

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

Alfred Kobsa

University of California, Irvine, CA, 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

TU Dortmund University, Germany

Madhu Sudan

Microsoft Research, Cambridge, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max Planck Institute for Informatics, Saarbruecken, Germany

Giambattista Amati Fabio Crestani (Eds.)

Advances in Information Retrieval Theory

Third International Conference, ICTIR 2011
Bertinoro, Italy, September 12-14, 2011
Proceedings

Volume Editors

Giambattista Amati
Fondazione Ugo Bordoni
Viale del Policlinico 147, 00161 Rome, Italy
E-mail: gba@fub.it

Fabio Crestani
University of Lugano, Faculty of Informatics
6900 Lugano, Switzerland
E-mail: fabio.crestani@usi.ch

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-23317-3 e-ISBN 978-3-642-23318-0
DOI 10.1007/978-3-642-23318-0
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011934976

CR Subject Classification (1998): H.3, H.2, I.2.3, I.2.6, F.2.2, H.4, H.5.2-4, I.7

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

© Springer-Verlag Berlin Heidelberg 2011

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.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

These proceedings contain the refereed papers and posters presented at the Third International Conference on the Theory of Information Retrieval (ICTIR 2011), held in Bertinoro, Italy, during September 12–14, 2011.

This biennial international conference provides an opportunity for the presentation of the latest work describing theoretical advances in the field of information retrieval (IR). The first ICTIR was held in Budapest in October 2007, organized by Keith van Rijsbergen, Sándor Dominich, Sándor Darányi, and Ferenc Kiss. The second ICTIR was held in Cambridge, UK, in September 2009. It was organized by Leif Azzopardi, Gabriella Kazai, Stephen Robertson, Stefan Rüger, Milad Shokouhi, Dawei Song, and Emine Yilmaz.

ICTIR was brought about by the growing interest in the consecutive workshops run at ACM SIGIR each year from 2000 until 2005 on mathematical and formal methods in IR. This initiative was in large part down to the determination of Sándor Dominich and his passion for all things formal and mathematical. The foundation and the success of ICTIR is a direct result of his commitment and dedication to fostering research and development into the theoretical underpinnings of IR. Sadly, his untimely passing away in 2008 means that he is unable to witness how the theory of IR unfolds in the future. Nonetheless, his belief in the importance of theory and his spirit in advocating the development of formal methods in IR lives on through this conference series. We are glad to continue this initiative and bring ICTIR to Italy, a country with a long series of contributions to the theoretical advances of IR.

The papers accepted for publication and presentation at ICTIR 2011 were selected from a total of 65 submissions. Each submission was subject to a double-blind reviewing process by at least three Program Committee members and was ranked according to its scientific quality, originality, and contribution to the theory of IR. The Committee decided to accept 38 papers, of these 25 (38%) were accepted as full papers and 13 (20%) as short papers. Most of the submitted papers (66%) were about foundations of IR, the remaining submissions were about techniques or topics that refer to wider contexts. The Technical Program thus ranged from topics related to query expansion, co-occurrence analysis, user and interactive modelling, system performance prediction and comparison, probabilistic approaches for ranking and modelling IR to, ultimately, topics related to interdisciplinary approaches or applications.

The program also included two invited talks by ChengXiang Zhai and Keith van Rijsbergen. We would like to thank the invited speakers for their thought-provoking talks on axiomatic analysis and optimization and quantum information theory. We hope they inspired the young as well as the most senior researchers to continue in their steps.

We are grateful to the members of the Program Committee for their time and effort in providing timely and high-quality reviews and feedback to authors. We would also like to thank all the authors who submitted their work for consideration and all the participants and student volunteers for their contributions and help.

Finally, we would like to say special thanks to the following organizations and individuals who helped to make ICTIR 2011 a success:

- The Fondazione Ugo Bordoni for its support and help during the organization phases of the conference.
- The University of Lugano (USI) for providing conference website design and hosting. We especially thank Giacomo Inches and Cristian Bianchi for their work on designing the website.
- Almaxwave for the kind and generous sponsorship.
- Microsoft Research for their continuous sponsorship of the conference.
- Yahoo Research for sponsoring the Best Student Paper Award and the Student Travel Award.
- The editorial staff at Springer for their agreement and assistance in publishing the conference as part of the *Lecture Notes in Computer Science* (LNCS) series.
- EasyChair for the support during the reviewing stages of the submitted papers.

