

Giambattista Amati
Claudio Carpineto
Giovanni Romano (Eds.)

LNCS 4425

Advances in Information Retrieval

29th European Conference on IR Research, ECIR 2007
Rome, Italy, April 2007
Proceedings

Giambattista Amati Claudio Carpineto
Giovanni Romano (Eds.)

Advances in Information Retrieval

29th European Conference on IR Research, ECIR 2007
Rome, Italy, April 2-5, 2007
Proceedings



Springer

Volume Editors

Giambattista Amati
Claudio Carpineto
Giovanni Romano
Fondazione Ugo Bordoni
Via Baldassarre Castiglione 59, 00142 Rome, Italy
E-mail: {gba, carpinet, romano}@fub.it

Library of Congress Control Number: 2007923290

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

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

ISSN	0302-9743
ISBN-10	3-540-71494-4 Springer Berlin Heidelberg New York
ISBN-13	978-3-540-71494-1 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media

springer.com

© Springer-Verlag Berlin Heidelberg 2007
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12038824 06/3142 5 4 3 2 1 0

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

University of Dortmund, Germany

Madhu Sudan

Massachusetts Institute of Technology, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Moshe Y. Vardi

Rice University, Houston, TX, USA

Gerhard Weikum

Max-Planck Institute of Computer Science, Saarbruecken, Germany

Preface

This volume contains the papers presented at ECIR 2007, the 29th European Conference on Information Retrieval. The conference was organized by the Fondazione Ugo Bordoni (FUB), in cooperation with the Information Retrieval Specialist Group of the British Computer Society (BCS-IRSG) and the ACM Special Interest Group on Information Retrieval (ACM SIGIR). It was held during April 2–5, 2007 in Rome, hosted by the National Research Council of Italy (CNR).

The conference was initially established by BCS-IRSG in the United Kingdom under the name “Annual Colloquium on Information Retrieval Research”. When the involvement of Continental Europe became more important, it was decided to alternate the conference venue between the UK and Continental Europe. Starting from 2001, the conference was renamed ECIR. In the last few years, ECIR has grown steadily, becoming the major European forum for research on information retrieval.

ECIR 2007 has dramatically confirmed the growth trend, with a high record number of submissions. In four years the number of full-paper submissions to ECIR has nearly tripled, going from 88 (2004, Sunderland) to 124 (2005, Santiago de Compostela), then to 175 (2006, London), and now to 220 (2007, Rome). ECIR 2007 has also attracted more and more people from outside Europe, thus making the conference a large, truly international event. Furthermore, in spite of these major changes, the traditional strong focus on students has been well preserved, for their participation has been massive.

The large number of sponsors of ECIR 2007 both reflects the increasing importance of the conference and is a key to its success. The sponsorship contribution was primarily used to support student attendance and to invite three internationally renowned researchers to give a keynote speech, namely, Andrei Broder (Yahoo! Research), Stephen Robertson (Microsoft Research Cambridge and City University London), and Marco Gori (University of Siena).

Turning to the reviewing process, the Program Committee (PC) of ECIR 2007 was formed by 147 members, 41 of whom were non-European. To increase consistency of refereeing, the PC was split in subgroups of homogeneous expertise with three members each, and each subgroup was then assigned a bunch of related papers.

In response to the call for papers, 220 submissions were received. Each submission was normally reviewed by three PC members. In case of a strong disagreement, the reviewers were given a chance to change their reviews before the PC meeting. The reviews were then thoroughly discussed by the 19 participants in the PC meeting, which was held at FUB in Rome. Decisions were based on the scores assigned by the reviewers and on the reviews themselves. When the reviews of a paper were not sufficient to make a decision, the paper went through an extra review at the PC meeting. Out of the 220 submissions,

42 were accepted for presentation at the conference. Interestingly, 23 of these had a full-time student as the primary author.

In addition, at the PC meeting, it was decided to create a short-paper session because there were many papers with good scores that would have been rejected due to high competition and limited capacity for oral presentation at the conference. Nineteen papers were accepted as short papers. Short papers were given 8 pages in the proceeding (instead of 12), with the short-paper session consisting of a brief oral presentation of all short papers followed by a poster-style exhibition.

There was also a separate call for posters. Each of the 72 poster submissions was normally reviewed by three PC members, as with paper submissions. Twenty-one posters were accepted for presentation.

The topics covered by the accepted papers span well-established as well as emerging research areas, with a concentration on indexing techniques, distributed information retrieval, and query processing. The 42 full papers have been grouped in the following way: Theory and Design (5), Efficiency (5), Peer-to-Peer Networks (4), Result Merging (2), Queries (4), Relevance Feedback (2), Evaluation (2), Classification and Clustering (4), Filtering (4), Topic Identification (2), Expert Finding (2), XML (2), Web IR (2), Multimedia IR (2).

