

Hong Kong University of Science and Technology
School of Business and Management
Department of Information and Systems Management

Seminar Announcement

Mining the Web for Personalized and Specialized Information

by

***Mr Michael Chiu-Lung Chau
Department of Management Information Systems
Eller College of Business and Public Administration
The University of Arizona***

**28 January 2003 (Tuesday)
11:00am – 12:00noon
ISMT Conference Room (Room 4379, lifts 17-18)**

~~~~~ All interested are welcome ~~~~~

**Abstract**

The size of the Web is growing exponentially and the number of indexable pages on the Web has exceeded 2 billion. Search engines find it increasingly difficult to keep an up-to-date and comprehensive search index, resulting in low precision and low recall rates. Consequently, many useful information and knowledge are hidden in the Web and it is difficult for users to find them. Another problem with existing search engines is that most of them only return a list of search results; personalized and in-depth analyses of the search results are usually not available. Searching and mining knowledge from the Web, therefore, becomes an interesting research topic. Two possible solutions to these problems are personalized search agents (a.k.a. spiders or crawlers) and specialized search engines. Both approaches provide users with customized search and analysis capabilities to extract knowledge in specific domains from the Web. This project focuses on the development of specialized search engines. A feature-based machine learning approach that incorporates Web content and Web structure analysis was proposed to perform Web page filtering in creating such engines. In the proposed approach, pre-defined features were extracted from each Web page and used as the input to a neural network and a support vector machine text classifier. The proposed approach was compared with traditional Web page filtering approaches and the results demonstrated that the proposed approach achieved statistically better performance in Web page filtering, especially when the number of training documents is small. We believe the techniques can be used in various applications such as knowledge management and e-commerce systems.

**Biography**

Mr Michael Chau is a doctoral candidate under the supervision of Professor Hsinchun Chen in the Department of Management Information Systems at the University of Arizona, where he is also a research associate in the Artificial Intelligence Lab and the COPLINK Center. He received a Bachelor of Science degree with first class honor in computer science and information systems from the University of Hong Kong in 1998 and joined the doctoral program in Arizona in the same year. Michael's research interests include Web applications, knowledge management, information retrieval, electronic commerce, intelligent agents, and financial information systems. He has participated actively in several projects funded by the National Science Foundation (NSF) and the National Institute of Health (NIH). The areas covered by these projects include digital library, knowledge management, Web agents, data and text mining, and medical informatics. His research has been (or will be) published in journals such as *IEEE Computer*, *Decision Support Systems (DSS)*, and *Journal of the American Society of Information Science and Technology (JASIST)*. He has served as the reviewer for various journals and conferences, including *Information Systems Research (ISR)* and *ACM Transactions on Information Systems (TOIS)*. He is also the co-founder of Zenic Technologies, Inc., a Tucson-based startup company that specializes in document management and knowledge management consulting.