

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

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 Alvaro 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, Yunqnan Zhang, and Jing Chen</i>	

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

Performance Comparison of Clustered and Replicated Information Retrieval Systems	124
<i>Fidel Cacheda, Victor Carneiro, Vassilis Plachouras, and Iadh Ounis</i>	
A Study of a Weighting Scheme for Information Retrieval in Hierarchical Peer-to-Peer Networks	136
<i>Massimo Melucci and Alberto Poggiani</i>	
A Decision-Theoretic Model for Decentralised Query Routing in Hierarchical Peer-to-Peer Networks	148
<i>Henrik Nottelmann and Nbrbert Fuhr</i>	
Central-Rank-Based Collection Selection in Uncooperative Distributed Information Retrieval	160
<i>Milad Shokouhi</i>	

!

Result Merging

Results Merging Algorithm Using Multiple Regression Models	173
<i>George Paltoglou, Michail Salampasis, and Maria Satratzemi</i>	
Segmentation of Search Engine Results for Effective Data-Fusion	185
<i>Milad Shokouhi</i>	

Queries

Query Hardness Estimation Using Jensen-Shannon Divergence Among Multiple Scoring Functions	198
<i>Javed A. A slam and Virgil Pavlu</i>	
Query Reformulation and Refinement Using NLP-Based Sentence Clustering	210
<i>Frederic Roulland, Aaron Kaplan, Stefania, Castellani, Claude Roux, Antonietta Grasso, Karin Pettersson, and, Jacki O'Neill</i>	
Automatic Morphological Query Expansion Using Analogy-Based Machine Learning	222
<i>Fabienne Moreau, Vincent Claveau, and Pascale Sebillot</i>	
Advanced Structural Representations for Question Classification and Answer Re-ranking	234
<i>Silvia Quarteroni, Alessandro Moschitti, Suresh Manandhar, and, Roberto Basili</i>	

Relevance Feedback

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

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

Evaluation

A Retrieval Evaluation Methodology for Incomplete Relevance Assessments	271
<i>Mark Baillie, Lei Azzopardi, and Ian Ruthven</i>	
Evaluating Query-Independent Object Features for Relevancy Prediction	283
<i>Andres R. Masegosa, Ildejo Joho, and Joernon Ad. 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>Sylvaw, 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>Alexandras 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
<i>Jason Skomorowski and, Olga, Vechtomova</i>	
Expert Finding	
Probabilistic Models for Expert Finding	418
<i>Hui Fang and ChengXiang Zhai</i>	
Using Relevance Feedback in Expert Search	431
<i>Craig Macdonald, and Iadh Owens</i>	
XML IR	
Using Topic Shifts for Focussed Access to XML Repositories.	444
<i>Elham Ashoori and Mounia Lalmas</i>	
Feature- and Query-Based Table of Contents Generation for XML Documents	456
<i>Zoltan Szldvik, Anastasios Tombros, and Mounia Laimas</i>	
Web IR	
Setting Per-field Normalisation Hyper-parameters for the Named-Page Finding Search Task	468
<i>Ben He and Iadh Ounis</i>	
Combining Evidence for Relevance Criteria: A Framework and Experiments in Web Retrieval	481
<i>Theodora Tsikrika and Mounia Lalmas</i>	
Multimedia IR	
Classifier Fusion for SVM-Based Multimedia Semantic Indexing	494
<i>Stéphane Ayache, Georges Quenot, and Jerome Gensel</i>	
Search of Spoken Documents Retrieves Well Recognized Transcripts	505
<i>Mark Sanderson and Xiao Mang Shou</i>	
Short Papers	
Natural Language Processing for Usage Based Indexing of Web Resources	517
<i>Anne Boyer and Armelle Brim,</i>	
Harnessing Trust in Social Search	525
<i>Peter Briggs and Barry Smyth</i>	
How to Compare Bilingual to Monolingual Cross-Language Information Retrieval	533
<i>Franco Crivellan, Giorgio Maria Di Nunzio, and, Nicola, Ferro</i>	
Multilingual Text Classification Using Ontologies	541
<i>Gerard de Melo and Stefan Siersdorfer</i>	

