Lecture Notes in Computer Science 2633
Edited by G. Goos, J. Hartmanis, and J. van Leeuwen

Springer
Berlin
Heidelberg
New York
Barcelona
Hong Kong
London
Milan
Paris
Tokyo

Fabrizio Sebastiani (Ed.)

Advances in Information Retrieval

25th European Conference on IR Research, ECIR 2003
Pisa, Italy, April 14-16, 2003
Proceedings

Springer

Volume Editor

Fabrizio Sebastiani
Istituto di Scienza e Tecnologie dell'Informazione
Consiglio Nazionale delle Ricerche
Via Giuseppe Moruzzi, 1, 56124 Pisa, Italy
E-mail: fabrizio@iei.pi.cnr.it

Cataloging-in-Publication Data applied for

A catalog record for this book is available from the Library of Congress.

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

CR Subject Classification (1998): H.3, H.2, I.2.3, I.2.6, H.4, H.5.4, I.7

ISSN 0302-9743
ISBN 3-540-01274-5 Springer-Verlag Berlin Heidelberg New York

Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

http://www.springer.de

Typesetting: Camera-ready by author, data conversion by DA-TeX Gerd Blumenstein
Printed on acid-free paper SPIN: 10924907 06/3142 5 4 3 2 1 0

Preface

The European Conference on Information Retrieval Research, now in its 25th "Silver Jubilee" edition, was initially established by the Information Retrieval Specialist Group of the British Computer Society (BCS-IRSG) under the name "Annual Colloquium on Information Retrieval Research," and was always held in the United Kingdom until 1997. Since 1998 the location of the colloquium has alternated between the United Kingdom and the rest of Europe, in order to reflect the growing European orientation of the event. For the same reason, in 2001 the event was renamed "European Annual Colloquium on Information Retrieval Research." Since 2002, the proceedings of the Colloquium have been published by Springer-Verlag in their Lecture Notes in Computer Science series.

In 2003 BCS-IRSG decided to rename the event "European Conference on Information Retrieval Research," in order to reflect what the event had slowly turned into, i.e., a full-blown conference with a European program committee, strong peer reviewing, and a (mostly) European audience.

However, ECIR still retains the strong student focus that has characterized the Colloquia since their inception: student fees are kept particularly low, a student travel grant program is available in order to encourage students to attend the conference (and encourage student authors to present their papers personally), and a Best Student Paper Award is assigned (conversely, ECIR has no best paper award).

In terms of submissions, ECIR 2003 has been a record-breaking success, since 101 papers were submitted in response to the call for papers, which amounts to a 94% increase with respect to ECIR 2002 in Glasgow and a 260% increase with respect to ECIR 2001 in Darmstadt. All papers were reviewed by at least three reviewers. Out of the 101 submitted papers, 31 were selected as full papers for oral presentation, and 16 were selected as short papers for poster presentation. Students are very well represented, since 20 out of 31 full papers and 13 out of 16 short papers involve a full-time student, which means that the traditional student focus of the Colloquium has been well preserved.

The contributions in these proceedings are indicative of the wide range of issues being tackled in current IR research, and include both theoretical and experimental work in several media (text, hypertext, structured text, multilingual text, spoken text, images, and music) and in several tasks (search, retrieval, clustering, categorization, both content-based and collaborative filtering, summarization , information extraction, question answering, topic detection and tracking, and visualization), either in centralized or in distributed environments, and tackling either effectiveness or efficiency issues.

I want to thank, first of all, the authors who submitted their papers to ECIR 2003, and thus contributed to the creation of a strong, high-quality program, which allowed us to look forward to an exciting conference. I am also deeply indebted to all the colleagues who accepted to serve on the Program

Committee and to all the colleagues who acted as additional reviewers; thank you for all the good work, and also for meeting the tight reviewing deadlines imposed by the ECIR 2003 schedule. Many thanks also to Karen Spärck Jones and Alberto Del Bimbo for accepting to give the ECIR 2003 keynote speeches; to Keith van Rijsbergen, Maristella Agosti, Ayse Göker, Kees Koster, Peter Ingwersen, and Alan Smeaton, for accepting to join what promises to be a very interesting panel; to Sándor Dominich, Pia Borlund, and Joemon Jose for accepting to be on the Best Student Paper Award Committee; to Ayse Göker and the BCS-IRSG Steering Committee for their support; and to Richard van de Stadt, of Borbala Online Conference Service, for making the Cyberchair conference management software freely available, which greatly simplified all the tasks connected with the submission and the reviewing of the papers.

