

Memoria Anual Núcleo Milenio
Centro de Investigación de la Web
Center for Web Research
2003

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: rbaeza@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 Millennium Nucleus
Centro de Investigación de la Web
Center for Web Research
2003

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

Name and Signature of Project Leader:

Ricardo Baeza-Yates

Hosting Institution: Department of Computing Sciences,
Universidad de Chile

Contents

- PRESENTATION 4**
- SUMMARY 4**
- INTRODUCTION 4**
- TECHNOLOGICAL AND SCIENTIFIC RESEARCH 5**
 - A) DATABASES AND INFORMATION RETRIEVAL 5
 - B) DISTRIBUTED SYSTEMS AND NETWORKS 7
- EDUCATION AND STUDENTS 10**
- COLLABORATION NETWORKS 10**
- OUTREACH 13**
- ADMINISTRATION 15**
 - Administrative Staff 15
- FINANCES 16**
- ANNEXES 17**

Presentation

The second year of the project has been dedicated to the consolidation of the research work, by the team of researchers and students involved, developing new concepts and algorithms. The Center currently involves 8 researchers and almost 40 students. The research field 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. Throughout the year 2003, many seminars and other events were organized, designed to attract motivated students to the subjects of the project.

We have increased the number of PhD students in the Center from 3 during 2002 to 10 this year. Also, some invited professors came to present different subjects and technologies to our group, and to establish cooperation relationships with their institutions.

Summary

During this second year, the nucleus published 14 articles in international journals (13 of them ISI), 1 book chapter, 15 articles in international conferences, and 10 articles in national conferences or as technical reports. We also edited two conference proceedings, published abroad. Six grants for PhD students and four for MSc students were funded. Around 20 seminars were organized, by a similar number of invited professors.

The center organized six short courses, one international conference (LA-Web 2003) and one national meeting (Information Society track inside the Chile Digital 2010 conference). A national contest about the Web of the Future was organized for all student levels and a new concept called Digital Window was deployed between Arica and Santiago for about a month to bridge the digital divide.

Introduction

The Center for Web Research was created to study 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. 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, as meta-data and the Semantic Web. Finally, we also include here Web Mining.
- 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, 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: Ricardo Baeza-Yates, Gonzalo Navarro, Andrea Rodriguez, Javier Ruiz del Solar. PhD Students: R. Paredes, K. Figueroa, D. Arroyuelo. MSc Student: F. Godoy. Undergraduate Students: R. Verschae, D. Yaksic. In general, this area covers all the multimedia search problems. The main subjects are Space Searching algorithms and Face recognition algorithms.

Publications:

1. Baeza-Yates, Ricardo: Information retrieval in the web beyond current search engines. *International Journal Approximate Reasoning* 34 (2003), 97-104 (ISI).
2. Ricardo Baeza-Yates, Joaquim Gabarro, and Xavier Messeguer. Fringe Analysis of Synchronized Parallel Insertion Algorithms in 2-3 Trees. *Theoretical Computer Science* 299, pp. 231-271, 2003 (ISI).
3. Jayme Szwarcfiter, Gonzalo Navarro, Ricardo Baeza-Yates, Joísa de S. Oliveira, Walter Cunto and Nivio Ziviani. Optimal Binary Search Trees with Costs Depending on the Access Paths. *Theoretical Computer Science* 290(3), pp. 1799-1814, 2003 (ISI).
4. Baeza-Yates, Ricardo; Chavez, Edgar and Crochemore, Máxime (editors): *Combinatorial Pattern Matching 2003*. *Combinatorial Pattern Matching Proceedings*, Lecture Notes in Computer Science 2676, Springer, 2003 (ISI).
5. Ricardo Baeza-Yates, Benjamin Bustos, Edgar Chávez, Norma Herrera and Gonzalo Navarro. Clustering in Metric Spaces and Its Application to Information Retrieval. In W. Wu, H. Xiong and S. Shekhar (editors), *Clustering and Information Retrieval*. Kluwer Academic Publishers, 2003. ISBN 1-4020-7682-7.
6. Benjamin Bustos, Gonzalo Navarro and Edgar Chávez. Pivot Selection Techniques for Proximity Searching in Metric Spaces. *Pattern Recognition Letters* 24(14):2357-2366 (ISI).
7. Veli Mäkinen, Gonzalo Navarro and Esko Ukkonen. Approximate Matching of Run-Length Compressed Strings. *Algorithmica* 35:347-369, 2003 (ISI).
8. Edgar Chávez and Gonzalo Navarro. Probabilistic Proximity Search: Fighting the Curse of Dimensionality in Metric Spaces. *Information Processing Letters (IPL)* 85:39-46, 2003 (ISI).
9. Gonzalo Navarro and Nora Reyes. Improved Deletions in Dynamic Spatial Approximation Trees. *Proc. SCCC'03*, IEEE CS Press, pages 13-22.

