Lecture Notes in Computer Science 4425

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
Advances in Information Retrieval

29th European Conference on IR Research, ECIR 2007
Rome, Italy, April 2-5, 2007
Proceedings
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 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
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
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 Chiaromella, 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 Fünnkranz, 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, IRIT, 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
# Table of Contents

## Keynote Talks

The Next Generation Web Search and the Demise of the Classic IR Model .......................................................... 1  
*Andrei Broder*

The Last Half-Century: A Perspective on Experimentation in Information Retrieval ............................................ 2  
*Stephen Robertson*

Learning in Hyperlinked Environments ................................. 3  
*Marco Gori*

## Theory and Design

A Parameterised Search System ........................................ 4  
*Roberto Cornacchia and Arjen P. de Vries*

Similarity Measures for Short Segments of Text .................. 16  
*Donald Metzler, Susan Dumais, and Christopher Meek*

Multinomial Randomness Models for Retrieval with Document Fields ........................................................... 28  
*Vassilis Plachouras and Iadh Ounis*

On Score Distributions and Relevance ................................ 40  
*Stephen Robertson*

Modeling Term Associations for Ad-Hoc Retrieval Performance Within Language Modeling Framework ................. 52  
*Xing Wei and W. Bruce Croft*

## Efficiency

Static Pruning of Terms in Inverted Files ........................... 64  
*Roi Blanco and Álvaro Barreiro*

Efficient Indexing of Versioned Document Sequences ............ 76  
*Michael Herscovici, Ronny Lempel, and Sivan Yogev*

Light Syntactically-Based Index Pruning for Information Retrieval .... 88  
*Christina Lioma and Iadh Ounis*

Sorting Out the Document Identifier Assignment Problem ........ 101  
*Fabrizio Silvestri*

Efficient Construction of FM-index Using Overlapping Block Processing for Large Scale Texts ........................................ 113  
*Di Zhang, Yunquan Zhang, and Jing Chen*
# Table of Contents

## 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 Sébille*

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*
<table>
<thead>
<tr>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevance Feedback Using Weight Propagation Compared with Information-Theoretic Query Expansion</td>
<td>258</td>
</tr>
<tr>
<td><em>Fadi Yamout, Michael Oakes, and John Tait</em></td>
<td></td>
</tr>
<tr>
<td>Evaluation</td>
<td></td>
</tr>
<tr>
<td>A Retrieval Evaluation Methodology for Incomplete Relevance Assessments</td>
<td>271</td>
</tr>
<tr>
<td><em>Mark Baillie, Leif Azzopardi, and Ian Ruthven</em></td>
<td></td>
</tr>
<tr>
<td>Evaluating Query-Independent Object Features for Relevancy Prediction</td>
<td>283</td>
</tr>
<tr>
<td><em>Andres R. Masegosa, Hideo Joho, and Joemon M. Jose</em></td>
<td></td>
</tr>
<tr>
<td>Classification and Clustering</td>
<td></td>
</tr>
<tr>
<td>The Utility of Information Extraction in the Classification of Books</td>
<td>295</td>
</tr>
<tr>
<td><em>Tom Betts, Maria Milosevijevic, and Jon Oberlander</em></td>
<td></td>
</tr>
<tr>
<td>Combined Syntactic and Semantic Kernels for Text Classification</td>
<td>307</td>
</tr>
<tr>
<td><em>Stephan Bloehdorn and Alessandro Moschitti</em></td>
<td></td>
</tr>
<tr>
<td>Fast Large-Scale Spectral Clustering by Sequential Shrinkage</td>
<td>319</td>
</tr>
<tr>
<td><em>Tie-Yan Liu, Huai-Yuan Yang, Xin Zheng, Tao Qin, and Wei-Ying Ma</em></td>
<td></td>
</tr>
<tr>
<td>A Probabilistic Model for Clustering Text Documents with Multiple Fields</td>
<td>331</td>
</tr>
<tr>
<td><em>Shanfeng Zhu, Ichigaku Takigawa, Shuqin Zhang, and Hiroshi Mamitsuka</em></td>
<td></td>
</tr>
<tr>
<td>Filtering</td>
<td></td>
</tr>
<tr>
<td>Personalized Communities in a Distributed Recommender System</td>
<td>343</td>
</tr>
<tr>
<td><em>Sylvain Castagnos and Anne Boyer</em></td>
<td></td>
</tr>
<tr>
<td>Information Recovery and Discovery in Collaborative Web Search</td>
<td>356</td>
</tr>
<tr>
<td><em>Maurice Coyle and Barry Smyth</em></td>
<td></td>
</tr>
<tr>
<td>Collaborative Filtering Based on Transitive Correlations Between Items</td>
<td>368</td>
</tr>
<tr>
<td><em>Alexandros Nanopoulos</em></td>
<td></td>
</tr>
<tr>
<td>Entropy-Based Authorship Search in Large Document Collections</td>
<td>381</td>
</tr>
<tr>
<td><em>Ying Zhao and Justin Zobel</em></td>
<td></td>
</tr>
<tr>
<td>Topic Identification</td>
<td></td>
</tr>
<tr>
<td>Use of Topicality and Information Measures to Improve Document Representation for Story Link Detection</td>
<td>393</td>
</tr>
<tr>
<td><em>Chirag Shah and Koji Eguchi</em></td>
<td></td>
</tr>
</tbody>
</table>
**XVI Table of Contents**

