# Preface

The research in data mining and information retrieval on structured and unstructured data has been growing tremendously in recent years. Unlike structured data commonly found in database or data warehouse, unstructured data does not have a rigid schema. Various kinds of text data such as online news, emails, memos, and equipment maintenance logs are examples of unstructured data. Other examples are data composed of a combination of structured database records and free texts such as data found in biomedical domain, automated diagnosis, and customer relationship management applications. Mining useful information or extracting knowledge from structured and unstructured data has received much attention recently. Information retrieval from unstructured data especially text data has also drawn much interests. A variety of models and techniques have been developed. Computational technique plays an important role in developing intelligent mining and information retrieval methods for handling and processing the information content found in the data.

This special issue features four high-quality papers addressing significant research related to the broad area of computational informatics in data mining and information retrieval.

"A Data Warehousing and Data Mining Framework for Web Usage Management" by Wu, Ng, and Huang, presents an integrated data warehousing and data mining framework for website management. This paper also proposes a pattern discovery method for analyzing Web user behavior. It tackles the problem of dynamic user pattern discovery from a large amount of clickstreams. It employs mathematical and statistical models to monitor the changing patterns. An efficient data model for aggregating user access sessions is developed. Experiments on a sports website have been conducted to demonstrate the effectiveness of the model. The results indicate that the proposed integrated model is useful and efficient.

"A New Data Mining Approach to Predicting Matrix Condition Numbers", by Xu and Zhang, proposes a novel approach for the estimation of the condition number of general sparse matrices. Instead of direct calculation using numerical methods, they employ data mining techniques for estimating the condition number. In particular, they make use of support vector regression (SVR) in their approach. Some feature selection methods have also been investigated for further reduction of the time cost and precision accuracy. They develop a feature selection criterion combining the weights from SVR with the weights from comparison of matrices with their preconditioned counterparts. The experimental results demonstrate that the response time of the prediction technique is fifteen times faster than the direct computation method for many large matrices. Therefore, the proposed solution is suitable for online applications which require the estimation of the condition number of general sparse matrices.

"Estimating Timestamp From Incomplete News Corpus", by Uejima, Miura, and Shioya, investigates the issue of estimating the timestamp information for a news article. Sometimes, the timestamps of news stories are not given explicitly. If the timestamp information can be inferred automatically, one can build intelligent applications such as topic and trend analysis as well as temporal-based summarization. The authors propose a method for learning temporal information and topic information by means of both Expectation-Maximization (EM) algorithm and incremental clustering. The effectiveness of the proposed method is validated by experiments on a real-world