

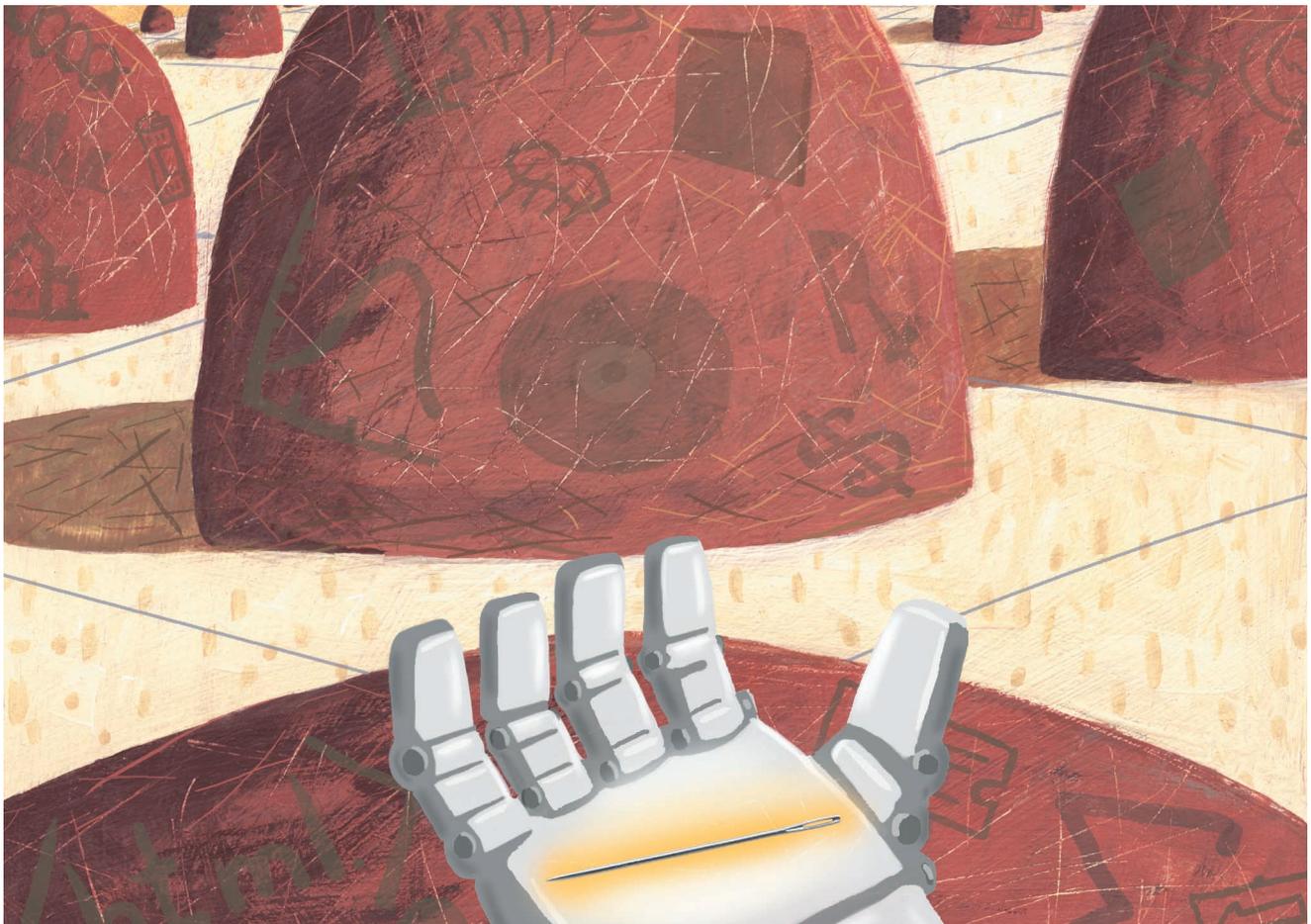
Mining Actionable Knowledge on the Web

Qiang Yang, *Hong Kong University of Science and Technology*

Craig A. Knoblock, *University of Southern California*

Xindong Wu, *University of Vermont*

The Web—its resources and users—offers a wealth of information for data mining and knowledge discovery. Up to now, a great deal of work has been done applying data mining and machine learning methods to discover novel and useful knowledge on the Web. However, many techniques aim only at extracting knowledge for human



users to view and use. Recently, more and more work addresses mining the Web for knowledge that computer systems will use. You can apply such actionable knowledge back to the Web for measurable performance improvements. This special issue of *IEEE Intelligent Systems* features five articles that address the problem of actionable Web mining.