**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 Szlávik, 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 Tsikrika 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*
Using Visual-Textual Mutual Information and Entropy for Inter-modal Document Indexing ................................. 549
Jean Martinet and Shin’ichi Satoh

A Study of Global Inference Algorithms in Multi-document Summarization ......................................................... 557
Ryan McDonald

Document Representation Using Global Association Distance Model ......................................................... 565
José E. Medina-Pagola, Ansel Y. Rodríguez, Abdel Hechavarría, and José Hernández Palancar

Sentence Level Sentiment Analysis in the Presence of Conjuncts Using Linguistic Analysis ........................................ 573
Arun Meena and T.V. Prabhakar

PageRank: When Order Changes ........................................ 581
Massimo Melucci and Luca Pretto

Model Tree Learning for Query Term Weighting in Question Answering ..................................................... 589
Christof Monz

Examining Repetition in User Search Behavior ........................................ 597
Mark Sanderson and Susan Dumais

Popularity Weighted Ranking for Academic Digital Libraries .......... 605
Yang Sun and C. Lee Giles

Naming Functions for the Vector Space Model ........................................ 613
Yannis Tzitzikas and Yannis Theoharis

Effective Use of Semantic Structure in XML Retrieval ................. 621
Roelof van Zwol and Tim van Loosbroek

Searching Documents Based on Relevance and Type ................. 629
Jun Xu, Yunbo Cao, Hang Li, Nick Craswell, and Yalou Huang

Investigation of the Effectiveness of Cross-Media Indexing ............. 637
Murat Yakıcı and Fabio Crestani

Improve Ranking by Using Image Information ................................. 645
Qing Yu, Shuming Shi, Zhiwei Li, Ji-Rong Wen, and Wei-Ying Ma

N-Step PageRank for Web Search ........................................ 653
Li Zhang, Tao Qin, Tie-Yan Liu, Ying Bao, and Hang Li

Authorship Attribution Via Combination of Evidence ................. 661
Ying Zhao and Phil Vines

