

Fabrizio Sebastiani (Ed.)

Advances in Information Retrieval

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



Springer

Table of Contents

Invited Papers

- Document Retrieval: Shallow Data, Deep Theories;
Historical Reflections, Potential Directions 1
Karen Spärck Jones
- Annotation and Retrieval of Structured Video Documents 12
Marco Bertini, Alberto Del Bimbo, and Walter Nunziati

Papers

IR and the Web

- Improving the Evaluation of Web Search Systems 25
Cathal Gurrin and Alan F. Smeaton
- When Are Links Useful? Experiments in Text Classification. 41
Michelle Fisher and Richard Everson
- Hierarchical Classification of HTML Documents with WebClassII 57
Michelangelo Ceci and Donato Malerba

Retrieval of Structured Documents

- Hierarchical Indexing and Flexible Element Retrieval
for Structured Document 73
Hang Cui, Ji-Rong Wen, and Tat-Seng Chua
- Construction of a Test Collection for the Focussed Retrieval
of Structured Documents 88
Gabriella Kazai, Mounia Lalmas, and Jane Reid
- User Behaviour in the Context of Structured Documents 104
Karen Finesilver and Jane Reid

Collaborative Filtering and Text Mining

Attaining Fast and Successful Searches in E-commerce Environments	120
<i>Raz Lin, Sarit Kraus, and Jeffrey Tew</i>	
Learning User Similarity and Rating Style for Collaborative Recommendation	135
<i>Lily F. Tian and Kwok-Wai Cheung</i>	
Spoken Information Extraction from Italian Broadcast News	146
<i>Vanessa Sandrini and Marcello Federico</i>	

Text Representation and Natural Language Processing

Taming Wild Phrases	161
<i>Cornelis H. A. Koster and Mark Seutter</i>	
Stemming and Decompounding for German Text Retrieval	177
<i>Martin Braschler and Bärbel Ripplinger</i>	
Question Answering System for Incomplete and Noisy Data: Methods and Measures for Its Evaluation	193
<i>Lili Aunimo, Oskari Heinonen, Reeta Kuuskoski, Juha Makkonen, Renaud Petit, and Otso Virtanen</i>	
Term Proximity Scoring for Keyword-Based Retrieval Systems	207
<i>Yves Rasolofo and Jacques Savoy</i>	

Formal Models and Language Models for IR

Propositional Logic Representations for Documents and Queries: A Large-Scale Evaluation	219
<i>David E. Losada and Alvaro Barreiro</i>	
From Uncertain Inference to Probability of Relevance for Advanced IR Applications	235
<i>Henrik Nottelmann and Norbert Fuhr</i>	
Topic Detection and Tracking with Spatio-Temporal Evidence	251
<i>Juha Makkonen, Helena Ahonen-Myka, and Marko Salmenkivi</i>	

Machine Learning and IR

Clustering and Visualization in a Multi-lingual Multi-document Summarization System	266
<i>Hsin-Hsi Chen, June-Jei Kuo, and Tsei-Chun Su</i>	
A Hybrid Relevance-Feedback Approach to Text Retrieval	281
<i>Zhao Xu, Xiaowei Xu, Kai Yu, and Volker Tresp</i>	
Experiments with Document Archive Size Detection	294
<i>Shengli Wu, Forbes Gibb, and Fabio Crestani</i>	

Text Categorization 1

- Using Kullback-Leibler Distance for Text Categorization 305
Brigitte Bigi
- Discretizing Continuous Attributes in AdaBoost for Text Categorization .. 320
Pio Nardiello, Fabrizio Sebastiani, and Alessandro Sperduti
- Combining Naive Bayes and n -Gram Language Models
 for Text Classification 335
Fuchun Peng and Dale Schuurmans

Usability, Interactivity, and Visualization Issues in IR

- WebDocBall: A Graphical Visualization Tool for Web Search Results 351
Jesús Vegas, Pablo de la Fuente, and Fabio Crestani
- Relevance Feedback for Content-Based Image Retrieval:
 What Can Three Mouse Clicks Achieve? 363
Daniel C. Heesch and Stefan Rürger
- Query-Based Document Skimming: A User-Centred Evaluation
 of Relevance Profiling 377
David J. Harper, Ivan Koychev, and Yixing Sun

