

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

University of Dortmund, 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 of Computer Science, Saarbruecken, Germany

Gary Geunbae Lee Dawei Song
Chin-Yew Lin Akiko Aizawa
Kazuko Kuriyama Masaharu Yoshioka
Tetsuya Sakai (Eds.)

Information Retrieval Technology

5th Asia Information Retrieval Symposium, AIRS 2009
Sapporo, Japan, October 21-23, 2009
Proceedings

Volume Editors

Gary Geunbae Lee

Pohang University of Science and Technology, E-mail: gblee@postech.ac.kr

Dawei Song

The Robert Gordon University, E-mail: d.song@rgu.ac.uk

Chin-Yew Lin

Microsoft Research Asia, E-mail: cyl@microsoft.com

Akiko Aizawa

National Institute of Informatics, E-mail: aizawa@nii.ac.jp

Kazuko Kuriyama

Shirayuri College, E-mail: kuriyama@shirayuri.ac.jp

Masaharu Yoshioka

Hokkaido University, E-mail: yoshioka@ist.hokudai.ac.jp

Tetsuya Sakai

Microsoft Research Asia, E-mail: tetsuyasakai@acm.org

Library of Congress Control Number: 2009935701

CR Subject Classification (1998): H.3, H.4, F.2.2, E.1, E.2

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

ISSN 0302-9743

ISBN-10 3-642-04768-8 Springer Berlin Heidelberg New York

ISBN-13 978-3-642-04768-8 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.com

© Springer-Verlag Berlin Heidelberg 2009

Printed in Germany

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

Printed on acid-free paper SPIN: 12769179 06/3180 5 4 3 2 1 0

Preface

Asia Information Retrieval Symposium (AIRS) 2009 was the fifth AIRS conference in the series established in 2004. The first AIRS was held in Beijing, China, the second in Jeju, Korea, the third in Singapore and the fourth in Harbin, China. The AIRS conferences trace their roots to the successful Information Retrieval with Asian Languages (IRAL) workshops, which started in 1996.

The AIRS series aims to bring together international researchers and developers to exchange new ideas and the latest results in information retrieval. The scope of the conference encompassed the theory and practice of all aspects of information retrieval in text, audio, image, video, and multimedia data.

AIRS 2009 received 82 submissions, from which we carefully selected 18 regular papers (22%) and 20 (24%) posters through a double-blind reviewing process. We are pleased to report that the conference proceedings include contributions from not only Asian countries, but also from Finland, Italy, Australia, UK and USA.

We are grateful to Elizabeth Liddy, chair of ACM SIGIR, for accepting to be the honorary conference chair, and to Hokkaido University for hosting the conference. We thank the Information Retrieval Facility, Microsoft Research Asia, Ricoh, Ltd., Global COE Program “Center for Next-Generation Information Technology Based on Knowledge Discovery and Knowledge Federation” and Sapporo International Communication Plaza Foundation for sponsoring the conference, and Springer for publishing the conference proceedings as part of their Lecture Notes in Computer Science (LNCS) series. We also thank ACM SIGIR and IPSJ SIGFI for giving the conference an “in cooperation with” status.

We thank the publication co-chairs, Akiko Aizawa and Kazuko Kuriyama, for compiling the camera-ready papers and liaising with Springer, and the finance chair, Takuya Kida, for successfully managing all money matters. We also thank the publicity co-chairs, Atsushi Fujii, Ian Soboroff, Dawei Song and William Webber, for advertising AIRS, and the area chairs and other program committee members for their high-quality and punctual reviews.

Finally, we acknowledge, and are inspired by, the many authors who submitted papers and who continue to contribute to this Asian community of IR research and development.

