Available online at www.sciencedirect.com
ScienceDirect
Information Processing and Management 43 (2007) 297–301
INFORMATION PROCESSING & MANAGEMENT
www.elsevier.com/locate/infoproman

Editorial

Special issue on AIRS2005: Information retrieval research in Asia

1. Introduction

Asia Information Retrieval Symposium (AIRS) was first established in 2004 by the Asian information retrieval community after the success of the International Workshop on Information Retrieval with Asian Languages (IRAL) series, held in Taejon (1996), Tsukuba (1997), Singapore (1998), Taipei (1999), Hong Kong (2000) and Sapporo (2003). AIRS aims to bring together international researchers and developers to exchange new ideas and latest achievements in the field of information retrieval (IR), especially in Asia. The scope of the symposium covers applications, systems, technologies and theoretical aspects of information retrieval in text, audio, image, video and multi-media data.

The information retrieval research activities in Asia mark a sharp and steady increase not only in the number of submissions to the series of the workshops and symposiums but also in qualities. For example, AIRS2005 (Gary Geunbae Lee, Akio Yamada, Helen Meng, & Sung Hyon Myaeng, 2005) received 136 submissions from all over the world including Asia, North America, Europe, Australia, and Africa, from which only 32 papers (23%) were presented in oral sessions and 36 papers in poster sessions (26%). The low acceptance rates witness the success and stability of the future information retrieval research activities in Asia.

This special issue contains the best papers carefully selected from AIRS2005, and the selected papers were revised and extended based on the reviews of the world class researchers. While the papers have a broad coverage in IR without a specific concentration often associated with a special issue, the fact that they are from a group of researchers in a specific region makes this special issue interesting. They not only represent most significant IR research in recent years but also reflect the R&D interests in Asia. We hope that this special issue helps the international IR community, including the researchers and developers in Asia, increase awareness of IR research activities in Asia and stimulates collaborations across different regions.

2. Papers in this issue

Technical areas covered in this special issue of "Information Retrieval Research in Asia" are very diverse including IR theories and models, text mining, web related techniques, multimedia retrieval and experimentation/evaluation issues. The papers are grouped with their corresponding topics as follows.

2.1. IR theories and models: language modeling and document ranking

Seung-Hoon Na, In-Su Kang, Ji-Eun Roh, Jong-Hyeok Lee: An Empirical Study of Query Expansion and Cluster-Based Retrieval in Language Modeling Approach. The term mismatch problem in information retrieval is very critical and has been major hurdles for good IR performance. This paper first performs an empirical study on query expansion using parsimony and cluster-based retrieval in terms of retrieval performance. They

0306-4573/$ - see front matter
doi:10.1016/j.ipm.2006.08.002

found that query expansion is generally more effective than cluster-based retrieval in resolving the word-mismatch problem, and their combinations are more effective when each method significantly improves the baseline performance.

Lingpeng Yang, Donghong Ji, Munkew Leong: Document reranking by term distribution and maximal marginal relevance for Chinese Information Retrieval. This paper proposes a document reranking method based on a specific term weighting scheme, which integrates both local and global distribution of terms as well as document frequency, document positions and term length. The new weighting scheme was used to replace the *idf* scheme in MMR (maximal marginal relevance), and was shown to outperform relevant methods consistently by evaluation on the NTCIR-3 Chinese document collection.

2.2. Text mining issues: classification, clustering, question answering, topic detection and tracking

June-Jei Kuo and Hsin-Hsi Chen: Cross Document Event Clustering Using Knowledge Mining from Co-Reference Chains. Topic detection and tracking must handle the co-reference relations between the entities across several documents. This paper proposes a metric of normalized-chain edit-distance to mine the controlled vocabularies from cross-document co-reference chains, and suggests a novel threshold model to use the controlled vocabularies for event clustering.

Francois Paradis, Jian-Yun Nie: Contextual Feature Selection for Text Classification. This paper presents a simple approach for the classification of ''noisy'' documents using bigrams and named entities. The approach combines conventional feature selection with a contextual approach to filter out passages around selected features.

Kyoung-Soo Han, Young-In Song, Sang-Bum Kim, and Hae-Chang Rim: *Answer Extraction and Ranking Strategies for Definitional Question Answering Using Linguistic Features and Definition Terminology*. Definitional questions refer to the defining questions such as ''What are fractals?'' and have several different characteristics from the factoid questions. This paper proposes answer extraction and ranking strategies for definitional question answering using linguistic features and definition terminology. This paper uses several diverse techniques to overcome the limitations of the error-prone NLP techniques such as simple anaphora resolution, syntactic patterns, and definitional terminology to generate more effective answers.

Yun Jin, Sung Hyon Myaeng, Yuchul Jung: Use of Place Information for Improved Event Tracking. Even though topic detection and tracking (TDT) can benefit from an explicit use of time and place information, the place information has been rarely used unlike the time information. This paper focused on the place information for more effective TDT tasks. News articles were analyzed for their characteristics of place information, and a new topic tracking method was proposed to incorporate the analysis results.

Akinori Fujino, Naonori Ueda, Kazumi Saito: A hybrid generative/discriminative approach to text classification with additional information. Web pages and technical papers consist of many additional components such as titles, links, and authors besides main text, and this paper presents a hybrid text classifier based on probabilistic generative and discriminative approaches. The formulation considers individual component generative models and constructs a classifier by combining these individual trained models based on the maximum entropy principle.

Yongwook Yoon, Gary Geunbae Lee: Efficient Implementation of Associative Classifiers for Document Classification. Associative classifiers have many favorable characteristics such as rapid training, good classification accuracy, and excellent interpretation ability, but have been rarely used for text classification tasks. This paper applies the associative classification techniques to document classification problem, and proposes a feature selection method based on dimensionality reduction and a new efficient method for storing and pruning the rules to handle the huge amount of classification rules.

2.3. Web related issues: web mining, web search engine and FAQ retrieval

Kui-Lam Kwok, Laszlo Grunfeld, Peter Deng: Employing Web Mining and Data Fusion to Improve Weak Ad Hoc Retrieval. Improving the weak query and retrieval effectiveness is an important issue for ad-hoc retrieval and this paper proposes a data fusion of sufficiently *dissimilar* retrieval lists to improve the weak query results.