I am extremely grateful to the companies and institutions who sponsored ECIR 2003: Elsevier, the Associazione Italiana di Informatica e Calcolo Automatico (AICA), the European Information Retrieval Specialist Group of the Council of European Professional Informatics Societies (CEPIS-EIRSG), Libero, Fast Search & Transfer, Canon Research Europe, Microsoft Research, Sharp Laboratories of Europe, IBM Research, and DataPort–AppleCentre. Their generous contribution allowed the Organizing Committee to keep the registration fees low and to enable a strong student grant program.

Finally, I want to extend a special word of thanks to the people who helped me in the organization of the conference: to Patrizia Andronico, who designed the ECIR 2003 official poster and website; to Francesca Borri, who took care of local arrangements with professionalism and much more beyond the call of duty; to the ECIR 2003 webmaster Claudio Gennaro, who painstakingly dealt with all the system issues related to the management of Cyberchair; and to our student volunteers Henri Avancini, Leonardo Candela, and Franca Debole, who helped at various stages of the organization. It is thanks to all of them if the organization of ECIR 2003 was not just hard work, but also a pleasure.

January 2003 Fabrizio Sebastiani

Organization

ECIR 2003 was organized by the Istituto di Scienze e Tecnologie dell'Informazione of the Consiglio Nazionale delle Ricerche (CNR, the Italian National Council of Research), with the collaboration of the Istituto di Informatica e Telematica of the CNR, and under the auspices of the Information Retrieval Specialist Group of the British Computer Society (BCS-IRSG).

Organizing Committee

Chair:	Fabrizio Sebastiani, National Council of Research, Italy
Web & Graphic Design:	Patrizia Andronico, National Council of Research, Italy
Local Arrangements:	Francesca Borri, National Council of Research, Italy
Webmaster :	Claudio Gennaro, National Council of Research, Italy

Program Committee

Fabrizio Sebastiani, National Council of Research, Italy (Chair)

Alan Smeaton, Dublin City University, Ireland
Alessandro Sperduti, University of Padova, Italy
Andreas Rauber, Vienna University of Technology, Austria
Arjen de Vries, Centre for Mathematics and Computer Science, The Netherlands
Avi Arampatzis, University of Nijmegen, The Netherlands
Ayse Göker, Robert Gordon University, UK
Barry Smyth, University College Dublin, Ireland
Carol Peters, National Council of Research, Italy
Claudio Carpineto, Fondazione Ugo Bordoni, Italy
David Carmel, IBM Research, Israel
David Harper, Robert Gordon University, UK
Djoerd Hiemstra, University of Twente, The Netherlands
Dunja Mladenić, Jožef Stefan Institute, Slavenia
Edda Leopold, Fraunhofer Institute, Germany
Eero Sormunen, University of Tampere, Finland
Fabio Crestani, University of Strathclyde, UK
Gabriella Pasi, National Council of Research, Italy
Gareth Jones, University of Exeter, UK
Gianni Amati, Fondazione Ugo Bordoni, Italy
Giuseppe Amato, National Council of Research, Italy
Gloria Bordogna, National Council of Research, Italy

