Lecture Notes in Computer Science 3772

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Mariano Consens   Gonzalo Navarro (Eds.)

String Processing and Information Retrieval

12th International Conference, SPIRE 2005
Buenos Aires, Argentina, November 2-4, 2005
Proceedings

Springer

Volume Editors

Mariano Consens
University of Toronto
Department of Mechanical and Industrial Engineering
Department of Computer Science
Toronto, Canada
E-mail: consens@cs.toronto.edu

Gonzalo Navarro
University of Chile
Center for Web Research, Dept. of Computer Science, Chile
E-mail: gnavarro@dcc.uchile.cl

Library of Congress Control Number: 2005934415

CR Subject Classification (1998): H.3, H.2.8, I.2, E.1, E.5, F.2.2

ISSN 0302-9743
ISBN-10 3-540-29740-5 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-29740-6 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springeronline.com

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11575832 06/3142 5 4 3 2 1 0

Preface

The papers contained in this volume were presented at the 12th edition of the International Symposium on String Processing and Information Retrieval (SPIRE), held November 2–4, 2005, in Buenos Aires, Argentina. They were selected from 102 papers submitted from 25 countries in response to the Call for Papers. A total of 27 submissions were accepted as full papers, yielding an acceptance rate of about 26%. In view of the large number of good-quality submissions the conference program also included 17 short papers that also appear in the proceedings. In addition, the Steering Committee invited the following speakers: Prabhakar Raghavan (Yahoo! Research, USA), Paolo Ferragina (University of Pisa, Italy), and Gonzalo Navarro (University of Chile, Chile).

Papers solicited for SPIRE 2005 were meant to constitute original contributions to areas such as string processing (dictionary algorithms, text searching, pattern matching, text compression, text mining, natural language processing, and automata-based string processing); information retrieval languages, applications, and evaluation (IR modeling, indexing, ranking and filtering, interface design, visualization, cross-lingual IR systems, multimedia IR, digital libraries, collaborative retrieval, Web-related applications, XML, information retrieval from semi-structured data, text mining, and generation of structured data from text); and interaction of biology and computation (sequencing and applications in molecular biology, evolution and phylogenetics, recognition of genes and regulatory elements, and sequence-driven protein structure prediction).

SPIRE has its origins in the South American Workshop on String Processing (WSP). Since 1998 the focus of the conference was broadened to include information retrieval. Starting in 2000, Europe has been the conference venue on even years. The first 11 meetings were held in Belo Horizonte (Brazil, 1993), Valparaíso (Chile, 1995), Recife (Brazil, 1996), Valparaíso (Chile, 1997), Santa Cruz (Bolivia, 1998), Cancún (Mexico, 1999), A Coruña (Spain, 2000), Laguna San Rafael (Chile, 2001), Lisboa (Portugal, 2002), Manaus (Brazil, 2003), and Padova (Italy, 2004).

SPIRE 2005 was held in tandem with LA-WEB 2005, the 3rd Latin American Web Congress, with both conferences sharing a common day in Web Retrieval.

SPIRE 2005 was sponsored by Centro Latinoamericano de Estudios en Informática (CLEI), Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo (CYTED), Center for Web Research (CWR, University of Chile), and Sociedad Argentina de Informática e Investigación Operativa (SADIO).

We thank the local organizers for their support in the organization of SPIRE and the members of the Program Committee and the additional reviewers for providing timely and detailed reviews of the submitted papers and for their active participation in the email discussions that took place before we could assemble the final program. Finally, we would like to thank Ricardo Baeza-Yates, who, on behalf of the Steering Committee, invited us to chair the Program Committee.

November 2005

Mariano P. Consens,
Gonzalo Navarro

SPIRE 2005 Organization

Steering Committee

Ricardo Baeza-Yates (Chair)	ICREA-Universitat Pompeu Fabra (Spain) and Universidad de Chile (Chile)
Alberto Apostolico	Università di Padova (Italy) and Georgia Tech (USA)
Alberto Laender	Universidade Federal de Minas Gerais (Brazil)
Massimo Melucci	Università di Padova (Italy)
Edleno de Moura	Universidade Federal do Amazonas (Brazil)
Mario Nascimento	University of Alberta (Canada)
Arlindo Oliveira	INESC (Portugal)
Berthier Ribeiro-Neto	Universidade Federal de Minas Gerais (Brazil)
Nivio Ziviani	Universidade Federal de Minas Gerais (Brazil)

