

Hwee Tou Ng Mun-Kew Leong
Min-Yen Kan Donghong Ji (Eds.)

Information Retrieval Technology

Third Asia Information Retrieval Symposium, AIRS 2006
Singapore, October 16-18, 2006
Proceedings

 Springer

Table of Contents

Session 1A: Text Retrieval

Query Expansion with ConceptNet and WordNet: An Intrinsic Comparison	1
<i>Ming-Hung Hsu, Ming-Feng Tsai, Hsin-Hsi Chen</i>	
Document Similarity Search Based on Manifold-Ranking of TextFiles	14
<i>Xiaojun Wan, Jianwu Yang, Jianguo Xiao</i>	
Adapting Document Ranking to Users' Preferences Using Click-Through Data	26
<i>Min Zhao, Hang Li, Adwait Ratnaparkhi, Hsiao-Wuen Hon, Jue Wang</i>	

Session 1B: Search and Extraction

A PDD-Based Searching Approach for Expert Finding in Intranet Information Management	43
<i>Yupeng Fu, Rongjing Xiang, Min Zhang, Yiqun Liu, Shaoping Ma</i>	
A Supervised Learning Approach to Entity Search	54
<i>Guoping Hu, Jingjing Liu, Hang Li, Yunbo Cao, Jian-Yun Nie, Jianfeng Gao</i>	
Hierarchical Learning Strategy in Relation Extraction Using Support Vector Machines	67
<i>GuoDong Zhou, Min Zhang, Guohong Fu</i>	

Session 1C: Text Classification and Indexing

Learning to Separate Text Content and Style for Classification	79
<i>Dell Zhang, Wee Sun Lee</i>	
Using Relative Entropy for Authorship Attribution	92
<i>Ying Zhao, Justin Zobel, Phil Vines</i>	
Efficient Query Evaluation Through Access-Reordering	106
<i>Steven Garcia, Andrew Turpin</i>	

Session 1D: Text Clustering

Natural Document Clustering by Clique Percolation in Random Graphs	119
<i>Wei Gao, Kam-Fai Wong</i>	
Text Clustering with Limited User Feedback Under Local Metric Learning	132
<i>Ruizhang Huang, Zhigang Zhang, Wai Lam</i>	
Toward Generic Title Generation for Clustered Documents	145
<i>Yuen-Hsien Tseng, Chi-Jen Lin, Hsiu-Han Chen, Yu-I Lin</i>	

Session 1E: Information Retrieval Models

Word Sense Language Model for Information Retrieval	158
<i>Liqi Gao, Yu Zhang, Ting Liu, Guiping Liu</i>	
Statistical Behavior Analysis of Smoothing Methods for Language Models of Mandarin Data Sets	172
<i>Ming-Shing Yu, Feng-Long Huang, Piyu Tsai</i>	
No Tag, a Little Nesting, and Great XML Keyword Search	187
<i>Lingbo Kong, Shiwei Tang, Dongqing Yang, Tengjiao Wang, Jun Gao</i>	

Session 2A: Web Information Retrieval

Improving Re-ranking of Search Results Using Collaborative Filtering . . .	205
<i>Rohini U, Vamshi Ambati</i>	
Learning to Integrate Web Catalogs with Conceptual Relationships in Hierarchical Thesaurus	217
<i>Jui-Chi Ho, Ing-Xiang Chen, Cheng-Zen Yang</i>	
Discovering Authoritative News Sources and Top News Stories	230
<i>Yang Hu, Mingjing Li, Zhiwei Li, Wei-ying Ma</i>	

Session 2B: Cross-Language Information Retrieval

Chinese Question-Answering: Comparing Monolingual with English-Chinese Cross-Lingual Results	244
<i>Kui-Lam Kwok, Peter Deng</i>	

Translation of Unknown Terms Via Web Mining for Information Retrieval	258
<i>Qing Li, Sung Hyon Myaeng, Yun Jin, Bo-Yeong Kang</i>	

A Cross-Lingual Framework for Web News Taxonomy Integration	270
<i>Cheng-Zen Yang, Che-Min Chen, Ing-Xiang Chen</i>	

Session 2C: Question Answering and Summarization

Learning Question Focus and Semantically Related Features from Web Search Results for Chinese Question Classification	284
<i>Shu-Jung Lin, Wen-Hsiang Lu</i>	

Improving the Robustness to Recognition Errors in Speech Input Question Answering	297
<i>Hideki Tsutsui, Toshihiko Manabe, Mika Fukui, Tetsuya Sakai, Hiroko Fujii, Koji Urata</i>	

An Adjacency Model for Sentence Ordering in Multi-document Summarization	313
<i>Yu Nie, Donghong Ji, Lingpeng Yang</i>	

Session 2D: Natural Language Processing

Poor Man's Stemming: Unsupervised Recognition of Same-Stem Words	323
<i>Harald Hammarström</i>	

NAYOSE: A System for Reference Disambiguation of Proper Nouns Appearing on Web Pages	338
<i>Shingo Ono, Minoru Yoshida, Hiroshi Nakagawa</i>	

Efficient and Robust Phrase Chunking Using Support Vector Machines	350
<i>Yu-Chieh Wu, Jie-Chi Yang, Yuc-Shi Lee, Show-Jane Yen</i>	

Session 2E: Evaluation

Statistical and Comparative Evaluation of Various Indexing and Search Models	362
<i>Samir Abdou, Jacques Savoy</i>	

Bootstrap-Based Comparisons of IR Metrics for Finding One Relevant Document	374
<i>Tetsuya Sakai</i>	