10. Veli Mäkinen, Gonzalo Navarro and Esko Ukkonen. Matching Numeric Strings under Noise. Proc. PSC'03, pages 99-110.
 11. Diego Arroyuelo, Francisca Muñoz, Gonzalo Navarro and Nora Reyes. Memory-Adaptative Dynamic Spatial Approximation Trees. Proc. SPIRE'03, pages 360-368 (short paper). LNCS 2857, Springer (ISI).
 12. Gonzalo Navarro and Rodrigo Paredes. Practical Construction of Metric t-Spanners. Proc. ALENEX'03, pages 69-81. SIAM Press.
 13. Veli Mäkinen, Gonzalo Navarro and Esko Ukkonen. Algorithms for Transposition Invariant String Matching. Proc. STACS'03, pages 191-202. LNCS 2607, Springer (ISI).
 14. Jaimes, A., Ruiz-del-Solar, J., Verschae, R., Yaksic, D., Baeza-Yates, R., Davis, E., Castillo, C. On the Image Content of the Web in Chile. Proc. of the First Latin American Web Congress, 72 – 83, Nov. 10 – 12, 2003, Santiago, Chile.
 15. Maldonado-Naude, F.; Sánchez, A.; Baeza-Yates, Ricardo. : Using Hermes-F: Experiences with a Framework for Developing Information Retrieval Applications. Fourth Mexican International Conference on Computer Science (EWC 2003, Sept. 8-12, Flaxcala Mexico) IEEE Computer Society Press, pp. 101-108.
 16. Roberto Uribe and Gonzalo Navarro. Una Estructura Dinámica para Búsqueda en Espacios Métricos. In Spanish. Proc. XI Jornadas Chilenas de Computación. In CD-ROM.
 17. Diego Arroyuelo, Gonzalo Navarro and Nora Reyes. Fully Dynamic and Memory-Adaptative Spatial Approximation Trees. Proc. CACIC'03, pages 1502-1513.
 18. Diego Arroyuelo, Verónica Ludueña, Nora Reyes and Gonzalo Navarro. Búsquedas en Bases de Datos no Convencionales. In Spanish. Proc. WICC'03, pages 46-50.
- Semantic Web, Web Mining and User Interfaces: Researchers: Ricardo Baeza-Yates, Claudio Gutiérrez, Carlos Hurtado. Ph.D. Students: C. Rivera. Undergraduate Students: E. Krsulovic, M. Calderón, C. Vásquez, R. Lemus, F. Saint-Jean, B. Poblete. 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, visualizations, and formats (such as XML) are also explored. This area also includes mining the structure and usage of the Web.

Publications:

19. Baeza-Yates, Ricardo; Lemus, Ricardo; Dulce Ponceleon and Savitha Srinivasan: WISDNA: An Information Visualization Paradigm for XML. First Latin American Web Conference, IEEE Cs. Press, Santiago Chile, November 2003
20. Alonso, Omar and Baeza-Yates, Ricardo: Alternative Implementation Techniques for Web Text Visualization. First Latin American Web Conference, IEEE Cs. Press, Santiago Chile, November 2003