The success of ECIR 2007 was due to a huge number of people and several organizations who were involved in the various stages of the whole process. We thank the researchers who submitted their results, the members of the Program Committee for reviewing many papers in a short time, the members of the local Organizing Committee for their hard work over many months, the members of the Award Committees for choosing the best paper and the the best student paper, the keynote speakers for accepting our invitation, the Italian National Research Council for providing the conference venue, the sponsoring organizations for providing the money. We are deeply indebted to all of them. We would also like to thank the members of BCS-IRSG and the organizers of ECIR 2006 for their useful help. Finally, a special thanks to Fondazione Ugo Bordoni, whose support was invaluable for running the whole conference.

We would like to conclude this preface on a more personal note. Although the amount of work required was definitely harder than expected, we now feel that it was worth it! Organizing and running ECIR 2007 was a great experience for us. We are very happy especially because of our own historical involvement in ECIR. We have seen how the conference has grown over time and we are proud to have contributed to making it a key event in the information retrieval field.

Enjoy the papers!

January 2007

Giambattista Amati
Claudio Carpineto
Giovanni Romano

Organization

ECIR 2007 was organized by the Fondazione Ugo Bordoni (FUB), in collaboration with the Information Retrieval Specialist Group of the British Computer Society (BCS-IRSG) and the ACM Special Interest Group on Information Retrieval (ACM-SIGIR).

General and Program Chairs

Giambattista Amati, Fondazione Ugo Bordoni, Rome, Italy
Claudio Carpineto, Fondazione Ugo Bordoni, Rome, Italy
Giovanni Romano, Fondazione Ugo Bordoni, Rome, Italy

Local Organization Committee

Marilena Carletti, Fondazione Ugo Bordoni
Annalisa Filardo, Fondazione Ugo Bordoni
Sara Saverione, Fondazione Ugo Bordoni
Stefania Vinci, Fondazione Ugo Bordoni
Guido Zampilloni, Fondazione Ugo Bordoni

Program Committee

Maristella Agosti, University of Padova, Italy
James Allan, University of Massachusetts, Amherst, USA
Massih-Reza Amini, Université Pierre et Marie Curie, France
Einat Amitay, IBM Research Lab, Haifa, Israel
Giuseppe Attardi, University of Pisa, Italy
Leif Azzopardi, University of Strathclyde, UK
Ricardo Baeza-Yates, Yahoo! Research, Barcelona, Spain
Alex Bailey, Google, UK
Álvaro Barreiro, Universidade da Coruña, Spain
Roberto Basili, University of Tor Vergata, Roma, Italy
Holger Bast, Max Planck Institute for Informatics, Germany
Micheline Beaulieu, University of Sheffield, UK
Michel Beigbeder, École Nationale Supérieure, Saint-Etienne, France
Nicholas Belkin, Rutgers University, USA
Gloria Bordogna, IDPA CNR, Italy
Theo Bothma, University of Pretoria, South Africa

VIII Organization

Mohand Boughanem, University Paul Sabatier, France
Giorgio Brajnik, Università di Udine, Italy
Martin Braschler, Zurich University of Applied Sciences, Switzerland
Peter Bruza, Queensland University of Technology, Australia
Stefan Büttcher, University of Waterloo, Canada
Wray Buntine, Helsinki Institute of Information Technology, Finland
Fidel Cacheda, Universidade da Coruña, Spain
Jamie Callan, Carnegie Mellon University, USA
Caterina Caracciolo, FAO, Italy
David Carmel, IBM Research Lab, Haifa, Israel
Carlos Castillo, Yahoo! Research, Spain
Yves Chiararella, IMAG, Grenoble, France
Paul-Alexandru Chirita, L3S Research Center, Hannover, Germany
Charles L. A. Clarke, University of Waterloo, Canada
Paul Clough, University of Sheffield, UK
Gordon Cormack, University of Waterloo, Canada
Nick Craswell, Microsoft, USA
Fabio Crestani, University of Strathclyde, UK
Bruce Croft, University of Massachusetts, Amherst, USA
Pablo De la Fuente, Universidad de Valladolid, Spain
Renato De Mori, University of Avignon, France
Maarten de Rijke, University of Amsterdam, The Netherlands
Arjen de Vries, CWI, The Netherlands
Marcello Federico, Istituto Trentino di Cultura, Italy
Ronen Feldman, Bar-Ilan University, Israel
Juan Manuel Fernández Luna, University of Granada, Spain
Paolo Ferragina, University of Pisa, Italy
Nicola Ferro, University of Padova, Italy
Edward Fox, Virginia Tech, USA
Johannes Fürnkranz, TU Darmstadt, Germany
Norbert Fuhr, University of Duisburg-Essen, Germany
Patrick Gallinari, LIP6, Université Pierre et Marie Curie, France
Éric Gaussier, Xerox Research Centre Europe, France
C. Lee Giles, Pennsylvania State University, USA
Mark Girolami, University of Glasgow, UK
Ayse Goker, Robert Gordon University, UK
Julio Gonzalo, UNED, Spain
Margaret Graham, Northumbria University, UK
Warren Greiff, The MITRE Corporation, USA
David Grossman, Illinois Institute of Technology, USA
Antonio Gulli, ASK, Italy
Cathal Gurrin, Dublin City University, Ireland
Preben Hansen, SICS, Sweden
David Hawking, CSIRO ICT Centre, Australia