September 2011

Giambattista Amati
Fabio Crestani

Organization

ICTIR 2011 was jointly organized by the Fondazione Ugo Bordoni and the University of Lugano (Università della Svizzera Italiana). It was held at the the University Residential Centre of Bertinoro in Italy.

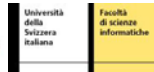
Conference and Program Chairs

Giambattista Amati
Fabio Crestani

Fondazione Ugo Bordoni, Italy
University of Lugano, Switzerland

Sponsors

Fondazione Ugo Bordoni
University of Lugano
Almawave
Microsoft Research
Yahoo! Research



Program Committee

Maristella Agosti	Università degli Studi di Padova, Italy
Alvaro Barreiro	University of A Coruña, Spain
Marco Bianchi	Fondazione Ugo Bordoni, Italy
Roi Blanco	Yahoo! Research, Spain
Paolo Boldi	Università degli Studi di Milano, Italy
Gloria Bordogna	National Research Council, Italy
Giorgio Brajnik	Università degli Studi di Udine, Italy
Peter Bruza	Queensland University of Technology, Australia
Wray Buntine	NICTA and ANU, Australia
Fidel Cacheda	University of A Coruña, Spain
Fazli Can	Bilkent University, Turkey
Mark Carman	Monash University, Australia
Carpineto Claudio	Fondazione Ugo Bordoni, Italy
Stephane Clinchant	Xerox Research Centre Europe, France
Bruce Croft	University of Massachusetts Amherst, USA
Maarten De Rijke	University of Amsterdam, The Netherlands
Arjen De Vries	CWI, The Netherlands
Hui Fang	University of Delaware, USA
Norbert Fuhr	University of Duisburg-Essen, Germany
Giorgio Gambosi	Università di Roma Tor Vergata, Italy
Eric Gaussier	University Joseph Fourier, France
Claudia Hauff	Delft University of Technology, The Netherlands
Ben He	Graduate University of Chinese Academy of Sciences, China
Eduard Hoenkamp	Queensland University of Technology, Australia
Jimmy Huang	York University, Canada
Qiang Huang	University of East Anglia, UK
Theo Huibers	University of Twente, The Netherlands
Giacomo Inches	University of Lugano, Switzerland
Peter Ingwersen	Royal School of Library and Information Science, Denmark
Kalervo Jarvelin	University of Tampere, Finland
Gareth Jones	Dublin City University, Ireland
Joemon Jose	University of Glasgow, UK
Donald Kraft	Louisiana State University, USA
Mounia Lalmas	Yahoo! Research, Spain
Wai Lam	The Chinese University of Hong Kong, China
Monica Landoni	University of Lugano, Switzerland