Evaluating Topic Difficulties from the Viewpoint of Query Term Expansion 390
Masaharu Yoshioka

Session 3A: Multimedia Information Retrieval

Incorporating Prior Knowledge into Multi-label Boosting for Cross-Modal Image Annotation and Retrieval 404
Wei Li, Maosong Sun

A Venation-Based Leaf Image Classification Scheme 416
Jin-Kyu Park, Ecn-Jun Hwang, Yunyoung Nam

Pic-A-Topic: Gathering Information Efficiently from Recorded TV Shows on Travel 429
Tetsuya Sakai, Tatsuya Uehara, Kazuo Sumita, Taishi Shimomori

A Music Retrieval System Based on Query-by-Singing for Karaoke Jukebox 445
Hung-Ming Yu, Wei-Ho Tsai, Hsin-Min Wang

Special Session: Medical Image Retrieval

A Semantic Fusion Approach Between Medical Images and Reports Using UMLS 460
Daniel Racoceanu, Caroline Lacoste, Roxana Teodorescu, Nicolas Vuilleminot

Automated Object Extraction for Medical Image Retrieval Using the Insight Toolkit (ITK)..... 476
Henning Müller, Joris Heuberger, Adrien Depeursinge, Antoine Geissbühler

Stripe: Image Feature Based on a New Grid Method and Its Application in ImageCLEF..... 489
Bo Qiu, Daniel Racoceanu, Chang Sheng Xu, Qi Tian

Poster Session

An Academic Information Retrieval System Based on Multiagent Framework 497
Toru Abe, Yuu Chiba, Suoya, Baoning Li, Tetsuo Kinoshita

Comparing Web Logs: Sensitivity Analysis and Two Types of Cross-Analysis	508
<i>Nikolai Buzikashvili</i>	
Concept Propagation Based on Visual Similarity	514
<i>Jean-Pierre Chevallet, Nicolas Maillot, Joo-Hwee Lim</i>	
Query Structuring with Two-Stage Term Dependence in the Japanese Language	522
<i>Koji Eguchi, W. Bruce Croft</i>	
Automatic Expansion of Abbreviations in Chinese News Text	530
<i>Guohong Fu, Kang-Kwong Luke, GuoDong Zhou, Ruifeng Xu</i>	
A Novel Ant-Based Clustering Approach for Document Clustering	537
<i>Yulan He, Siu Cheung Hui, Yongxiang Sim</i>	
Evaluating Scalability in Information Retrieval with Multigraded Relevance	545
<i>Amélie Imafouo, Michel Beigbeder</i>	
Text Mining for Medical Documents Using a Hidden Markov Model	553
<i>Hyeju Jang, Sa Kwang Song, Sung Hyon Myaeng</i>	
Multi-document Summarization Based on Unsupervised Clustering	560
<i>Paul Ji</i>	
A Content-Based 3D Graphic Information Retrieval System	567
<i>Yonghwan Kim, Soochan Hwang</i>	
Query Expansion for Contextual Question Using Genetic Algorithms	574
<i>Yasutomo Kimura, Kenji Araki</i>	
Fine-Grained Named Entity Recognition Using Conditional Random Fields for Question Answering	581
<i>Changki Lee, Yi-Gyu Hwang, Hyo-Jung Oh, Soojong Lim, Jeong Heo, Chung-Hee Lee, Hyeon-Jin Kim, Ji-Hyun Wang, Myung-Gil Jang</i>	
A Hybrid Model for Sentence Ordering in Extractive Multi-document Summarization	588
<i>Deri Liu, Zengchang Zhang, Yanxiang He, Donghong Ji</i>	
Automatic Query Type Identification Based on Click Through Information	593
<i>Yiqun Liu, Min Zhang, Liyun Ru, Shaoping Ma</i>	

Japanese Question-Answering System for Contextual Questions Using Simple Connection Method, Decreased Adding with Multiple Answers, and Selection by Ratio	601
<i>Masaki Murata, Masao Utiyama, Hitoshi Isahara</i>	
Multi-document Summarization Using a Clustering-Based Hybrid Strategy	608
<i>Yu Nie, Donghong Ji, Lingpeng Yang, Zhengyu Niu, Tingting He</i>	
A Web User Preference Perception System Based on Fuzzy Data Mining Method	615
<i>Wei-Shen Tai, Chen-Tung Chen</i>	
An Analysis on Topic Features and Difficulties Based on Web Navigational Retrieval Experiments	625
<i>Masao Takaku, Keizo Oyama, Akiko Aizawa</i>	
Towards Automatic Domain Classification of Technical Terms: Estimating Domain Specificity of a Term Using the Web	633
<i>Takehito Utsuro, Mitsuhiro Kida, Masatsugu Tonoike, Satoshi Sato</i>	
Evaluating Score Normalization Methods in Data Fusion	642
<i>Shengli Wu, Fabio Crestani, Yixin Bi</i>	
WIDIT: Integrated Approach to HARD Topic Search	649
<i>Kiduk Yang, Ning Yu, Hui Zhang, Shahrier Akram, Ivan Record</i>	
Automatic Query Expansion Using Data Manifold	659
<i>Lingpeng Yang, Donghong Ji, Yu Nie, Tingting He</i>	
An Empirical Comparison of Translation Disambiguation Techniques for Chinese-English Cross-Language Information Retrieval	666
<i>Ying Zhang, Phil Vines, Justin Zobel</i>	
Web Mining for Lexical Context-Specific Paraphrasing	673
<i>Shiqi Zhao, Ting Liu, Xincheng Yuan, Sheng Li, Yu Zhang</i>	
Author Index	681