Ben He, University of Glasgow, UK
William Hersh, Oregon Health Sciences University, USA
Djoerd Hiemstra, University of Twente, The Netherlands
Eduard Hoenkamp, University of Nijmegen, The Netherlands
Andreas Hotho, Universitt Kassel, Germany
Xiaohua Hu, Drexel University, USA
Theo Huibers, University of Twente, The Netherlands
Peter Ingwersen, Royal School of Library and Information Science, Denmark
Mario J. Gaspar da Silva, University of Lisbon, Portugal
Kalervo Jarvelin, University of Tampere, Finland
Gareth Jones, Dublin City University, Ireland
Joemon Jose, University of Glasgow, UK
Jaap Kamps, University of Amsterdam, The Netherlands
Jussi Karlgren, SICS, Sweden
Gabriella Kazai, Microsoft, Cambridge, UK
Manolis Koubarakis, Technical University of Crete, Greece
Wessel Kraaij, TNO TPD, The Netherlands
S. Ravi Kumar, Yahoo! Research, USA
Kui-Lam Kwok, Queens College, City University of New York, USA
Mounia Lalmas, Queen Mary, University of London, UK
Monica Landoni, University of Strathclyde, UK
Birger Larsen, Royal School of Library and Information Science, Denmark
Mun-Kew Leong, Laboratories of Information Technology, Singapore
David Lewis, David D. Lewis Consulting, USA
Xuelong Li, School of Computer Science and Information Systems, London, UK
Yang Lingpeng, Institute for Infocomm Research, Singapore
Christina Lioma, University of Glasgow, UK
David Losada, University of Santiago de Compostela, Spain
Craig Macdonald, University of Glasgow, UK
Andrew MacFarlane, City University, London, UK
Marco Maggini, University of Siena, Italy
Bernardo Magnini, Istituto Trentino di Cultura, Italy
Massimo Melucci, University of Padova, Italy
Alessandro Micarelli, University of Roma Tre, Italy
Stefano Mizzaro, University of Udine, Italy
Dunja Mladenic, Jozef Stefan Institute, Slovenia
Marie-Francine Moens, Katholieke Universiteit Leuven, Belgium
Alistair Moffat, University of Melbourne, Australia
Josiane Mothe, IIRIT, France
Gheorghe Muresan, Rutgers University, USA
Jian-Yun Nie, University of Montreal, Canada
Michael Oakes, University of Sunderland, UK
Stanislaw Osinski, Poznan Supercomputing and Networking Center, Poland
Iadh Ounis, University of Glasgow, UK

Marius Pasca, Google, USA
 Gabriella Pasi, University of Milano - Bicocca, Italy
 Jan O. Pedersen, Yahoo!, USA
 Nils Pharo, Oslo University College, Norway
 Vassilis Plachouras, University of Glasgow, UK
 Jay Ponte, Google, USA
 Victor Poznanski, Sharp, UK
 Andreas Rauber, Vienna University of Technology, Austria
 Stephen Robertson, Microsoft Research, UK
 Thomas Rolleke, Queen Mary, University of London, UK
 Stefan Rüger, Imperial College, UK
 Song Ruihua, Microsoft Research Asia, China
 Ian Ruthven, University of Strathclyde, UK
 Dominich Sandor, University of Veszprem, Hungary
 Ralf Schenkel, Max-Planck-Institut für Informatik, Germany
 Hinrich Schuetze, University of Stuttgart, Germany
 Giovanni Semeraro, University of Bari, Italy
 Stefan Siersdorfer, Max Planck Institute for Computer Science, Germany
 Fabrizio Silvestri, ISTI CNR, Italy
 Alan Smeaton, Dublin City University, Ireland
 Barry Smyth, University College Dublin, Ireland
 Vaclav Snasel, VSB-Technical University Ostrava, Czech Republic
 Eero Sormunen, University of Tampere, Finland
 Amanda Spink, Queensland University of Technology, Australia
 John Tait, University of Sunderland, UK
 Martin Theobald, Max Planck Institute for Informatics, Germany
 Ulrich Thiel, Fraunhofer IPSI, Germany
 Anastasios Tombros, Queen Mary, University of London, UK
 Andrew Tomkins, Yahoo!, USA
 Stephen Tomlinson, Hummingbird, USA
 Pertti Vakkari, University of Tampere, Finland
 Keith van Rijsbergen, University of Glasgow, UK
 Olga Vechtomova, University of Waterloo, Canada
 Dawid Weiss, Poznan University of Technology, Poland
 Ross Wilkinson, CSIRO ICT Centre, Australia
 Wensi Xi, Google, USA
 Tao Yang, ASK, USA
 Elad Yom-Tov, IBM Research Lab, Haifa, Israel
 Hugo Zaragoza, Yahoo! Research, Barcelona, Spain
 Hua-Jun Zeng, Microsoft Research Asia, China
 ChengXiang Zhai, University of Illinois at Urbana-Champaign, USA
 Dell Zhang, University of London, UK
 Justin Zobel, RMIT, Australia

