

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

New York University, NY, 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

Gary Geunbae Lee Akio Yamada
Helen Meng Sung Hyon Myaeng (Eds.)

Information Retrieval Technology

Second Asia Information Retrieval Symposium, AIRS 2005
Jeju Island, Korea, October 13-15, 2005
Proceedings



Springer

Volume Editors

Gary Geunbae Lee

Pohang University of Science and Technology (POSTECH)

Department of Computer Science and Engineering

San 31, Hyoja-Dong, Nam-gu, Pohang, 790-784, South Korea

E-mail: gblee@postech.ac.kr

Akio Yamada

Media and Information Research Laboratories, NEC Corp. Shimonumabe 1753

Nakahara-ku, Kawasaki 211-8666, Japan

E-mail: a-yamada@da.jp.nec.com

Helen Meng

The Chinese University of Hong Kong

Human-Computer Communications Laboratory

Department of Systems Engineering and Engineering Management

Shatin, NT, Hong Kong SAR, China

E-mail: hmmeng@se.cuhk.edu.hk

Sung Hyon Myaeng

Information and Communications University (ICU)

Munji-dong, Yuseong-gu, Daejeon, 305-714, South Korea

E-mail: myaeng@icu.ac.kr

Library of Congress Control Number: 2005933614

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

ISSN 0302-9743

ISBN-10 3-540-29186-5 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-29186-2 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 is a part of Springer Science+Business Media

springeronline.com

© Springer-Verlag Berlin Heidelberg 2005

Printed in Germany

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

Printed on acid-free paper SPIN: 11562382 06/3142 5 4 3 2 1 0

Preface

Asia Information Retrieval Symposium (AIRS) was established in 2004 by the Asian information retrieval community after the successful series of Information Retrieval with Asian Languages (IRAL) workshops held in six different locations in Asia, starting from 1996. The AIRS symposium aims to bring together international researchers and developers to exchange new ideas and the latest results in the field of information retrieval (IR). The scope of the symposium covers applications, systems, technologies and theoretical aspects of information retrieval in text, audio, image, video and multi-media data.

We are very pleased to report that we saw a sharp and steady increase in the number of submissions and their qualities, compared with previous IRAL workshop series. We received 136 submissions from all over the world including Asia, North America, Europe, Australia, and even Africa, from which 32 papers (23%) were presented in oral sessions and 36 papers in poster sessions (26%). We also held a special session called “Digital Photo Albuming,” where 4 oral papers and 3 posters were presented. It was a great challenge and hard work for the program committee to select the best among the excellent papers. The high acceptance rates witness the success and stability of the AIRS series. All the papers and posters are included in this LNCS (Lecture Notes in Computer Science) proceedings volume, which is SCI-indexed.

The technical program included two keynote talks by Prof. Walter Bender and Prof. Eduard Hovy respectively, and a total of 10 oral sessions with two in parallel including one special session and also two poster sessions.

The technical and social programs, which we are proud of, were made possible by the hard working people behind the scene. In addition to the Program Committee members, we are thankful to the Organizing Committee, Poster Chair (Helen Meng), Demo/Exhibition Chair (Myung-Gil Jang), Publicity Chairs (Hsin-his Chen, MunChurl Kim, Mun-kew Leong, Tetsuya Sakai, Bing Swen, Robert Luk), Sponsor Chair (Hee Suk Lim), Treasurer/Secretary (Bo Yeong Kang), and Qing Li for conference website management.

We also thank the sponsoring organizations Toshiba Corporation, Japan; LexisNexis, USA; INEK, Korea; and Daum Soft, Korea for their support, as well as the Information and Communication University and the Korea Institute of Science and Technology Information for hosting the conference, and Springer for publishing these proceeding in their LNCS series.

We believe that this conference has already set a very high standard for a regionally oriented conference, especially in Asia, and hope that you enjoyed the conference as well as the beautiful scenery of Jeju island.