21. Claudio Gutiérrez, Carlos Hurtado, Alberto Mendelzon. Formal Aspects of Querying RDF Databases. First VLDB International Workshop on Semantic Web and Databases (SWDB), Berlin, September 2003.
22. Claudio Gutiérrez, Carlos Hurtado. RDF and Inconsistency. Dagstuhl-Seminar Nr. 03241, Inconsistency Tolerance. June 2003.
23. Carlos Hurtado, Claudio Gutiérrez, Alberto Mendelzon. Capturing Summarizability with Integrity Constraints in OLAP. Technical Report, Departamento de Ciencias de la Computación, Universidad de Chile, 2003, TR/DCC-2003-6.
24. Carlos Hurtado, Claudio Gutiérrez, Computing Cube View Dependences in OLAP Datacubes, 15th IEEE International Conference on Scientific and Statistical Database Management (SSDBM), July 2003.
25. E. Krsulovic-Morales, C. Gutierrez, Structuring information on the Web from below: The case of Educational Organizations in Chile, First Latin American Web Congress, Santiago, Chile, November 10-12, 2003.
26. Baeza-Yates, Ricardo and Saint Jean, Felipe: Análisis de Consultas a un buscador y su aplicación a la jerarquización de páginas web. BID, número 10, July 2003, Barcelona, Spain.
27. Baeza-Yates, Ricardo and Saint Jean, Felipe: A Three Level Search Engine Index based in Query Log Distribution. String Processing and Information Retrieval 10th International Symposium, SPIRE 2003 Manaus, Brasil, October 2003 . Proceedings, LNCS 2857 (ISI).
28. Baeza-Yates, Ricardo and Poblete, Barbara: Evolution of the Web Structure. International World Wide Web Conference, Budapest, Hungary, May 2003 (extended poster)
29. Baeza-Yates, R., Poblete, B. and Saint-Jean, Felipe. Evolution of the Chilean Web: 2001-2002 (In Spanish), Jornadas Chilenas de Computacion, Chillan, Chile, November 2003.
30. Baeza-Yates, Ricardo and Poblete, Barbara: Evolution of the Chilean Web Structure Composition. First Latin American Web Conference, IEEE CS Press, Santiago Chile, November 2003
31. E. Krsulovic, C. Gutierrez, Infraestructura para la incorporación de metadatos en sitios de Departamentos Universitarios Chilenos; III Workshop de Ingeniería de Software, Chillán 2003.
32. E. Krsulovic, C. Gutierrez, Propuesta para la creación de una Ontología sobre Departamentos Universitarios de Computación en Chile , V Congreso Chileno de Educación Superior en Computación, 2003.

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.

- Crawling and Ranking 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 build efficient search engines. Many tradeoffs arise when a search system is designed, and we have proposed new technologies to improve the efficiency and accuracy of the searching technologies.

Publications:

33. Carlos Castillo, Cooperation Schemes between Web Servers and Crawlers (short paper) First Latin American Web Conference, IEEE CS Press, Santiago Chile, November 2003, pages 31--35.

- Mobile Agents and Distributed Programming: Researcher: José Piquer. PhD students: Eric Tanter, Juan C. Vidal, Javier Bustos. We have developed a Reflection programming platform for Java (Reflex) to enable an easier agent development platform and to explore a reflexive model to link behavioral reflection and Aspect Oriented Programming. Juan C. Vidal is finishing his thesis on agent load balancing and Javier Bustos is starting a research area on Active Object load balancing for Grid Computing.

Publications:

34. Juan C. Vidal, J. Piquer, "Modeling a Distributed Multi-Agent System to Agent Distributing based on Behavior Patterns", 7th IASTED International Conference on Internet and Multimedia Systems and Applications (IMSA 2003), Honolulu, Hawaii, August 2003.

35. Eric Tanter, Jacques Noyé, Denis Caromel, Pierre Cointe, "Partial Behavioral Reflection: Spatial and Temporal Selection of Reification", Proceedings of the 18th Annual ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications (OOPSLA 2003), October 2003, Anaheim, USA. Published as ACM SIGPLAN Notices, volume 38, number 11, November 2003 (ISI)

36. Eric Tanter and Peter Ebraert, "A Flexible Approach to Interactive Runtime Inspection", In 1st Workshop on Advancing the State-of-the-Art in Runtime Inspection (ASARTI 2003), in conjunction with the 17th European Conference on Object-Oriented Programming (ECOOP 2003), July 2003, Darmstadt, Germany.

- Parallelism and Simulation: Researchers: Mauricio Marin, Gonzalo Navarro. MSc Student: J. Canuman. Undergraduate Students: Miguel Mora, Eduardo Salinas, Robinson Lienlaf, Josselyn Vega, Daniel Gonzalez, José Alvarado, Maribel Araya, Javier Leon, Cristian Martinez. 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 parallelization 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. Is being pursued.

Publications:

37. Mauricio Marín and Gonzalo Navarro. Distributed Query Processing using Suffix Arrays. Proc. SPIRE'03, pages 311-325. LNCS 2857 (ISI).