Paper Awards (Sponsored by Yahoo! Research)

Best Paper Award Committee

Fabio Crestani, University of Strathclyde, UK (Chair)

David Losada, University of Santiago de Compostela, Spain

Hinrich Schuetze, University of Stuttgart, Germany

Best Student Paper Award Committee

Maristella Agosti, University of Padova, Italy (Chair)

David Lewis, David D. Lewis Consulting, USA

Giovanni Semeraro, University of Bari, Italy

Additional Reviewers

Eija Airio, University of Tampere, Finland

Mark Baillie, University of Strathclyde, UK

Nuno Cardoso, University of Lisboa, Portugal

Mauro Cettolo, Istituto Trentino di Cultura, Italy

Marco Degemmis, University of Bari, Italy

Gianna Del Corso, University of Pisa, Italy

Chris H.Q. Ding, Lawrence Berkeley National Laboratory, USA

Nicola Fanizzi, University of Bari, Italy

Stefano Ferilli, University of Bari, Italy

Claudio Giuliano, Istituto Trentino di Cultura, Italy

Milen Kouylekov, Istituto Trentino di Cultura, Italy

Jie Lu, Carnegie Mellon University, USA

Pasquale Lops, University of Bari, Italy

Joao Magalhaes, Imperial College, UK

Inderjeet Mani, The MITRE Corporation, USA

Vanessa Murdock, Yahoo! Research, Barcelona, Spain

Matteo Negri, Istituto Trentino di Cultura, Italy

Domenico Redavid, University of Bari, Italy

James Thom, RMIT, Australia

Lecture Notes in Computer Science

For information about Vols. 1–4328

please contact your bookseller or Springer

Vol. 4444: T. Reps, M. Sagiv, J. Bauer (Eds.), *Program Analysis and Compilation, Theory and Practice*. X, 361 pages. 2007.

Vol. 4429: R. Lu, J.H. Siekmann, C. Ullrich (Eds.), *Cognitive Systems*. X, 161 pages. 2007. (Sublibrary LNAI).

Vol. 4425: G. Amati, C. Carpineto, G. Romano (Eds.), *Advances in Information Retrieval*. XIX, 759 pages. 2007.

Vol. 4424: O. Grumberg, M. Huth (Eds.), *Tools and Algorithms for the Construction and Analysis of Systems*. XX, 738 pages. 2007.

Vol. 4423: H. Seidl (Ed.), *Foundations of Software Science and Computational Structures*. XVI, 379 pages. 2007.

Vol. 4422: M.B. Dwyer, A. Lopes (Eds.), *Fundamental Approaches to Software Engineering*. XV, 440 pages. 2007.

Vol. 4421: R. De Nicola (Ed.), *Programming Languages and Systems*. XVII, 538 pages. 2007.

Vol. 4420: S. Krishnamurthi, M. Odersky (Eds.), *Compiler Construction*. XIV, 233 pages. 2007.

Vol. 4419: P.C. Diniz, E. Marques, K. Bertels, M.M. Fernandes, J.M.P. Cardoso (Eds.), *Reconfigurable Computing: Architectures, Tools and Applications*. XIV, 391 pages. 2007.

Vol. 4418: A. Gagalowicz, W. Philips (Eds.), *Advances in Computer Vision and Computer Graphics*. XV, 620 pages. 2007.

Vol. 4416: A. Bemporad, A. Bicchi, G. Buttazzo (Eds.), *Hybrid Systems: Computation and Control*. XVII, 797 pages. 2007.

Vol. 4415: P. Lukowicz, L. Thiele, G. Tröster (Eds.), *Architecture of Computing Systems - ARCS 2007*. X, 297 pages. 2007.