August 2005

Gary G. Lee and Akio Yamada
Sung Hyon Myaeng and Dong In Park

Organization

General Conference Chairs

Sung Hyon Myaeng (ICU, Korea)

Dong In Park (KISTI, Korea)

Program Chairs

Gary G. Lee (POSTECH, Korea)

Akio Yamada (NEC, Japan)

Poster Chair

Helen Meng (Chinese University of Hong Kong, China)

Demo/Exhibition Chair

Myung-Gil Jang (ETRI, Korea)

Publicity Chairs

Hsin-hsi Chen (National Taiwan University, Taiwan)

MunChurl Kim (ICU, Korea)

Mun-kew Leong (I2R, Singapore)

Tetsuya Sakai (Toshiba, Japan)

Bing Swen (Peking University, China)

Robert Luk (Hong Kong Polytechnic University, China)

Sponsor Chair

Hee Suk Lim (Hansin University)

Treasurer/Secretary

Bo Yeong Kang (ICU, Korea)

Steering Committee

Jun Adachi, National Institute of Informatics, Japan

Hsin-Hsi Chen, National Taiwan University, Taiwan

Lee-Feng Chien, Academia Sinica, Taiwan

Gary Geunbae Lee, POSTECH, Korea

Mun-Kew Leong, Institute for Infocomm Research, Singapore

Helen Meng, The Chinese University of Hong Kong, China

Sung Hyon Myaeng, Information and Communication University, Korea

Hwee Tou Ng, National University of Singapore, Singapore

Tetsuya Sakai, Toshiba, Japan

Kam-Fai Wong, The Chinese University of Hong Kong, China

Ming Zhou, Microsoft Research Asia, China

Organization Committee

Bo Yeong Kang (Information & Communications University)

Munchurl Kim (Information & Communications University)

Myung-Gil Jang (ETRI)

Pyung Kim (KISTI)

Qing Li (Information & Communications University)

Hee Suk Lim (Hansin University)

Dong-Chul Lee (Cheju National University)

Youngjoong Ko (Dong-A University)

Hanmin Jung (KISTI)

Kyung Soon Lee (Chonbuk National University)

Hosted by

Information and Communications University

Korea Institute of Science and Technology Information

Program Committee

Hani Abu-Salem, DePaul University, USA

Terumasa Aoki, University of Tokyo, Japan

Kohtaro Asai, Mitsubishi Electric, Japan

Ricardo Baeza-Yates, ICREA-UPF, Spain & CWR, Univ. of Chile

Hsin-Hsi Chen, National Taiwan University, Taiwan

Lee-Feng Chien, Academia Sinica, Taiwan

Tat-Seng Chua, National University of Singapore, Singapore

Fabio Crestani, University of Strathclyde, Scotland, UK

Jianfeng Gao, Microsoft Research Asia, China

Tetsuya Ishikawa, University of Tsukuba, Japan

Min-Yen Kan, National University of Singapore, Singapore

Noriko Kando, National Institute of Informatics, Japan

Christopher Khoo, Nanyang Technological University, Singapore

Munchurl Kim, ICU, Korea

Hoi-Rin Kim, ICU, Korea

Juntae Kim, Dongguk University, Korea

Kazuaki Kishida, Surugadai University, Japan

Kui-Lam Kwok, Queens College, City University of New York, USA

Wai Lam, Chinese University of Hong Kong, China

Gary Geunbae Lee, POSTECH, Korea

JaeHo Lee, University of Seoul, Korea

Jae Sung Lee, Chungbuk National University, Korea
Jong-Hyeok Lee, POSTECH, Korea
Mun-Kew Leong, Institute for Infocomm Research, Singapore
Gina-Anne Levow, University of Chicago, USA
Hang Li, Microsoft Research Asia, China
Ee-Peng Lim, Nanyang Technological University, Singapore
Chin-Yew Lin, University of Southern California, USA
Robert Luk, Hong Kong Polytechnic University, China
Helen Meng, The Chinese University of Hong Kong, China
Hiroshi Murase, Nagoya University, Japan
Hiroshi Nakagawa, University of Tokyo, Japan
Hwee Tou Ng, National University of Singapore, Singapore
Jian-Yun Nie, University of Montreal, Canada
Sam Gyun Oh, Sungkyunkwan University, Korea
Jon Patrick, University of Sydney, Australia
Fuchun Peng, BBN Technologies, USA
Hae-Chang Rim, Korea University, Korea
YongMan Ro, ICU, Korea
Tetsuya Sakai, Toshiba Corporate R&D Center, Japan
Shin'ichi Satoh, National Institute of Informatics, Japan
John R. Smith, IBM T. J. Watson Research Center, USA
Padmini Srinivasan, University of Iowa, USA
Masaru Sugano, KDDI R&D Laboratories, Japan
Aixin Sun, University of New South Wales, Australia
Maosong Sun, Tsinghua University, China
Ulrich Thiel, Fraunhofer IPSI, Germany
Takenobu Tokunaga, Tokyo Institute of Technology, Japan
Hsin-Min Wang, Academia Sinica, Taiwan
Kam-Fai Wong, CUHK, China
Ross Wilkinson, CSIRO, Australia
Lide Wu, Fudan University, China
Seong-Joon Yoo, Sejong University, Korea
ChengXiang Zhai, University of Illinois, Urbana Champaign, USA
Min Zhang, Tsinghua University, China
Ming Zhou, Microsoft Research Asia, China
Justin Zobel, RMIT, Australia