38. Mauricio Marín and Gonzalo Navarro. Suffix Arrays in Parallel. Proc. EuroPar'03, pages 338-341 (short paper). LNCS 2790, Springer (**ISI**).
39. M. Marin, J. Vega and R. Miranda, "Comparative Study of Parallel Suffix Arrays Algorithms", Workshop Chileno de Sistemas Distribuidos y Paralelismo, JCC 2003, Chillan, Noviembre 2003, Chile.
40. M. Printista, G.V. Costa and M. Marin, "Un enfoque paralelo para el procesamiento de consultas sobre Base de Datos Textuales", IX Jornadas Iberoamericanas de Informatica, Cartagena de Indias, Colombia, 2003, ISBN: 84-9602307-9.
41. P. Millado, D. Laguna, O. Sofia and M. Marin, "Representacion visual para la administracion del procesamiento paralelo de consultas SQL", Workshop Argentino de Sistemas Distribuidos y Paralelismo, CACIC 03, La Plata, Argentina 2003.

Finally, we edited one conference proceeding that includes all the topics already mentioned:

42. R. Baeza-Yates, D. Schwabe, J. Piquer, N. Ziviani, L. Olsina (editors), Proceedings of the First Latin-American Web Conference (LA-WEB 2003), IEEE Computer Science Press, CA, USA, November 2003.

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 2003, one post-doctorate position, three visiting professors (visits for at least a month), 6 PhD students and 3 MSc were financed by the project, at the different departments involved. Also, 3 research assistants have been financed at University of Chile, both at the Computer Science and the Electrical Engineering Depts.

During 2003, the following students and young researchers were financed by the Center:

- Post-docs: Carlos Hurtado
- Visiting professors: Alberto Mendelzon (Canada), Alexander Jaimes (USA), Diego Gomez (Colombia)
- PhD grants: Eric Tanter, Rodrigo Paredes, Karina Figueroa, Carlos Castillo, Diego Arroyuelo, Cuauhtemoc Rivera.
- MSc grants: Ernesto Krsulovic, Simón Smith, Bárbara Poblete, Francisco Godoy.
- Research Assistants: Rodrigo Verschae, Dinko Yaksic, Eduardo Rodríguez.

Also, working on projects for the Center, but with external funding:

- PhD students: Juan Carlos Vidal, Javier Bustos, Nora Reyes, Marcelo Mendoza.
- MSc students: Norma Herrera, Roberto Uribe, José Canumán, Mary Carmen Jarur.
- Undergraduate Students: Miguel Mora, Eduardo Salinas, Robinson Lienlaf, Josselyn Vega, Daniel Gonzalez, José Alvarado, Maribel Araya, Javier Leon, Cristian Martinez, Felipe Saint-Jean, Ricardo Lemus, Carolina Bonacic, Veronica Gil, Laura Galarce, Cristian Galvez, Rodrigo Frez, Manuel Osorio.

Among these students, three of them finished their M.Sc. Thesis (Francisco Godoy, Norma Herrera, and Mary Carmen Jarur) and two of them their Engineering Degree (Felipe Saint-Jean, Dinko Yaksic) during 2003.

The distribution of students in the three universities involved are:

- Ph.D.: All at the University of Chile with the following procedence: Chile (4), Argentina (2), Mexico (2), France (1) and Colombia (1).
- M.Sc.: Five at the University of Chile (all Chileans except one from Argentina, two of them coming from University of Magallanes), and three at University of Concepcion (all Chileans except one from Peru).
- Undergraduate: Nine from University of Magallanes and eight from University of Chile.

Collaboration Networks

Besides the interaction inside the project (Universities of Chile, Concepción and Magallanes), the Center is currently collaborating with the following institutions:

- University of Toronto, Canada (Alberto Mendelzon)
- Polytechnic University of Catalunya, Spain (Joaquim Gabarro and Xavier Messeguer).
- Federal University of Rio de Janeiro, Brazil (Jayme Szwarcfiter)
- Federal University of Minas Gerais, Brazil (Nivio Ziviani)
- Universidad de Las Americas, Mexico (Alfredo Sanchez)
- Universidad Michoacana, Mexico (Edgar Chavez)
- University of Helsinki, Finland (Veli Makinen, Esko Ukkonen)
- Universidad Nacional de San Luis, Argentina (Nora Reyes, Norma Herrera)
- Universidad Nacional de La Patagonia, Argentina (M. Printista, G. Costa)
- University of Marne-la-Vallee, France (Maxime Crochemore)
- University of Paris, France (Denis Caromel, Pierre Cointe, Jacques Noye, Peter Ebraert)
- IBM Almaden Research Center, USA (Dulce Ponceleon, Savitha Srinivasan)
- Oracle Corporation, USA (Omar Alonso)

The project leader also visited Stanford University (USA) and University of Melbourne (Australia) to start new research collaborations.