Iadh Ounis, University of Glasgow, UK
Ian Ruthven, University of Strathclyde, UK
Ion Androutsopoulos, Athens University of Economics and Business, Greece
Jacques Savoy, University of Neuchâtel, Switzerland
Jane Reid, Queen Mary University of London, UK
Joemon Jose, University of Glasgow, UK
Johannes Fürnkrantz, Austrian Research Institute for Artificial Intelligence, Austria
John Tait, University of Sunderland, UK
Josiane Mothe, Paul Sabatier University, France
Jussi Karlgren, Swedish Institute of Computer Science, Sweden
Kalervo Järvelin, University of Tampere, Finland
Kees Koster, University of Nijmegen, The Netherlands
Keith van Rijsbergen, University of Glasgow, UK
Marcello Federico, ITC-irst, Italy
Marco Gori, University of Siena, Italy
Mark Girolami, University of Paisley, UK
Mark Sanderson, University of Sheffield, UK
Marko Grobelnik, Jožef Stefan Institute, Slavenia
Martin Braschler, Eurospider Information Technology AG, Switzerland
Massimo Melucci, University of Padova, Italy
Micheline Beaulieu, University of Sheffield, UK
Mohand Boughanem, Paul Sabatier University, France
Monica Landoni, University of Strathclyde, UK
Mounia Lalmas, Queen Mary University of London, UK
Nicholas Kushmerick, University College Dublin, Ireland
Noam Slonim, Hebrew University, Israel
Norbert Fuhr, University of Duisburg-Essen, Germany
Pasquale Savino, National Council of Research, Italy
Patrick Gallinari, University of Paris 6, France
Pavel Zezula, Masaryk University, Czech Republic
Peter Ingwersen, Royal School of Library and Information Science, Denmark
Pia Borlund, Royal School of Library and Information Science, Denmark
Sándor Dominich, University of Veszprém, Hungary
Stavros Christodoulakis, Technical University of Crete, Greece
Umberto Straccia, National Council of Research, Italy
Wessel Kraaij, TNO TPD, The Netherlands
Yoelle Maarek, IBM Research, Israel
Yves Chiaramella, CLIPS-IMAG, France

Best Student Paper Award Committee

Sándor Dominich, University of Veszprém, Hungary (Chair)
Pia Borlund, Royal School of Library and Information Science, Denmark
Joemon Jose, University of Glasgow, UK

Additional Reviewers

Alexei Vinokourov
Athanasios Kehagias
Behzad Shahraray
Bernardo Magnini
Birger Larsen
Carlo Meghini
Caspar Treijtel
ChengXiang Zhai
Claus-Peter Klas
Donna Harman
Erik Thorlund Jepsen
Eugenio Di Sciascio
Fabio Aiolli
Fabio Paternò
Franca Debole
Franciska De Jong
Franco Scarselli
Gilles Hubert
Giorgio Satta
Giuseppe Attardi
Gregory Grefenstette
Henri Avancini
Henrik Nottelmann
Isabelle Moulinier
Janez Brank
Jean Pierre Chevallet
José Maria Gómez-Hidalgo
Jurij Leskovec
Leonardo Candela
Lynda Lechani
Marc Seutter
Maria Elena Renda
Marie-France Bruandet
Massimiliano Pontil
Michelangelo Diligenti
Monica Bianchini
Natasa Milić-Frayling
Nicola Orio
Norbert Gövert
Paolo Ferragina
Renée Pohlmann
Ryen White
Simon Tong
SK Michael Wong
Tassos Tombros
Thorsten Joachims
Tony Rose
Victor Lavrenko
Yoshi Gotoh

Previous Venues of ECIR

2002 Glasgow, UK
2001 Darmstadt, Germany
2000 Cambridge, UK
1999 Glasgow, UK
1998 Grenoble, France
1997 Aberdeen, UK
1996 Manchester, UK
1995 Crewe, UK
1994 Drymen, UK
1993 Glasgow, UK
1992 Lancaster, UK
1991 Lancaster, UK
1990 Huddersfield, UK
1989 Huddersfield, UK
1988 Huddersfield, UK
1987 Glasgow, UK
1986 Glasgow, UK
1985 Bradford, UK
1984 Bradford, UK
1983 Sheffield, UK
1982 Sheffield, UK
1981 Birmingham, UK
1980 Leeds, UK
1979 Leeds, UK

Main Corporate Sponsor

ELSEVIER

Gold Sponsor

AICA
Associazione Italiana
per l'Informatica
ed il Calcolo Automatico

Sponsors

CEPIS

LIBERO

Microsoft Research

CRE
Canon Research Centre Europe

fast

SHARP
Laboratories of
Europe Limited

IBM

Data Port
Apple Solution Experts

Table of Contents

Invited Papers

Papers

IR and the Web

Retrieval of Structured Documents

Collaborative Filtering and Text Mining

Text Representation and Natural Language Processing

Formal Models and Language Models for IR

Machine Learning and IR

Text Categorization 1

Usability, Interactivity, and Visualization Issues in IR

Text Categorization 2

Architectural Issues and Efficiency Issues in IR

Posters