Additional Reviewers

Yee Seng Chan, Tee Kiah Chia, Yunbo Cao, Jie Tang, Jun Xu

Table of Contents

Session 1A: Relevance/Retrieval Models

The Reliability of Metrics Based on Graded Relevance <i>Tetsuya Sakai</i>	1
Improving Weak Ad-Hoc Retrieval by Web Assistance and Data Fusion <i>Kui-Lam Kwok, Laszlo Grunfeld, Peter Deng</i>	17
Query Expansion with the Minimum Relevance Judgments <i>Masayuki Okabe, Kyoji Umemura, Seiji Yamada</i>	31
Improved Concurrency Control Technique with Lock-Free Querying for Multi-dimensional Index Structure <i>Myung-Keun Kim, Hae-Young Bae</i>	43

Session 1B: Multimedia IR

A Color-Based Image Retrieval Method Using Color Distribution and Common Bitmap <i>Chin-Chen Chang, Tzu-Chuen Lu</i>	56
A Probabilistic Model for Music Recommendation Considering Audio Features <i>Qing Li, Sung Hyon Myaeng, Dong Hai Guan, Byeong Man Kim</i> ...	72
VisMed: A Visual Vocabulary Approach for Medical Image Indexing and Retrieval <i>Joo-Hwee Lim, Jean-Pierre Chevallet</i>	84
Object Identification and Retrieval from Efficient Image Matching: Snap2Tell with the STOIC Dataset <i>Jean-Pierre Chevallet, Joo-Hwee Lim, Mun-Kew Leong</i>	97

Session 2A: Natural Language Processing in IR

Extracting the Significant Terms from a Sentence-Term Matrix by Removal of the Noise in Term Usage <i>Changbeom Lee, Hoseop Choe, Hyukro Park, Cheolyoung Ock</i>	113
---	-----

Cross Document Event Clustering Using Knowledge Mining from
Co-reference Chains
June-Jei Kuo, Hsin-Hsi Chen 121

Filtering Contents with Bigrams and Named Entities to Improve Text
Classification
François Paradis, Jian-Yun Nie 135

Session 2B: Enabling Technology

The Spatial Indexing Method for Supporting a Circular Location
Property of Object
Hwi-Joon Seon, Hong-Ki Kim 147

Supervised Categorization of JavaScript™ Using Program Analysis
Features
Wei Lu, Min-Yen Kan 160

Effective and Scalable Authorship Attribution Using Function Words
Ying Zhao, Justin Zobel 174

Session 3A: Web IR

Learning to Integrate Web Taxonomies with Fine-Grained Relations: A
Case Study Using Maximum Entropy Model
Chia-Wei Wu, Tzong-Han Tsai, Wen-Lian Hsu 190

WIDIT: Fusion-Based Approach to Web Search Optimization
Kiduk Yang, Ning Yu 206

Transactional Query Identification in Web Search
In-Ho Kang 221

Session 3B: Question Answering

Improving FAQ Retrieval Using Query Log Clustering in Latent
Semantic Space
Harksoo Kim, Hyunjung Lee, Jungyun Seo 233

Phrase-Based Definitional Question Answering Using Definition
Terminology
*Kyoung-Soo Han, Young-In Song, Sang-Bum Kim,
Hae-Chang Rim* 246

Enhanced Question Answering with Combination of Pre-acquired Answers	
<i>Hyo-Jung Oh, Chung-Hee Lee, Hyeon-Jin Kim, Myung-Gil Jang</i>	260

Session 4A: Document/Query Models

An Empirical Study of Query Expansion and Cluster-Based Retrieval in Language Modeling Approach	
<i>Seung-Hoon Na, In-Su Kang, Ji-Eun Roh, Jong-Hyeok Lee</i>	274
Effective Query Model Estimation Using Parsimonious Translation Model in Language Modeling Approach	
<i>Seung-Hoon Na, In-Su Kang, Ji-Eun Roh, Jong-Hyeok Lee</i>	288
Chinese Document Re-ranking Based on Term Distribution and Maximal Marginal Relevance	
<i>Lingpeng Yang, Donghong Ji, Munkew Leong</i>	299
On Effectiveness Measures and Relevance Functions in Ranking INEX Systems	
<i>Huyen-Trang Vu, Patrick Gallinari</i>	312