Posters

Cross-Document Entity Tracking ........................................ 670
Roxana Angheluta and Marie-Francine Moens
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enterprise People and Skill Discovery Using Tolerant Retrieval and Vision</td>
<td>674</td>
</tr>
<tr>
<td>Jan Brunnert, Omar Alonso, and Dirk Riehle</td>
<td></td>
</tr>
<tr>
<td>Experimental Results of the Signal Processing Approach to Distributional Clustering of Terms on Reuters-21578 Collection</td>
<td>678</td>
</tr>
<tr>
<td>Marta Capdevila Dalmau and Oscar W. Márquez Flórez</td>
<td></td>
</tr>
<tr>
<td>Overall Comparison at the Standard Levels of Recall of Multiple Retrieval Methods with the Friedman Test</td>
<td>682</td>
</tr>
<tr>
<td>José M. Casanova, Manuel A. Presedo Quindimil, and Álvaro Barreiro</td>
<td></td>
</tr>
<tr>
<td>Building a Desktop Search Test-Bed</td>
<td>686</td>
</tr>
<tr>
<td>Sergey Chernov, Pavel Serdyukov, Paul-Alexandru Chirita, Gianluca Demartini, and Wolfgang Nejdl</td>
<td></td>
</tr>
<tr>
<td>Hierarchical Browsing of Video Key Frames</td>
<td>691</td>
</tr>
<tr>
<td>Gianluigi Ciocca and Raimondo Schettini</td>
<td></td>
</tr>
<tr>
<td>Active Learning with History-Based Query Selection for Text Categorisation</td>
<td>695</td>
</tr>
<tr>
<td>Michael Davy and Saturnino Luz</td>
<td></td>
</tr>
<tr>
<td>Fighting Link Spam with a Two-Stage Ranking Strategy</td>
<td>699</td>
</tr>
<tr>
<td>Guang-Gang Geng, Chun-Heng Wang, Qiu-Dan Li, and Yuan-Ping Zhu</td>
<td></td>
</tr>
<tr>
<td>Improving Naive Bayes Text Classifier Using Smoothing Methods</td>
<td>703</td>
</tr>
<tr>
<td>Feng He and Xiaoqing Ding</td>
<td></td>
</tr>
<tr>
<td>Term Selection and Query Operations for Video Retrieval</td>
<td>708</td>
</tr>
<tr>
<td>Bouke Huurnink and Maarten de Rijke</td>
<td></td>
</tr>
<tr>
<td>An Effective Threshold-Based Neighbor Selection in Collaborative Filtering</td>
<td>712</td>
</tr>
<tr>
<td>Taek-Hun Kim and Sung-Bong Yang</td>
<td></td>
</tr>
<tr>
<td>Combining Multiple Sources of Evidence in XML Multimedia Documents: An Inference Network Incorporating Element Language Models</td>
<td>716</td>
</tr>
<tr>
<td>Zhigang Kong and Mounia Lalmas</td>
<td></td>
</tr>
<tr>
<td>Language Model Based Query Classification</td>
<td>720</td>
</tr>
<tr>
<td>Andreas Merkel and Dietrich Klakow</td>
<td></td>
</tr>
<tr>
<td>Integration of Text and Audio Features for Genre Classification in Music Information Retrieval</td>
<td>724</td>
</tr>
<tr>
<td>Robert Neumayer and Andreas Rauber</td>
<td></td>
</tr>
<tr>
<td>Retrieval Method for Video Content in Different Format Based on Spatiotemporal Features</td>
<td>728</td>
</tr>
<tr>
<td>Xuefeng Pan, Jintao Li, Yongdong Zhang, Sheng Tang, and Juan Cao</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Combination of Document Priors in Web Information Retrieval</td>
<td>732</td>
</tr>
<tr>
<td>Jie Peng and Iadh Ounis</td>
<td></td>
</tr>
<tr>
<td>Enhancing Expert Search Through Query Modeling</td>
<td>737</td>
</tr>
<tr>
<td>Pavel Serdyukov, Sergey Chernov, and Wolfgang Nejdl</td>
<td></td>
</tr>
<tr>
<td>A Hierarchical Consensus Architecture for Robust Document Clustering</td>
<td>741</td>
</tr>
<tr>
<td>Xavier Sevillano, Germán Cobo, Francesc Alías, and Joan Claudi Socoró</td>
<td></td>
</tr>
<tr>
<td>Summarisation and Novelty: An Experimental Investigation</td>
<td>745</td>
</tr>
<tr>
<td>Simon Sweeney, Fabio Crestani, and David E. Losada</td>
<td></td>
</tr>
<tr>
<td>A Layered Approach to Context-Dependent User Modelling</td>
<td>749</td>
</tr>
<tr>
<td>Elena Vildjiounaite and Sanna Kallio</td>
<td></td>
</tr>
<tr>
<td>A Bayesian Approach for Learning Document Type Relevance</td>
<td>753</td>
</tr>
<tr>
<td>Peter C.K. Yeung, Stefan Büttcher, Charles L.A. Clarke, and Maheedhar Kolla</td>
<td></td>
</tr>
<tr>
<td>Author Index</td>
<td>757</td>
</tr>
</tbody>
</table>