Vol. 4414: S. Hochreiter, R. Wagner (Eds.), *Bioinformatics Research and Development*. XVI, 482 pages. 2007. (Sublibrary LNBI).

Vol. 4410: A. Branco (Ed.), *Anaphora: Analysis, Algorithms and Applications*. X, 191 pages. 2007. (Sublibrary LNAI).

Vol. 4407: G. Puebla (Ed.), *Logic-Based Program Synthesis and Transformation*. VIII, 237 pages. 2007.

Vol. 4405: L. Padgham, F. Zambonelli (Eds.), *Agent-Oriented Software Engineering VII*. XII, 225 pages. 2007.

Vol. 4403: S. Obayashi, K. Deb, C. Poloni, T. Hiroyasu, T. Murata (Eds.), *Evolutionary Multi-Criterion Optimization*. XIX, 954 pages. 2007.

Vol. 4399: X. Llorà, T. Kovacs, K. Takadama, P.L. Lanzi, S.W. Wilson, W. Stolzmann (Eds.), *Learning Classifier Systems*. XII, 345 pages. 2007. (Sublibrary LNAI).

Vol. 4397: C. Stephanidis, M. Pieper (Eds.), *Universal Access in Ambient Intelligence Environments*. XV, 467 pages. 2007.

Vol. 4396: J. García-Vidal, L. Cerdà-Alabern (Eds.), *Wireless Systems and Mobility in Next Generation Internet*. IX, 271 pages. 2007.

Vol. 4395: M. Daydé, J.M.L.M. Palma, Á.L.G.A. Coutinho, E. Pacitti, J.C. Lopes (Eds.), *High Performance Computing for Computational Science - VEC- PAR 2006*. XXIV, 721 pages. 2007.

Vol. 4394: A. Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing*. XVI, 648 pages. 2007.

Vol. 4393: W. Thomas, P. Weil (Eds.), *STACS 2007*. XVIII, 708 pages. 2007.

Vol. 4392: S.P. Vadhan (Ed.), *Theory of Cryptography*. XI, 595 pages. 2007.

Vol. 4391: Y. Stylianou, M. Faundez-Zanuy, A. Esposito (Eds.), *Progress in Nonlinear Speech Processing*. XII, 269 pages. 2007.

Vol. 4390: S.O. Kuznetsov, S. Schmidt (Eds.), *Formal Concept Analysis*. X, 329 pages. 2007. (Sublibrary LNAI).

Vol. 4389: D. Weyns, H.V.D. Parunak, F. Michel (Eds.), *Environments for Multi-Agent Systems III*. X, 273 pages. 2007. (Sublibrary LNAI).

Vol. 4385: K. Coninx, K. Luyten, K.A. Schneider (Eds.), *Task Models and Diagrams for Users Interface Design*. XI, 355 pages. 2007.

Vol. 4384: T. Washio, K. Satoh, H. Takeda, A. Inokuchi (Eds.), *New Frontiers in Artificial Intelligence*. IX, 401 pages. 2007. (Sublibrary LNAI).

Vol. 4383: E. Bin, A. Ziv, S. Ur (Eds.), *Hardware and Software, Verification and Testing*. XII, 235 pages. 2007.

Vol. 4381: J. Akiyama, W.Y.C. Chen, M. Kano, X. Li, Q. Yu (Eds.), *Discrete Geometry, Combinatorics and Graph Theory*. XI, 289 pages. 2007.

Vol. 4380: S. Spaccapietra, P. Atzeni, F. Fages, M.-S. Hacid, M. Kifer, J. Mylopoulos, B. Pernici, P. Shvaiko, J. Trujillo, I. Zaihrayeu (Eds.), *Journal on Data Semantics VIII*. XV, 219 pages. 2007.

Vol. 4378: I. Virbitskaite, A. Voronkov (Eds.), *Perspectives of Systems Informatics*. XIV, 496 pages. 2007.

Vol. 4377: M. Abe (Ed.), *Topics in Cryptology – CT-RSA 2007*. XI, 403 pages. 2006.