August 2009

Gary Geunbae Lee
Dawei Song
Chin-Yew Lin
Masaharu Yoshioka
Tetsuya Sakai

Organization

Steering Committee

Hsin-Hsi Chen	National Taiwan University, Taiwan
Wai Lam	The Chinese University of Hong Kong, China
Gary Geunbae Lee	Pohang University of Science and Technology, Korea
Alistair Moffat	University of Melbourne, Australia
Hwee Tou Ng	National University of Singapore, Singapore
Tetsuya Sakai	Microsoft Research Asia, China
Dawei Song	The Robert Gordon University, UK
Masaharu Yoshioka	Hokkaido University, Japan

Advisory Board

Mun Kew Leong	National Library Board, Singapore
Sung Hyon Myaeng	Information and Communication University, Korea
Kam-Fai Wong	The Chinese University of Hong Kong, China

Honorary Conference Chair

Elizabeth Liddy	Syracuse University, USA
-----------------	--------------------------

General Co-chairs

Tetsuya Sakai	Microsoft Research Asia, China
Masaharu Yoshioka	Hokkaido University, Japan

Program Committee Co-chairs

Gary Geunbae Lee	Pohang University of Science and Technology, Korea
Chin-Yew Lin	Microsoft Research Asia, China
Dawei Song	The Robert Gordon University, UK

Organization Committee

Finance Chair	Takuya Kida, Hokkaido University, Japan
Publication Co-chair	Akiko Aizawa, National Institute of Informatics, Japan Kazuko Kuriyama, Shirayuri College, Japan
Publicity Co-chair	Atsushi Fujii, University of Tsukuba, Japan Ian Soboroff, National Institute of Standards and Technology, USA Dawei Song, The Robert Gordon University, UK William Webber, University of Melbourne, Australia

Area Chairs

Gareth Jones	Dublin City University, Ireland
Tie-Yan Liu	Microsoft Research Asia, China
David Losada	Universidad de Santiago de Compostela, Spain
Teruko Mitamura	Carnegie Mellon University, USA
Ian Ruthven	University of Strathclyde, UK
Andrew Turpin	RMIT, Australia
Min Zhang	Tsinghua University, China

Hosted by

Hokkaido University, Japan

Program Committee

Gianni Amati	Fondazione Ugo Bordoni, Italy
Leif Azzopardi	University of Glasgow, UK
Peter Bailey	Microsoft Research, USA
Mark Baillie	University of Strathclyde, UK
Alvaro Barreiro	University of A Coruña, Spain
Bodo Billerbeck	RMIT University, Australia
Roi Blanco	University of A Coruña, Spain
Rui Cai	Microsoft Research Asia, China
Enhong Chen	University of Science and Technology, China
Hsin-Hsi Chen	National Taiwan University, Taiwan
Xuqi Cheng	Chinese Academy of Sciences, China
Seungjin Choi	Pohang University of Science and Technology, Korea

Nick Craswell	Microsoft Research Cambridge, UK
Shane Culpepper	RMIT University, Australia
Efthimis N. Efthimiadis	University of Washington, USA
David Elswailer	Friedrich-Alexander-University of Erlangen-Nürnberg, Germany
Juan M. Fernández Luna	University of Granada, Spain
Colum Foley	Dublin City University, Ireland
Bin Gao	Microsoft Research Asia, China
Fred Gey	University of California, Berkeley, USA
Daqing He	University of Pittsburgh, USA
Ichiro Ide	Nagoya University, Japan
Hideo Joho	University of Glasgow, UK
Hiroshi Kanayama	Tokyo Research Laboratory, IBM, Japan
Jaana Kekäläinen	University of Tampere, Finland
Jeongwoo Ko	Google, USA
Wei Lai	Microsoft Research Asia, China
Mounia Lalmas	University of Glasgow, UK
Wai Lam	Chinese University of Hong Kong, Hong Kong
Ni Lao	Carnegie Mellon University, USA
Martha Larson	Technical University of Delft, The Netherlands
Ting Liu	Harbin Institute of Technology, China
Yiqun Liu	Tsinghua University, China
Andrew MacFarlane	City University of London, UK
Stephane Marchand-Maillet	University of Geneva, Switzerland
Massimo Melucci	University of Padova, Italy
Helen Meng	The Chinese University of Hong Kong, Hong Kong
Tatsunori Mori	Yokohama National University, Japan
Alessandro Moschitti	University of Trento, Italy
Jian-Yun Nie	University of Montreal, Canada
Eric Nyberg	Carnegie Mellon University, USA
Neil O'Hare	Dublin City University, Ireland
Nils Pharo	Oslo University College, Norway
Benjamin Piwowarski	University of Glasgow, UK
Dulce Ponceleón	IBM Almaden Research, USA
Thomas Roelleke	Queen Mary University of London, UK
Shinichi Satoh	National Institute of Informatics, Japan
Falk Scholer	RMIT University, Australia
Milad Shokouhi	Microsoft Research Cambridge, UK
Hideki Shima	Carnegie Mellon University, USA
Luo Si	Purdue University, USA
Mark Smucker	University of Waterloo, Canada