Session 4B: Special Session: Digital PhotoAlbum

Home Photo Categorization Based on Photographic Region Templates	
<i>Sang-Kyun Kim, Seungji Yang, Kyong Sok Seo, Yong Man Ro, Ji-Yeon Kim, Yang Suk Seo</i>	328
MPEG-7 Visual-Temporal Clustering for Digital Image Collections	
<i>Robert O'Callaghan, Mirosław Bober</i>	339
A Structured Learning Approach to Semantic Photo Indexing and Query	
<i>Joo-Hwee Lim, Jesse S. Jin, Suhuai Luo</i>	351
Image Retrieval Using Sub-image Matching in Photos Using MPEG-7 Descriptors	
<i>Min-Sung Ryu, Soo-Jun Park, Chee Sun Won</i>	366

Session 5A: TDT/Clustering

An Incremental Document Clustering for the Large Document Database	
<i>Kil Hong Joo, Won Suk Lee</i>	374

Subsite Retrieval: A Novel Concept for Topic Distillation <i>Tao Qin, Tie-Yan Liu, Xu-Dong Zhang, Guang Feng, Wei-Ying Ma</i>	388
A Rough Set-Based Fuzzy Clustering <i>Yaqin Zhao, Xianzhong Zhou, Guizhong Tang</i>	401
Effective Use of Place Information for Event Tracking <i>Yun Jin, Sung Hyon Myaeng, Mann-Ho Lee, Hyo-Jung Oh, Myung-Gil Jang</i>	410

Session 5B: Multimedia/Classification

A Classifier Design Based on Combining Multiple Components by Maximum Entropy Principle <i>Akinori Fujino, Naonori Ueda, Kazumi Saito</i>	423
A Query-by-Singing Technique for Retrieving Polyphonic Objects of Popular Music <i>Hung-Ming Yu, Wei-Ho Tsai, Hsin-Min Wang</i>	439
Integrating Textual and Visual Information for Cross-Language Image Retrieval <i>Wen-Cheng Lin, Yih-Chen Chang, Hsin-Hsi Chen</i>	454
Practical Application of Associative Classifier for Document Classification <i>Yongwook Yoon, Gary Geunbae Lee</i>	467

Poster and Demo Session 1

A Method for Query Expansion Using a Hierarchy of Clusters <i>Masaki Aono, Hironori Doi</i>	479
Chinese Question Classification from Approach and Semantic Views <i>Youzheng Wu, Jun Zhao, Bo Xu</i>	485
GJM-2: A Special Case of General Jelinek-Mercer Smoothing Method for Language Modeling Approach to Ad Hoc IR <i>Guodong Ding, Bin Wang</i>	491
The Empirical Impact of the Nature of Novelty Detection <i>Le Zhao, Min Zhang, Shaoping Ma</i>	497
Song Wave Retrieval Based on Frame-Wise Phoneme Recognition <i>Yuuichi Yaguchi, Ryuichi Oka</i>	503

<i>trackThem: Exploring a Large-Scale News Video Archive by Tracking Human Relations</i> <i>Ichiro Ide, Tomoyoshi Kinoshita, Hiroshi Mo, Norio Katayama, Shin'ichi Satoh</i>	510
Topic-Independent Web High-Quality Page Selection Based on K-Means Clustering <i>Canhui Wang, Yiqun Liu, Min Zhang, Shaoping Ma</i>	516
Improving Text Similarity Measurement by Critical Sentence Vector Model <i>Wei Li, Kam-Fai Wong, Chunfa Yuan, Wenjie Li, Yunqing Xia</i>	522
Audio Fingerprinting Scheme by Temporal Filtering for Audio Identification Immune to Channel-Distortion <i>Mansoo Park, Hoi-Rin Kim, Dong-Ho Shin, Seung Hyun Yang</i>	528
On the Chinese Document Clustering Based on Dynamical Term Clustering <i>Chih-Ming Tseng, Kun-Hsiu Tsai, Chiun-Chieh Hsu, His-Cheng Chang</i>	534
Integrating Heterogeneous Multimedia Resources <i>Seong Joon Yoo, Chull Hwan Song</i>	540
Calculating Webpage Importance with Site Structure Constraints <i>Hui-Min Yan, Tao Qin, Tie-Yan Liu, Xu-Dong Zhang, Guang Feng, Wei-Ying Ma</i>	546
Gene Ontology Classification of Biomedical Literatures Using Context Association <i>Ki Chan, Wai Lam</i>	552
An Examination of Feature Selection Frameworks in Text Categorization <i>Bong Chih How, Wong Ting Kiong</i>	558
Query Transitive Translation Using IR Score for Indonesian-Japanese CLIR <i>Ayu Purwarianti, Masatoshi Tsuchiya, Seiichi Nakagawa</i>	565
Development of a Meta Product Search Engine with Web Services <i>Wooju Kim, Daewoo Choi, Jongmyong Kim, Bonggyun Jin</i>	571
An Automatic Code Classification System by Using Memory-Based Learning and Information Retrieval Technique <i>Heui Seok Lim, Won Kyu Hoon Lee, Hyeon Chul Kim, Soon Young Jeong, Heon Chang Yu</i>	577