Raymond Lau	City University of Hong Kong, China
Christina Lioma	University of Stuttgart, Germany
David Losada	University of Santiago de Compostela, Spain
Robert Luk	The Hong Kong Polytechnic University, China
Craig Macdonald	University of Glasgow, UK
Andrew Macfarlane	City University London, UK
Massimo Melucci	Università degli Studi di Padova, Italy
Donald Metzler	University of Southern California, Information Sciences Institute, USA
Marie-Francine Moens	Katholieke Universiteit Leuven, Belgium
Boughanem Mohand	IRIT University Paul Sabatier Toulouse, France
Alessandro Moschitti	Università degli Studi di Trento, Italy
Jian-Yun Nie	Université de Montreal, Canada
Paul Ogilvie	LinkedIn, USA
Iadh Ounis	University of Glasgow, UK
Gabriella Pasi	Università degli Studi di Milano Bicocca, Italy
Vassilis Plachouras	Presans, France
Vijay Raghavan	University of Louisiana at Lafayette, USA
Andreas Rauber	Vienna University of Technology, Austria
Stephen Robertson	Microsoft Research Cambridge, UK
Thomas Roelleke	Queen Mary University of London, UK
Stefan Ruger	Knowledge Media Institute, UK
Ian Ruthven	University of Strathclyde, UK
Jacques Savoy	University of Neuchâtel, Switzerland
Giovanni Semeraro	Università degli Studi di Bari, Italy
Fabrizio Silvestri	National Research Council, Italy
Amanda Spink	Loughborough University, UK
Paul Thomas	CSIRO, Australia
Anastasios Tombros	Queen Mary University of London, UK
Theo Van Der Weide	Radboud University Nijmegen, The Netherlands
Keith Van Rijsbergen	University of Glasgow, UK
Olga Vechtomova	University of Waterloo, Canada
Emine Yilmaz	Microsoft Research Cambridge, UK
Chengxiang Zhai	University of Illinois, Urbana-Champaign, USA
Dell Zhang	Birkbeck College University of London, UK
Peng Zhang	Robert Gordon University, UK
Jianhan Zhu	University College London, UK
Guido Zuccon	University of Glasgow, UK

Additional Reviewers

Bonzanini, Marco
Calegari, Silvia
Caputo, Annalina
Daoud, Mariam
De Vine, Lance
Koopman, Bevan
Kopliku, Kopliku

Martinez-Alvarez, Miguel
Naji, Nada
Ozcan, Rifat
Parapar, Javier
Santos, Rodrygo
Yahyaei, Sirvan
Zhao, Jessie

Table of Contents

Invited Talks

Axiomatic Analysis and Optimization of Information Retrieval Models (Abstract)	1
<i>ChengXiang Zhai</i>	
What Is Quantum Information Retrieval? (Abstract)	2
<i>C.J. Keith van Rijsbergen</i>	

Predicting Query Performance

User Perspectives on Query Difficulty	3
<i>Christina Lioma, Birger Larsen, and Hinrich Schutze</i>	
A Unified Framework for Post-Retrieval Query-Performance Prediction	15
<i>Oren Kurland, Anna Shtok, David Carmel, and Shay Hummel</i>	
Predicting the Performance of Recommender Systems: An Information Theoretic Approach	27
<i>Alejandro Bellogín, Pablo Castells, and Iván Cantador</i>	

Latent Semantic Analysis and Word Co-occurrence Analysis

Trading Spaces: On the Lore and Limitations of Latent Semantic Analysis	40
<i>Eduard Hoenkamp</i>	
Quantum Latent Semantic Analysis	52
<i>Fabio A. González and Juan C. Caicedo</i>	
Pure High-Order Word Dependence Mining via Information Geometry	64
<i>Yueshan Hou, Liang He, Xiaozhao Zhao, and Dawei Song</i>	

Query Expansion and Re-ranking

Promoting Divergent Terms in the Estimation of Relevance Models	77
<i>Javier Parapar and Álvaro Barreiro</i>	
Is Document Frequency Important for PRF?	89
<i>Stéphane Clinchant and Eric Gaussier</i>	

Comparison of Information Retrieval Systems and Approximate Search

Model-Based Inference about IR Systems	101
<i>Ben Carterette</i>	
Selecting a Subset of Queries for Acquisition of Further Relevance Judgements	113
<i>Mehdi Hosseini, Ingemar J. Cox, Natasa Millic-Frayling, Vishwa Vinay, and Trevor Sweeting</i>	
On the Feasibility of Unstructured Peer-to-Peer Information Retrieval	125
<i>H. Asthana, Ruoxun Fu, and Ingemar J. Cox</i>	

Probability Ranking Principle and Alternatives

Can Information Retrieval Systems Be Improved Using Quantum Probability?	139
<i>Massimo Melucci</i>	
An Analysis of Ranking Principles and Retrieval Strategies	151
<i>Guido Zuccon, Leif Azzopardi, and C.J. Keith Van Rijsbergen</i>	
Towards a Better Understanding of the Relationship between Probabilistic Models in IR	164
<i>Robin Aly and Thomas Demeester</i>	