During the year, a coordination meeting (with all the researchers involved) was organized during the LA-Web'03 conference (november).

During the year 2003 we received 20 foreign researchers visiting the Center, most of them giving technical talks, creating the potential for more research collaborations:

- Christopher Miller, Smart Information Flow Technologies, USA; Associates with Etiquette; An Approach to Designing Agents We Can Live With (January)
- Mark Cieliebak, ETH Zurich, Switzerland, The Partial Digest Problem - An Intriguing Challenge in Computational Biology (March)
- Louis Rosenfeld, LouisRosenfeld.com, USA, Information Architecture in Large Companies (March)
- Veli Makinen, University of Helsinki, Finland, Searching Music: Transposition Invariant String Matching (March)
- Felipe Cabrera, Microsoft Corp, USA, The Future of Web Services (April)
- Alberto Mendelzon, University of Toronto, Canada, Man aging XML Documents (April)
- Michael J. Ferguson. Professor Emeritus, INRS-Télécommunications - Université de Québec, Canada; The Telephone, Internet, and the New Telecommunication Infrastructure (April)
- Victor Sandoval, Ecole Nationale des Ponts et Chaussées, Paris, France; Virtual Space and the Digital Divide (April)
- Alexander Jaimes, Columbia University, USA; Concepts and Techniques for Image and Video Retrieval using Visual Content (June)
- Alexander Jaimes, Columbia University, USA, Multimedia Retrieval using Visual Content in Chile: TREC 2003 and the Chilean Web (July)

- Luis Olsina, UNLP, Argentina. Métricas e Indicadores: Dos Conceptos Claves para Medición y Evaluación (July)
- Leopoldo Bertossi, Carleton University, Canada; Consistent Answers in Data Integration Systems (August)
- Nivio Ziviani, Univ. Federal de Minas Gerais, Brazil; Localization, Access and Organization of Information in the Web (November)
- Stefan Decker, ISI, Ireland. Semantic Web: Current Status and Challenges (invited talk to La-Web'03, November)
- Knut Magne Risvik, Yahoo!, USA: Multi-tier Architecture for Web Search Engines (invited talk to La-Web'03, November)
- Ramez Elmasri, Univ. of Texas at Arlington, USA: WebOntEx: Ontology Extraction from Web Pages (invited talk to La-Web'03, November)
- Daniel Schwabe, Catholic Univ. at Rio de Janeiro, Brazil: Co-chair of the La-Web Program Committee (November)
- Ehud Gudes, Dept. of Computer Science, Ben-Gurion University, Beer-Sheva, Israel; Computing Frequent Graph Patterns from Semi Structured Data (November)
- Petr Kroha, TU Chemnitz, Fac Computer Science, Germany, Features of a Web-Oriented Information System (November)
- Andrew Olson, Department of Computer & Information Science, Indiana University at Purdue, USA, Constructing Distributed Computing Systems with the UniFrame Process (December)

With INRIA we cooperated with funding of CONICYT (Project ProXiMos), enabling also a PhD student (Eric Tanter) from the Center to visit INRIA during 2003, and to sign a co-advisor agreement for a second PhD student who stayed 6 months there.

We also participated with the IBM Team in TRECVID 2003, the current standard benchmark for image and video retrieval held in USA every year.

Outreach

The main goal of the project's outreach is to promote web technology in the Chilean society and to develop some relations with the industry in Chile and abroad. During this second year, we organized several short courses to promote the knowledge and understanding of the underlying technology on the Web.

First, we organized the first Latin-American Web conference (LA-Web'03, <http://www.la-web.org>) under the sponsorship of the same organization that organizes the World Wide Web conference. This year the 2nd conference will be held in Brazil. The conference was attended by 120 persons coming from 18 countries, with half of them from Chile. From the rest, 20% were from other countries in Latin America, 15% from Europe, and another 15% from other countries. From those, 40 were graduate students and 20 undergraduate students.