Text Categorization 2

- Representative Sampling for Text Classification
 Using Support Vector Machines 393
Zhao Xu, Kai Yu, Volker Tresp, Xiaowei Xu, and Jizhi Wang
- Chinese Text Categorization Based on the Binary Weighting Model
 with Non-binary Smoothing 408
Xue Dejun and Sun Maosong
- A Study on Optimal Parameter Tuning for Rocchio Text Classifier 420
Alessandro Moschitti

Architectural Issues and Efficiency Issues in IR

- Optimization of Restricted Searches in Web Directories
 Using Hybrid Data Structures 436
Fidel Cacheda, Victor Carneiro, Carmen Guerrero, and Angel Viña
- Similarity Join in Metric Spaces 452
Vlastislav Dohnal, Claudio Gennaro, Pasquale Savino, and Pavel Zezula
- An Efficient Compression Code for Text Databases 468
*Nieves R. Brisaboa, Eva L. Iglesias, Gonzalo Navarro,
 and José R. Paramá*

Posters

Compressing Semistructured Text Databases	482
<i>Joaquín Adiego, Gonzalo Navarro, and Pablo de la Fuente</i>	
Vertical Searching in Juridical Digital Libraries	491
<i>Maria de Lourdes da Silveira, Berthier Ribeiro-Neto, Rodrigo de Freitas Vale, and Rodrigo Tôrres Assumpção</i>	
Corpus-Based Thesaurus Construction for Image Retrieval in Specialist Domains	502
<i>Khurshid Ahmad, Mariam Tariq, Bogdan Vrusias, and Chris Handy</i>	
Generating Extracts with Genetic Algorithms	511
<i>Enrique Alfonseca and Pilar Rodríguez</i>	
The ITC-irst News on Demand Platform	520
<i>Nicola Bertoldi, Fabio Brugnara, Mauro Cettolo, Marcello Federico, Diego Giuliani, Erwin Leeuwis, and Vanessa Sandrini</i>	
Evaluating Peer-to-Peer Networking for Information Retrieval within the Context of Meta-searching	528
<i>Iraklis A. Klampanos, James J. Barnes, and Joemon M. Jose</i>	
Parallel Computing for Term Selection in Routing/Filtering	537
<i>Andy MacFarlane, Stephen E. Robertson, and Julie A. McCann</i>	
A Weighting Scheme for Star-Graphs	546
<i>Jean Martinet, Iadh Ounis, Yves Chiaramella, and Philippe Mulhem</i>	
Phrase-Based Hierarchical Clustering of Web Search Results	555
<i>Irmina Mastowska</i>	
Aggregated Feature Retrieval for MPEG-7	563
<i>Jiamin Ye and Alan F. Smeaton</i>	
Document Retrieval in the Context of Question Answering	571
<i>Christof Monz</i>	
A Study of the Usefulness of Institutions' Acronyms as Web Queries	580
<i>Sándor Dominich, Júlia Góth, and Adrienn Skrop</i>	
Building a Hierarchy of Events and Topics for Newspaper Digital Libraries	588
<i>Aurora Pons-Porrata, Rafael Berlanga-Llavori, and José Ruiz-Shulcloper</i>	
A Machine Learning Approach for the Curation of Biomedical Literature ..	597
<i>Min Shi, David S. Edwin, Rakesh Menon, Lixiang Shen, Jonathan Y.K. Lim, Han Tong Loh, S. Sathiya Keerthi, and Chong Jin Ong</i>	
Automatic Construction of Theme Melody Index from Music Database for Fast Content-Based Retrievals	605
<i>Chang-Hwan Shin, Kyong-I Ku, KiChang Kim, and Yoo-Sung Kim</i>	

A Personalized Information Search Process Based
on Dialoguing Agents and User Profiling613
*Giovanni Semeraro, Marco Degemmis, Pasquale Lops, Ulrich Thiel,
and Marcello L'Abbate*

Author Index623