Program Committee Chairs

Mariano Consens	Dept. of Mechanical and Industrial Engineering Dept. of Computer Science University of Toronto, Canada
Gonzalo Navarro	Center for Web Research Dept. of Computer Science Universidad de Chile, Chile

Program Committee Members

Amihood Amir	Bar-Ilan University (Israel)
Alberto Apostolico	Università di Padova (Italy) and Georgia Tech (USA)
Ricardo Baeza-Yates	ICREA-Universitat Pompeu Fabra (Spain) and Universidad de Chile (Chile)
Nieves R. Brisaboa	Universidade da Coruña (Spain)
Edgar Chávez	Universidad Michoacana (Mexico)
Charles Clarke	University of Waterloo (Canada)
Bruce Croft	University of Massachussetts (USA)
Paolo Ferragina	Università di Pisa (Italy)
Norbert Fuhr	Universität Duisburg-Essen (Germany)
Raffaele Giancarlo	Università di Palermo (Italy)
Roberto Grossi	Università di Pisa (Italy)
Carlos Heuser	Universidade Federal de Rio Grande do Sul (Brazil)

Carlos Hurtado	Universidad de Chile (Chile)
Lucian Ilie	University of Western Ontario (Canada)
Panagiotis Ipeirotis	New York University (USA)
Juha Kärkkäinen	University of Helsinki (Finland)
Nick Koudas	University of Toronto (Canada)
Mounia Lalmas	Queen Mary University of London (UK)
Gad Landau	University of Haifa (Israel) and Polytechnic University (NY, USA)
Stefano Lonardi	University of California at Riverside (USA)
Yoelle Maarek	IBM Haifa Research Lab (Israel)
Veli Mäkinen	Bielefeld University (Germany)
Mauricio Marín	Universidad de Magallanes (Chile)
João Meidanis	UNICAMP (Brazil)
Massimo Melucci	Università di Padova (Italy)
Edleno de Moura	Universidade Federal do Amazonas (Brazil)
Ian Munro	University of Waterloo (Canada)
Arlindo Oliveira	INESC (Portugal)
Kunsoo Park	Seoul National University (Korea)
Prabhakar Raghavan	Yahoo Inc. (USA)
Berthier Ribeiro-Neto	Universidade Federal de Minas Gerais (Brazil)
Kunihiko Sadakane	Kyushu University (Japan)
Marie-France Sagot	INRIA (France)
João Setubal	Virginia Tech (USA)
Jayavel Shanmugasundaram	Cornell University (USA)
Ayumi Shinohara	Tohoku University (Japan)
Jorma Tarhio	Helsinki University of Technology (Finland)
Jeffrey Vitter	Purdue University (USA)
Hugh Williams	Microsoft Corporation (USA)
Hugo Zaragoza	Microsoft Research (UK)
Nivio Ziviani	Universidade Federal de Minas Gerais (Brazil)
Justin Zobel	RMIT (Australia)

External Reviewers

Jussara Almeida
Michela Bacchin
Ramurti Barbosa
Bodo Billerbeck
Sebastian Böcker
Michael Cameron
David Carmel
Luis Coelho
Marco Cristo
Giorgio Maria Di Nunzio
Alair Pereira do Lago
Shiri Dori
Celia Francisca dos Santos
Fan Yang
Feng Shao
Nicola Ferro
Kimmo Fredriksson
Gudrun Fisher
Paulo B. Golgher
Alejandro Hevia
Jie Zheng
Carmel Kent

Shahar Keret
Sascha Kriewel
Nicholas Lester
Julia Mixtacki
Henrik Nottelmann
Rodrigo Paredes
Hannu Peltola
Nadia Pisanti
Bruno Possas
Davi de Castro Reis
Luis Russo
Marinella Sciortino
Darren Shakib
S.M.M. (Saied) Tahaghoghi
Andrew Turpin
Ying Zhang
Tsvi Kopelowitz
Michael Laszlo
Saadia Malik
Viviane Moreira Orengo
Nicola Orio
Laxmi Parida
Patrícia Peres
Benjamin Piwowarski
Jussi Rautio
Nora Reyes
Klaus-Bernd Schürmann
Rahul Shah
Riva Shalom
Eric Tannier
Rodrigo Verschae

Local Organization

SADIO (Argentine Society for Informatics and Operations Research)

SADIO President	Gabriel Baum
Local Arrangements Chair	Héctor Monteverde
Steering Committee Liaison	Ricardo Baeza-Yates
Administrative Manager	Alejandra Villa

Table of Contents

String Processing and Information Retrieval 2005