Using Visual-Textual Mutual Information and Entropy for Inter-modal Document Indexing	549
<i>Jean Martinet and Shin'ichi Satoh</i>	
A Study of Global Inference Algorithms in Multi-document Summarization	557
<i>Ryan McDonald</i>	
Document Representation Using Global Association Distance Model	565
<i>Jose E. Medina-Pagola, Ansel Y. Rodriguez, Abdel Hechavarria, and Jose Hernandez Palancor</i>	
Sentence Level Sentiment Analysis in the Presence of Conjuncts Using Linguistic Analysis.	573
<i>Arun Meena and T. V. Prabhakar</i>	
PageRank: When Order Changes.	581
<i>Massimo Melucci and Luca Pretto</i>	
Model Tree Learning for Query Term Weighting in Question Answering.	589
<i>Christof Monz</i>	
Examining Repetition in User Search Behavior.	597
<i>Mark Sanderson and, Susan Durnais-</i>	
Popularity Weighted Ranking for Academic Digital Libraries	605
<i>Yang Sun and C. Lee Giles</i>	
Naming Functions for the Vector Space Model.	613
<i>Yannis Tzitzikas and Yannis Theoharis</i>	
Effective Use of Semantic Structure in XML Retrieval.	621
<i>Roelof van Zwol and Tim, van Loosbroek</i>	
Searching Documents Based on Relevance and Type.	629
<i>Jim Xu, Yunbo Cao, Hang Li, Nick' Craswell, and Yalou Huang</i>	
Investigation of the Effectiveness of Cross-Media Indexing.	637
<i>Murat Yakici and, Fabio Crestani</i>	
Improve Ranking by Using Image Information.	645
<i>Qing Yu, Shaming Shi, Zhiwei Li, Ji-Rong Wen, and, Wei- Yinq Ma</i>	
<i>N-Step PageRank for Web Search.</i>	653
<i>Li Zhang, Tao Qin, Tie-Yan Liu, Ying Bao, and, Hang Li</i>	
Authorship Attribution Via Combination of Evidence.	661
<i>Ying Zhao and, Phil Vines</i>	
 Posters	
Cross-Document Entity Tracking	670
<i>Roxana Angheluta and, Marie-Francine Moens</i>	

Enterprise People and Skill Discovery Using Tolerant Retrieval and Visualization	» 674
<i>Jan Brunnerl, Omar' Alonso, and, Dirk Riehle</i>	
Experimental Results of the Signal Processing Approach to Distributional Clustering of Terms on <i>Reuters-21578</i> Collection.	678
<i>Marta Capdevila Dalmau and Oscar W. Mdrquez Florez</i>	
Overall Comparison at the Standard Levels of R.ecall of Multiple Retrieval Methods with the Friedman Test . . . ,	682
<i>Jose M. Casanova, Manuel A. Presedo Quindimil, and Alvaro Barreiro</i>	
Building a Desktop Search Test-Bed	686
<i>Sergey Chernov, Pavel Serdyukov, Paul-Alexandrii Chirita, Gianluca Demartini, and Wolfgang Nejdl</i>	
Hierarchical Browsing of Video Key Frames.	691
<i>Gianluigi Ciocca and Raimondo Schettini</i>	
Active Learning with History-Based Query Selection for Text Categorisation	695
<i>Michael Davy and, Saturnino Luz</i>	
Fighting Link Spam with a Two-Stage Ranking Strategy.	699
<i>Guang-Gang Geng, Chun-Heng Wang, Qiu-Dan Li. and Yuan-Ping Zhu</i>	
Improving Naive Bayes Text Classifier Using Smoothing Methods.	703
<i>Feng He and Xiaoqing Ding</i>	
Term Selection and Query Operations for Video Retrieval.	708
<i>Bouke Huurnmk and, Maarlen de Rijke</i>	
An Effective Threshold-Based Neighbor Selection in Collaborative Filtering	712
<i>Ta.ek-Hun Kim and Sung-Bong Yang</i>	
Combining Multiple Sources of Evidence in XML Multimedia Documents: An Inference Network Incorporating Element Language Models	716
<i>Zhigang Kong and Mounia Lalmas</i>	
Language Model Based Query Classification	720
<i>Andreas Merkel and Dietrich Klakow</i>	
Integration of Text and Audio Features for Genre Classification in Music Information Retrieval.	724
<i>Robert Neumayer and Andreas Rauber</i>	
Retrieval Method for Video Content in Different Format Based on Spatiotemporal Features.	728
<i>Xuefeng Pan, Jmtao Li, Yongdong Zhang, Sheng Tang, and, Juan Cao</i>	

Combination of Document Priors in Web Information Retrieval	732
<i>Jie Peng and, Iadh Ounis</i>	
Enhancing Expert Search Through Query Modeling	737
<i>Pavel Serdyukov, Sergey Chernov, and Wolfgang Nejdl</i>	
A Hierarchical Consensus Architecture for Robust Document Clustering	741
<i>Xavier Sevillano, German Cobo, Francesc Alias, and Joan Claudi Socoro</i>	
Summarisation and Novelty: An Experimental Investigation	745
<i>Simon Sweeney, Fabio Crestani, and, David E. Losada</i>	
A Layered Approach to Context-Dependent User Modelling	749
<i>Elena Vildjiounaite and Sanna Kallio</i>	
A Bayesian Approach for Learning Document Type Relevance	753
<i>Peter C.K. Yeung, Stefan Buttcher, Charles L.A. Clarke, and Maheedhar Kolla</i>	
Author Index	757