During LA-Web 2003, we organized the following short courses, which were attended in total by more than 100 people, being the most popular the one in metadata which attracted forty librarians:

- Raghavan Srinivas, Sun Microsystems Inc, USA; Network and web services security concepts using Java.
- Bebo White, Stanford University, USA, and Ricardo Baeza-Yates, Chile. Web Accessibility, Findability and Mobility
- Eva Mendez Rodriguez, Universidad Carlos III. Madrid, Spain, Metadatos en Servicios y Sistemas de Información Web
- Henning Breuer, Germany and Cuauhtemoc Rivera, Mexico; Web Usability
- Ricardo Baeza-Yates, XML Information Retrieval

This conference had the co-funding of the Chilean Fulbright Commission and the Iberoamerican Cooperation Program (CYTED).

We also participated in the organization of the first joint-meeting on Information Society called Chile Digital 2010 (<http://www.chiledigital.cl>) taking care of the "Innovar, Investigar y Emprender" track. The track had an attendance ranging from 60 to 150 people.

During June 2003 we organized another short course entitled Information Architecture and Usability on the Web by Ricardo Baeza-Yates, Cuauhtemoc Rivera and Javier Velasco. About 60 people from different areas attended this course, which was repeated with the same success in July during the Systems Engineering Workshop in CasaPiedra, Santiago (Taller de Ingenieria de Sistemas).

We organized a national contest on "The Web of the Future", for primary and secondary school students as well as university students, which was very successful receiving submissions from Arica to Punta Arenas. The final result were published in the Center web site. This contest was co-funded by the Chilean Fulbright commission.

We cooperated with Universidad de Tarapacá to deploy a "Digital Window" concept, communicating Arica and Santiago in a live and permanent public video-Conference over the Internet, with several positional audio channels. This experiment devoted to bridge the digital divide was active at our Faculty (two weeks in October), the Presidential Palace (10 days in November) and the Diego Portales building (one week during Chile Digital 2010), and attracted extensive media coverage, including an article in the main page of the digital version of El Clarin, the main Argentinian

newspaper. In Arica the window was first at the Engineering faculty of University of Tarapaca and then moved to the Governors Building in the main square of that city. This activity was finalist in December of 2003 of the Chilean Digital Challenge Contest. This experience was possible with the cooperation of the Chilean Fulbright commission, the Corporation for the Development of Parinacota and Arica and the Pais Digital Foundation.

We acted as the executive secretariat for the Economy Ministry group for Electronic Document Standard. This work is almost finished, waiting for an official resolution to promote its use inside the government, deploying an official standard norm for digital documents by the INN. This norm will have an important impact on digital government and their interaction with other institutions.

At the beginning of 2003 we released the second study of the Chilean Web entitled Evolution of the Chilean Web: 2001-2002 by R. Baeza-Yates, B. Poblete, and F. Saint-Jean. This study is available at the Center Web site and has been downloaded profusely from inside and outside Chile.

During this year, we developed a Video over IP system for a network camera prototype for SmalCamera, an MIT startup in the US (www.smalcamera.com). The system is ready for production, but has not been released yet. We also collaborated with CASLAB (www.caslab.cl) on the topic of Web security.

The members of the Center also gave invited talks in several places outside and inside Chile as follows:

- Ricardo Baeza-Yates, Web Mining, Stanford University, USA (July 2003) and Melbourne University, Australia (November 2003)
- Ricardo Baeza-Yates, Approximate String Matching, Australasian Language Technology Association, Melbourne, Australia, December 2003.
- Claudio Gutierrez, La Web Semantica, Universidad de Talca, Chile, July 2003.
- E. Krsulovic, C. Gutierrez, Necesidad de Semántica en la Web Chilena, Encuentro Nacional ChileDigital2010, November 2003.
- Mauricio Marin, Paralelismo en Aplicaciones de la Web, Jornadas Chilenas de Computación, Chillan, November 2003 and Jornadas Iberoamericanas de Informática, Cartagena, Colombia, August 2003.
- Gonzalo Navarro, Current Challenges in Textual Databases, Encuentro Mexicano de Computacion, Mexico, September 2003.

We have continued improving our web site, including more information and documents in Spanish (www.ciw.cl) and English (www.cwr.cl).

We already sent to the Millennium Program all the articles published in local newspapers regarding outreach activities.

Administration

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, Postdoc, 100%

Administrative Staff

Javier Velasco, Outreach Coordinator, 50%
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, as well as most of the operational costs.