Ruihua Song	Microsoft Research Asia, China
Koichi Takeda	Tokyo Research Laboratory, IBM, Japan
Jamie Thom	RMIT University, Australia
Anastasios Tombros	Queen Mary, University of London, UK
Elaine Toms	Dalhousie University, Canada
Andrew Trotman	University of Otago, New Zealand
Anh Vo	University of Melbourne, Australia
Bin Wang	Institute of Computing Technology, Chinese Academy of Sciences, China
Richard Wang	Carnegie Mellon University, USA
William Webber	University of Melbourne, Australia
Iris Xie	University of Wisconsin-Milwaukee, USA
Guirong Xue	Shanghai Jiao-Tong University, China
Rong Yan	IBM T.J. Watson Research Center, USA
Qiang Yang	Hong Kong University of Science and Technology, Hong Kong
Hwan-Jo Yu	Pohang University of Science and Technology, Korea
Jun Zhao	Chinese Academy of Science, China
Le Zhao	Carnegie Mellon University, USA

Table of Contents

Regular Papers

Fully Automatic Text Categorization by Exploiting WordNet	1
<i>Jianqiang Li, Yu Zhao, and Bo Liu</i>	
A Latent Dirichlet Framework for Relevance Modeling	13
<i>Viet Ha-Thuc and Padmini Srinivasan</i>	
Assigning Location Information to Display Individuals on a Map for Web People Search Results	26
<i>Harumi Murakami, Yuya Takamori, Hiroshi Ueda, and Shoji Tatsumi</i>	
Web Spam Identification with User Browsing Graph	38
<i>Huijia Yu, Yiqun Liu, Min Zhang, Liyun Ru, and Shaoping Ma</i>	
Metric and Relevance Mismatch in Retrieval Evaluation	50
<i>Falk Scholer and Andrew Turpin</i>	
Test Collection-Based IR Evaluation Needs Extension toward Sessions – A Case of Extremely Short Queries	63
<i>Heikki Keskustalo, Kalervo Järvelin, Ari Pirkola, Tarun Sharma, and Marianne Lykke</i>	
Weighted Rank Correlation in Information Retrieval Evaluation	75
<i>Massimo Melucci</i>	
Extractive Summarization Based on Event Term Temporal Relation Graph and Critical Chain	87
<i>Maofu Liu, Wenjie Li, and Huijun Hu</i>	
Using an Information Quality Framework to Evaluate the Quality of Product Reviews	100
<i>You-De Tseng and Chien Chin Chen</i>	
Automatic Extraction for Product Feature Words from Comments on the Web	112
<i>Zhichao Li, Min Zhang, Shaoping Ma, Bo Zhou, and Yu Sun</i>	
Image Sense Classification in Text-Based Image Retrieval	124
<i>Yih-Chen Chang and Hsin-Hsi Chen</i>	
A Subword Normalized Cut Approach to Automatic Story Segmentation of Chinese Broadcast News	136
<i>Jin Zhang, Lei Xie, Wei Feng, and Yanning Zhang</i>	