- Vol. 4376: E. Frachtenberg, U. Schwiegelshohn (Eds.), *Job Scheduling Strategies for Parallel Processing. VII*, 257 pages. 2007.
- Vol. 4374: J.F. Peters, A. Skowron, I. Düntsch, J. Grzymala-Busse, E. Orłowska, L. Polkowski (Eds.), *Transactions on Rough Sets VI, Part I. XII*, 499 pages. 2007.
- Vol. 4373: K. Langendoen, T. Voigt (Eds.), *Wireless Sensor Networks. XIII*, 358 pages. 2007.
- Vol. 4372: M. Kaufmann, D. Wagner (Eds.), *Graph Drawing. XIV*, 454 pages. 2007.
- Vol. 4371: K. Inoue, K. Satoh, F. Toni (Eds.), *Computational Logic in Multi-Agent Systems. X*, 315 pages. 2007. (Sublibrary LNAI).
- Vol. 4370: P.P. Lévy, B. Le Grand, F. Poulet, M. Soto, L. Darago, L. Toubiana, J.-F. Vibert (Eds.), *Pixelization Paradigm. XV*, 279 pages. 2007.
- Vol. 4369: M. Umeda, A. Wolf, O. Bartenstein, U. Geske, D. Seipel, O. Takata (Eds.), *Declarative Programming for Knowledge Management. X*, 229 pages. 2006. (Sublibrary LNAI).
- Vol. 4368: T. Erlebach, C. Kaklamani (Eds.), *Approximation and Online Algorithms. X*, 345 pages. 2007.
- Vol. 4367: K. De Bosschere, D. Kaeli, P. Stenström, D. Whalley, T. Ungerer (Eds.), *High Performance Embedded Architectures and Compilers. XI*, 307 pages. 2007.
- Vol. 4366: K. Tuyls, R. Westra, Y. Saeys, A. Nowé (Eds.), *Knowledge Discovery and Emergent Complexity in Bioinformatics. IX*, 183 pages. 2007. (Sublibrary LNBI).
- Vol. 4364: T. Kühne (Ed.), *Models in Software Engineering. XI*, 332 pages. 2007.
- Vol. 4362: J. van Leeuwen, G.F. Italiano, W. van der Hoek, C. Meinel, H. Sack, F. Plášil (Eds.), *SOFSEM 2007: Theory and Practice of Computer Science. XXI*, 937 pages. 2007.
- Vol. 4361: H.J. Hoogeboom, G. Păun, G. Rozenberg, A. Salomaa (Eds.), *Membrane Computing. IX*, 555 pages. 2006.
- Vol. 4360: W. Dubitzky, A. Schuster, P.M.A. Sloot, M. Schroeder, M. Romberg (Eds.), *Distributed, High-Performance and Grid Computing in Computational Biology. X*, 192 pages. 2007. (Sublibrary LNBI).
- Vol. 4358: R. Vidal, A. Heyden, Y. Ma (Eds.), *Dynamical Vision. IX*, 329 pages. 2007.
- Vol. 4357: L. Buttyán, V. Gligor, D. Westhoff (Eds.), *Security and Privacy in Ad-Hoc and Sensor Networks. X*, 193 pages. 2006.
- Vol. 4355: J. Julliand, O. Kouchnarenko (Eds.), *B 2007: Formal Specification and Development in B. XIII*, 293 pages. 2006.
- Vol. 4354: M. Hanus (Ed.), *Practical Aspects of Declarative Languages. X*, 335 pages. 2006.
- Vol. 4353: T. Schwentick, D. Suciu (Eds.), *Database Theory – ICDT 2007. XI*, 419 pages. 2006.
- Vol. 4352: T.-J. Cham, J. Cai, C. Dorai, D. Rajan, T.-S. Chua, L.-T. Chia (Eds.), *Advances in Multimedia Modeling, Part II. XVIII*, 743 pages. 2006.
- Vol. 4351: T.-J. Cham, J. Cai, C. Dorai, D. Rajan, T.-S. Chua, L.-T. Chia (Eds.), *Advances in Multimedia Modeling, Part I. XIX*, 797 pages. 2006.
- Vol. 4349: B. Cook, A. Podelski (Eds.), *Verification, Model Checking, and Abstract Interpretation. XI*, 395 pages. 2007.
- Vol. 4348: S.T. Taft, R.A. Duff, R.L. Brukardt, E. Ploederer, P. Leroy (Eds.), *Ada 2005 Reference Manual. XXII*, 765 pages. 2006.
- Vol. 4347: J. Lopez (Ed.), *Critical Information Infrastructures Security. X*, 286 pages. 2006.
- Vol. 4346: L. Brim, B. Haverkort, M. Leucker, J. van de Pol (Eds.), *Formal Methods: Applications and Technology. X*, 363 pages. 2007.
- Vol. 4345: N. Maglaveras, I. Chouvarda, V. Koutkias, R. Brause (Eds.), *Biological and Medical Data Analysis. XIII*, 496 pages. 2006. (Sublibrary LNBI).
- Vol. 4344: V. Gruhn, F. Oquendo (Eds.), *Software Architecture. X*, 245 pages. 2006.
- Vol. 4342: H. de Swart, E. Orłowska, G. Schmidt, M. Roubens (Eds.), *Theory and Applications of Relational Structures as Knowledge Instruments II. X*, 373 pages. 2006. (Sublibrary LNAI).
- Vol. 4341: P.Q. Nguyen (Ed.), *Progress in Cryptology – VIETCRYPT 2006. XI*, 385 pages. 2006.
- Vol. 4340: R. Prodan, T. Fahringer, *Grid Computing. XXIII*, 317 pages. 2007.
- Vol. 4339: E. Ayguadé, G. Baumgartner, J. Ramanujam, P. Sadayappan (Eds.), *Languages and Compilers for Parallel Computing. XI*, 476 pages. 2006.
- Vol. 4338: P. Kalra, S. Peleg (Eds.), *Computer Vision, Graphics and Image Processing. XV*, 965 pages. 2006.
- Vol. 4337: S. Arun-Kumar, N. Garg (Eds.), *FSTTCS 2006: Foundations of Software Technology and Theoretical Computer Science. XIII*, 430 pages. 2006.
- Vol. 4336: V.R. Basili, D. Rombach, K. Schneider, B. Kitchenham, D. Pfahl, R.W. Selby, *Empirical Software Engineering Issues. XVII*, 193 pages. 2007.
- Vol. 4335: S.A. Brueckner, S. Hassas, M. Jelasity, D. Yamins (Eds.), *Engineering Self-Organising Systems. XII*, 212 pages. 2007. (Sublibrary LNAI).
- Vol. 4334: B. Beckert, R. Hähnle, P.H. Schmitt (Eds.), *Verification of Object-Oriented Software. XXIX*, 658 pages. 2007. (Sublibrary LNAI).
- Vol. 4333: U. Reimer, D. Karagiannis (Eds.), *Practical Aspects of Knowledge Management. XII*, 338 pages. 2006. (Sublibrary LNAI).
- Vol. 4332: A. Bagchi, V. Atluri (Eds.), *Information Systems Security. XV*, 382 pages. 2006.
- Vol. 4331: G. Min, B. Di Martino, L.T. Yang, M. Guo, G. Ruenger (Eds.), *Frontiers of High Performance Computing and Networking – ISPA 2006 Workshops. XXXVII*, 1141 pages. 2006.
- Vol. 4330: M. Guo, L.T. Yang, B. Di Martino, H.P. Zima, J. Dongarra, F. Tang (Eds.), *Parallel and Distributed Processing and Applications. XVIII*, 953 pages. 2006.
- Vol. 4329: R. Barua, T. Lange (Eds.), *Progress in Cryptology – INDOCRYPT 2006. X*, 454 pages. 2006.