Interdisciplinary Approaches

Cognitive Processes in Query Generation	176
<i>Claudia Hauff and Geert-Jan Houben</i>	
Protocol-Driven Searches for Medical and Health-Sciences Systematic Reviews	188
<i>Matt-Mouley Bouamrane, Craig Macdonald, Iadh Ounis, and Frances Mair</i>	
Enhanced Information Retrieval Using Domain-Specific Recommender Models	201
<i>Wei Li, Debasis Ganguly, and Gareth J.F. Jones</i>	

User and Relevance

Exploring Ant Colony Optimisation for Adaptive Interactive Search	213
<i>M-Dyaa Albakour, Udo Kruschwitz, Nikolaos Nanas, Dawei Song, Maria Fasli, and Anne De Roeck</i>	

Applying the User-over-Ranking Hypothesis to Query Formulation	225
<i>Matthias Hagen and Benno Stein</i>	

How to Count Thumb-Ups and Thumb-Downs: User-Rating Based Ranking of Items from an Axiomatic Perspective	238
<i>Dell Zhang, Robert Mao, Haitao Li, and Joanne Mao</i>	

Result Diversification and Query Disambiguation

Aggregated Search Result Diversification	250
<i>Rodrygo L.T. Santos, Craig Macdonald, and Iadh Ounis</i>	

Topical Categorization of Search Results Based on a Domain Ontology	262
<i>Silvia Calegari, Fabio Farina, and Gabriella Pasi</i>	

Towards Semantic Category Verification with Arbitrary Precision	274
<i>Dmitri Roussinov</i>	

Logical Operators and Descriptive Approaches

Negation for Document Re-ranking in Ad-hoc Retrieval	285
<i>Pierpaolo Basile, Annalina Caputo, and Giovanni Semeraro</i>	

A Descriptive Approach to Classification	297
<i>Miguel Martinez-Alvarez and Thomas Roelleke</i>	

Posters

Do Subtopic Judgments Reflect Diversity?	309
<i>John A. Akinyemi and Charles L.A. Clarke</i>	

On Upper Bounds for Dynamic Pruning	313
<i>Craig Macdonald, Nicola Tonellotto, and Iadh Ounis</i>	

A Comparative Study of Pseudo Relevance Feedback for Ad-hoc Retrieval	318
<i>Kai Hui, Ben He, Tiejian Luo, and Bin Wang</i>	

A Generic Data Model for Schema-Driven Design in Information Retrieval Applications	323
<i>Hany Azzam and Thomas Roelleke</i>	

A Query-Basis Approach to Parametrizing Novelty-Biased Cumulative Gain	327
<i>Teerapong Leelanupab, Guido Zuccon, and Joemon M. Jose</i>	

Investigating Query-Drift Problem from a Novel Perspective of Photon Polarization	332
<i>Peng Zhang, Dawei Song, Xiaozhao Zhao, and Yuerian Hou</i>	
Using Emotion to Diversify Document Rankings	337
<i>Yashar Moshfeghi, Guido Zuccon, and Joemon M. Jose</i>	
Improved Stable Retrieval in Noisy Collections	342
<i>Gianni Amati, Alessandro Celi, Cesidio Di Nicola, Michele Flammini, and Daniela Pavone</i>	
On the Use of Complex Numbers in Quantum Models for Information Retrieval	346
<i>Guido Zuccon, Benjamin Piwowarski, and Leif Azzopardi</i>	
A Query Performance Analysis for Result Diversification	351
<i>Jiyin He, Marc Bron, and Maarten de Rijke</i>	
Rare Disease Diagnosis as an Information Retrieval Task	356
<i>Radu Dragusin, Paula Petcu, Christina Lioma, Birger Larsen, Henrik Jørgensen, and Ole Winther</i>	
Distilling Relevant Documents by Means of Dynamic Quantum Clustering	360
<i>Emanuele Di Buccio and Giorgio Maria Di Nunzio</i>	
Adding Emotions to Pictures	364
<i>Claudia Hauff and Dolf Trieschnigg</i>	
Author Index	369