Japanese Spontaneous Spoken Document Retrieval Using NMF-Based Topic Models	149
<i>Xinhui Hu, Hideki Kashioka, Ryosuke Isotani, and Satoshi Nakamura</i>	
Finding ‘Lucy in Disguise’: The Misheard Lyric Matching Problem	157
<i>Nicholas Ring and Alexandra L. Uitdenbogerd</i>	
Selecting Effective Terms for Query Formulation	168
<i>Chia-Jung Lee, Yi-Chun Lin, Ruey-Cheng Chen, and Pu-Jen Cheng</i>	
Discovering Volatile Events in Your Neighborhood: Local-Area Topic Extraction from Blog Entries	181
<i>Masayuki Okamoto and Masaaki Kikuchi</i>	
A Unified Graph-Based Iterative Reinforcement Approach to Personalized Search	193
<i>Yunping Huang, Le Sun, and Zhe Wang</i>	
Exploiting Sentence-Level Features for Near-Duplicate Document Detection	205
<i>Jenq-Haur Wang and Hung-Chi Chang</i>	
Posters	
Language Models of Collaborative Filtering	218
<i>Jun Wang</i>	
Efficient Text Classification Using Term Projection	230
<i>Yabin Zheng, Zhiyuan Liu, Shaohua Teng, and Maosong Sun</i>	
IPHITS: An Incremental Latent Topic Model for Link Structure	242
<i>Huifang Ma, Weizhong Zhao, Zhixin Li, and Zhongzhi Shi</i>	
Supervised Dual-PLSA for Personalized SMS Filtering	254
<i>Wei-ran Xu, Dong-xin Liu, Jun Guo, Yi-chao Cai, and Ri-le Hu</i>	
Enabling Effective User Interactions in Content-Based Image Retrieval	265
<i>Haiming Liu, Srdan Zagorac, Victoria Uren, Dawei Song, and Stefan Rüger</i>	
Improving Text Rankers by Term Locality Contexts	277
<i>Rey-Long Liu and Zong-Xing Lin</i>	
Mutual Screening Graph Algorithm: A New Bootstrapping Algorithm for Lexical Acquisition	289
<i>Yuhan Zhang and Yanquan Zhou</i>	

Web Image Retrieval for Abstract Queries Using Text and Image Information	300
<i>Kazutaka Shimada, Suguru Ishikawa, and Tsutomu Endo</i>	
Question Answering Based on Answer Trustworthiness	310
<i>Hyo-Jung Oh, Chung-Hee Lee, Yeo-Chan Yoon, and Myung-Gil Jang</i>	
Domain Specific Opinion Retrieval	318
<i>Guang Qiu, Feng Zhang, Jiajun Bu, and Chun Chen</i>	
A Boosting Approach for Learning to Rank Using SVD with Partially Labeled Data	330
<i>Yuan Lin, Hongfei Lin, Zhihao Yang, and Sui Su</i>	
Opinion Target Network and Bootstrapping Method for Chinese Opinion Target Extraction	339
<i>Yunqing Xia, Boyi Hao, and Kam-Fai Wong</i>	
Automatic Search Engine Performance Evaluation with the Wisdom of Crowds	351
<i>Rongwei Cen, Yiqun Liu, Min Zhang, Liyun Ru, and Shaoping Ma</i>	
A Clustering Framework Based on Adaptive Space Mapping and Rescaling	363
<i>Yiling Zeng, Hongbo Xu, Jiafeng Guo, Yu Wang, and Shuo Bai</i>	
Research on Lesk-C-Based WSD and Its Application in English-Chinese Bi-directional CLIR	375
<i>Yuejie Zhang and Tao Zhang</i>	
Searching Polyphonic Indonesian Folksongs Based on N-gram Indexing Technique	387
<i>Aurora Marsye and Mirna Adriani</i>	
Study on the Click Context of Web Search Users for Reliability Analysis	397
<i>Rongwei Cen, Yiqun Liu, Min Zhang, Liyun Ru, and Shaoping Ma</i>	
Utilizing Social Relationships for Blog Popularity Mining	409
<i>Chih-Lu Lin, Hao-Lun Tang, and Hung-Yu Kao</i>	
S-node: A Small-World Navigation System for Exploratory Search	420
<i>Satoshi Shimada, Tomohiro Fukuhara, and Tetsuji Satoh</i>	
Efficient Probabilistic Latent Semantic Analysis through Parallelization	432
<i>Raymond Wan, Vo Ngoc Anh, and Hiroshi Mamitsuka</i>	
Author Index	445