

Searching and Mining the Web for Personalized and Specialized Information

Michael Chau

School of Business
The University of Hong Kong
Pokfulam, Hong Kong
mchau@business.hku.hk

With the rapid growth of the Web, users are often faced with the problem of information overload and find it difficult to search for relevant and useful information on the Web. Besides general-purpose search engines, there exist some alternative approaches that can help users perform searches on the Web more effectively and efficiently. Personalized search agents and specialized search engines are two such approaches. The goal of this research is to investigate how machine learning and artificial intelligence techniques can be used to improve these approaches.

A system development research process was adopted as the methodology in this study. In the first part of this research, five different personalized search agents, namely CI Spider, Meta Spider, Cancer Spider, Nano Spider, and Collaborative Spider, were developed. These spiders combine Web searching with various techniques such as noun phrasing, text clustering, and multi-agent technologies to help satisfy users' information needs in different domains and different contexts. Individual experiments were designed and conducted to evaluate the proposed approach and the experimental results showed that the prototype systems performed better than or comparable to traditional search methods.

The second part of the research aims to investigate how artificial intelligence techniques can be used to facilitate the development of specialized search engines. A Hopfield Net spider was proposed to locate from the Web URLs that are relevant to a given domain. A feature-based machine-learning text classifier also was proposed to perform filtering on Web pages. A prototype system was built for each approach. Both systems were evaluated and the results demonstrated that they both outperformed traditional approaches.

This research has two main contributions. Firstly, it demonstrated how machine learning and artificial intelligence techniques can be used to improve and enhance the development of personalized search agents and specialized search engines. Secondly, it provided a set of tools that can facilitate users in their Web searching and Web mining activities in various contexts.

Relevant publications are available on the Web at <http://www.business.hku.hk/~mchau/>