Table of Contents

Keynote Talks

The Next Generation Web Search and the Demise of the Classic IR Model	1
<i>Andrei Broder</i>	
The Last Half-Century: A Perspective on Experimentation in Information Retrieval	2
<i>Stephen Robertson</i>	
Learning in Hyperlinked Environments	3
<i>Marco Gori</i>	

Theory and Design

A Parameterised Search System	4
<i>Roberto Cornacchia and Arjen P. de Vries</i>	
Similarity Measures for Short Segments of Text	16
<i>Donald Metzler, Susan Dumais, and Christopher Meek</i>	
Multinomial Randomness Models for Retrieval with Document Fields ...	28
<i>Vassilis Plachouras and Iadh Ounis</i>	
On Score Distributions and Relevance	40
<i>Stephen Robertson</i>	
Modeling Term Associations for Ad-Hoc Retrieval Performance Within Language Modeling Framework	52
<i>Xing Wei and W. Bruce Croft</i>	

Efficiency

Static Pruning of Terms in Inverted Files	64
<i>Roi Blanco and Álvaro Barreiro</i>	
Efficient Indexing of Versioned Document Sequences	76
<i>Michael Herscovici, Ronny Lempel, and Sivan Yogev</i>	
Light Syntactically-Based Index Pruning for Information Retrieval	88
<i>Christina Lioma and Iadh Ounis</i>	
Sorting Out the Document Identifier Assignment Problem	101
<i>Fabrizio Silvestri</i>	
Efficient Construction of FM-index Using Overlapping Block Processing for Large Scale Texts	113
<i>Di Zhang, Yunquan Zhang, and Jing Chen</i>	

Peer-to-Peer Networks (In Memory of Henrik Nottelmann)

Performance Comparison of Clustered and Replicated Information Retrieval Systems 124
Fidel Cacheda, Victor Carneiro, Vassilis Plachouras, and Iadh Ounis

A Study of a Weighting Scheme for Information Retrieval in Hierarchical Peer-to-Peer Networks 136
Massimo Melucci and Alberto Poggiani

A Decision-Theoretic Model for Decentralised Query Routing in Hierarchical Peer-to-Peer Networks 148
Henrik Nottelmann and Norbert Fuhr

Central-Rank-Based Collection Selection in Uncooperative Distributed Information Retrieval 160
Milad Shokouhi

