50%  
 Claudio Gutiérrez, Associate Researcher, 25%  
 Javier Ruiz del Solar, Associate Researcher, 25%  
 Carlos Hurtado, Invited Professor, 100%

### Administrative Staff

Magna Bornand, Finance Assistant, 25%  
 Cuauthemoc Rivera, Webmaster, 25%  
 Guillermo Morales, Staff, 10%

The hosting institution provides offices, connectivity, equipment and support for the Center, and all the basic operational costs.