Finding New News: Novelty Detection in Broadcast News <i>Georgina Gaughan, Alan F. Smeaton</i>	583
Named Entity Tagging for Korean Using DL-CoTrain Algorithm <i>Byung-Kwan Kwak, Jeong-Won Cha</i>	589
Semantic Categorization of Contextual Features Based on Wordnet for G-to-P Conversion of Arabic Numerals Combined with Homographic Classifiers <i>Youngim Jung, Aesun Yoon, Hyuk-Chul Kwon</i>	595
 Poster and Demo Session 2	
Robust Matching Method for Scale and Rotation Invariant Local Descriptors and Its Application to Image Indexing <i>Kengo Terasawa, Takeshi Nagasaki, Toshio Kawashima</i>	601
Image Feedback Retrieval Based on Vector Space Model Transformation <i>Luo Xin, Shiro Ajioka, Masami Shishibori, Kenji Kita</i>	616
Indexing Structures for Content-Based Retrieval of Large Image Databases: A Review <i>Ling He, Lingda Wu, Yichao Cai, Yuchi Liu</i>	626
Document Similarity Search Based on Generic Summaries <i>Xiaojun Wan, Jianwu Yang</i>	635
KNM: A Novel Intelligent User Interface for Webpage Navigation <i>Shiguang Liu, Hua-Jun Zeng, Zheng Chen, Chuangbai Xiao, Wei-Ying Ma</i>	641
Towards Construction of Evaluation Framework for Query Expansion <i>Masaharu Yoshioka</i>	647
An Efficient Incremental Nearest Neighbor Algorithm for Processing <i>k</i> -Nearest Neighbor Queries with Visal and Semantic Predicates in Multimedia Information Retrieval System <i>Dong-Ho Lee, Dong-Joo Park</i>	653
Generating Term Transliterations Using Contextual Information and Validating Generated Results Using Web Corpora <i>Jin-Shea Kuo</i>	659

Handling Orthographic Varieties in Japanese IR: Fusion of Word-, N-Gram-, and Yomi-Based Indices Across Different Document Collections <i>Nina Kummer, Christa Womser-Hacker, Noriko Kando</i>	666
A Term Weighting Approach for Text Categorization <i>Kyung-Chan Lee, Seung-Shik Kang, Kwang-Soo Hahn</i>	673
A LF Based Answer Indexing Method for Encyclopedia Question-Answering System <i>Hyeon-Jin Kim, Ji-Hyun Wang, Chang-Ki Lee, Chung-Hee Lee, Myung-Gil Jang</i>	679
Approximate Phrase Match to Compile Synonymous Translation Terms for Korean Medical Indexing <i>Jae Sung Lee, Hye Mi Yun</i>	685
Protein Function Classification Based on Gene Ontology <i>Dae-Won Park, Hyoung-Sam Heo, Hyuk-Chul Kwon, Hea-Young Chung</i>	691
Extracting and Utilizing of IS-A Relation Patterns for Question Answering Systems <i>Bojun Shim, Youngjoong Ko, Jungyun Seo</i>	697
An Iterative Approach for Web Catalog Integration with Support Vector Machines <i>Ing-Xiang Chen, Jui-Chi Ho, Cheng-Zen Yang</i>	703
Fuzzy Post-clustering Algorithm for Web Search Engine <i>Younghee Im, Jiyoung Song, Daihee Park</i>	709
A Relational Nested Interval Encoding Scheme for XML Storage and Retrieval <i>Gap-Joo Na, Sang-Won Lee</i>	715
The Design of Webservices Framework Support Ontology Based Dynamic Service Composition <i>Seungkeun Lee, Sehoon Lee, Kiwook Lim, Junghyun Lee</i>	721
Privacy Preserving Decision Tree in Multi Party Environment <i>Eakalak Suthampan, Songrit Maneewongvatana</i>	727
Author Index	733