Result Merging

Results Merging Algorithm Using Multiple Regression Models 173
George Paltoglou, Michail Salampasis, and Maria Satratzemi

Segmentation of Search Engine Results for Effective Data-Fusion 185
Milad Shokouhi

Queries

Query Hardness Estimation Using Jensen-Shannon Divergence Among Multiple Scoring Functions 198
Javed A. Aslam and Virgil Pavlu

Query Reformulation and Refinement Using NLP-Based Sentence Clustering 210
Frédéric Roulland, Aaron Kaplan, Stefania Castellani, Claude Roux, Antonietta Grasso, Karin Pettersson, and Jacki O'Neill

Automatic Morphological Query Expansion Using Analogy-Based Machine Learning 222
Fabienne Moreau, Vincent Claveau, and Pascale Sébillot

Advanced Structural Representations for Question Classification and Answer Re-ranking 234
Silvia Quarteroni, Alessandro Moschitti, Suresh Manandhar, and Roberto Basili

Relevance Feedback

Incorporating Diversity and Density in Active Learning for Relevance Feedback 246
Zuobing Xu, Ram Akella, and Yi Zhang

Relevance Feedback Using Weight Propagation Compared with Information-Theoretic Query Expansion	258
<i>Fadi Yamout, Michael Oakes, and John Tait</i>	

Evaluation

A Retrieval Evaluation Methodology for Incomplete Relevance Assessments	271
<i>Mark Baillie, Leif Azzopardi, and Ian Ruthven</i>	
Evaluating Query-Independent Object Features for Relevancy Prediction	283
<i>Andres R. Masegosa, Hideo Joho, and Joemon M. Jose</i>	

Classification and Clustering

The Utility of Information Extraction in the Classification of Books	295
<i>Tom Betts, Maria Milosavljevic, and Jon Oberlander</i>	
Combined Syntactic and Semantic Kernels for Text Classification	307
<i>Stephan Bloehdorn and Alessandro Moschitti</i>	
Fast Large-Scale Spectral Clustering by Sequential Shrinkage Optimization	319
<i>Tie-Yan Liu, Huai-Yuan Yang, Xin Zheng, Tao Qin, and Wei-Ying Ma</i>	
A Probabilistic Model for Clustering Text Documents with Multiple Fields	331
<i>Shanfeng Zhu, Ichigaku Takigawa, Shuqin Zhang, and Hiroshi Mamitsuka</i>	

Filtering

Personalized Communities in a Distributed Recommender System	343
<i>Sylvain Castagnos and Anne Boyer</i>	
Information Recovery and Discovery in Collaborative Web Search	356
<i>Maurice Coyle and Barry Smyth</i>	
Collaborative Filtering Based on Transitive Correlations Between Items	368
<i>Alexandros Nanopoulos</i>	
Entropy-Based Authorship Search in Large Document Collections	381
<i>Ying Zhao and Justin Zobel</i>	

Topic Identification

Use of Topicality and Information Measures to Improve Document Representation for Story Link Detection	393
<i>Chirag Shah and Koji Eguchi</i>	

Ad Hoc Retrieval of Documents with Topical Opinion 405
Jason Skomorowski and Olga Vechtomova

Expert Finding

Probabilistic Models for Expert Finding 418
Hui Fang and ChengXiang Zhai

Using Relevance Feedback in Expert Search 431
Craig Macdonald and Iadh Ounis

XML IR

Using Topic Shifts for Focussed Access to XML Repositories 444
Elham Ashoori and Mounia Lalmas

Feature- and Query-Based Table of Contents Generation for XML Documents 456
Zoltán Szilávik, Anastasios Tombros, and Mounia Lalmas

Web IR

Setting Per-field Normalisation Hyper-parameters for the Named-Page Finding Search Task..... 468
Ben He and Iadh Ounis

Combining Evidence for Relevance Criteria: A Framework and Experiments in Web Retrieval 481
Theodora Tsirikika and Mounia Lalmas

Multimedia IR

Classifier Fusion for SVM-Based Multimedia Semantic Indexing..... 494
Stéphane Ayache, Georges Quénot, and Jérôme Gensel

Search of Spoken Documents Retrieves Well Recognized Transcripts 505
Mark Sanderson and Xiao Mang Shou

Short Papers

Natural Language Processing for Usage Based Indexing of Web Resources 517
Anne Boyer and Armelle Brun

Harnessing Trust in Social Search 525
Peter Briggs and Barry Smyth

How to Compare Bilingual to Monolingual Cross-Language Information Retrieval..... 533
Franco Crivellari, Giorgio Maria Di Nunzio, and Nicola Ferro

Multilingual Text Classification Using Ontologies 541
Gerard de Melo and Stefan Siersdorfer