“VISCORS: A Visual-Content Recommender for the Mobile Web,” by Chan Young Kim, Jae Kyu Lee, Yoon Ho Cho, and Deok Hwan Kim, presents an algorithm for applying collaborative filtering to deliver Web content to mobile users. Their algorithm combines two information-filtering techniques: collaborative filtering and content-based image retrieval. The resulting system recommends multimedia Web files to mobile users by learning similar users’ interests and Web pages with similar contents and by adaptively acquiring mobile-users’ preferences.

“Collaborative Filtering with Maximum Entropy,” by Dmitry Pavlov, Eren Manavoglu, David M. Pennock, and C. Lee Giles, presents a novel maximum-entropy approach for generating online recommendations as a user navigates through a collection of documents. Unlike previous approaches in which each document is viewed as a set of unordered keywords, in this article, the authors’ approach represents document access records as a collection of ordered sequences of document requests. A statistical model is then learned to allow collaborative filtering that makes recommendations based on users’ interests. It presents a clustering algorithm to help scale up the recommender algorithm and demonstrates that it compares well with other similarity-based recommender systems.

“Mining Web Pages for Data Records,” by Bing Liu, Robert Grossman, and Yanhong Zhai, presents a method for extracting structured data records from unstructured data sources such as HTML files. These data records often present their host pages’ essential information in the form of product lists and service catalogs. Unlike previous supervised-learning methods, which require substantial human effort, the authors propose a more effective automatic technique based on string matching. Additionally, their technique can mine both contiguous and noncontiguous data records. Existing techniques can’t mine noncontiguous data records. Their technique can

help to automatically compose Web services that communicate with each other to complete sophisticated computational tasks.

With a similar goal in mind, “OLERA: Semisupervised Web-Data Extraction with Visual Support,” by Chia-Hui Chang and Shih-Chien Kuo, presents an algorithm for extracting semistructured content from Web documents. The authors focus on reducing human involvement as much as possible while retaining much of the information extraction’s accuracy. To achieve this, they apply a technique that combines rule learning and string matching, and measured it against data sets of various complexities.

“Exploiting Interclass Rules for Focused Crawling” by Ismail Sengör Altıngövdde and Özgür Ulusoy presents an algorithm for a focused crawler that finds Web pages on a particular target topic. A crucial issue for a focused crawler is the underlying heuristic used for deciding the page to visit next. The authors propose a rule-based approach to improve a baseline focused crawler’s harvest rate and coverage. They explore rules derived from interclass (topic) linkage patterns for deciding the crawler’s next move. The rule-based approach also enhances the baseline crawler in supporting tunneling. You can use the focused-crawler algorithm to find topic-specific Web pages and Web sites that other data mining algorithms can later mine or to help focus a search engine’s results.

Web mining is a research frontier for both data mining and Web information exploration and has seen sustained interest in recent years. Actionable Web mining research is only just beginning to bear fruit—with many productive years ahead. We hope that you’ll enjoy this issue and join us in our efforts in data mining and Web information exploration to answer questions such as

- How can we use knowledge about user behaviors and usage patterns to build adaptive user interfaces and high-performance network systems?
- How can we use Web site structure, content, and linkage information to improve Web navigation and search?
- How can we wrap Web pages to build more convenient Web services? ■

The Authors



Qiang Yang is a faculty member of the Computer Science Department at the Hong Kong University of Science and Technology. His research interests include artificial intelligence planning, machine learning and case-based reasoning, data mining, and wireless computing. He received his PhD in computer science from the University of Maryland, College Park. He is a member of the AAAI and a senior member of the IEEE. Contact him at the Dept. of Computer Science, Hong Kong Univ. of Science and Technology, Clearwater Bay, Kowloon, Hong Kong; qyang@cs.ust.hk; www.cs.ust.hk/~qyang.



Craig A. Knoblock is a senior project leader at the Information Sciences Institute and a research associate professor in computer science at the University of Southern California. His research interests include information integration, automated planning, machine learning, and constraint reasoning and the application of these techniques to geospatial and biological data integration. He received his PhD in computer science from Carnegie Mellon University. He is a member of the AAAI and ACM. Contact him at the Univ. of Southern California, 4676 Admiralty Way, Ste. 1001, Marina del Rey, CA 90292; knoblock@isi.edu, www.isi.edu/~knoblock.



Xindong Wu is a professor and the chair of the Department of Computer Science at the University of Vermont. His research interests include data mining, knowledge-based systems, and Web information exploration. He received his PhD in artificial intelligence from the University of Edinburgh. He is the executive editor of *Knowledge and Information Systems*, founder and current steering-committee chair of the IEEE International Conference on Data Mining, a series editor of the Springer Book Series on Advanced Information and Knowledge Processing, and the chair of the IEEE Computer Society Technical Committee on Computational Intelligence. He will be the editor in chief of *IEEE Transactions on Knowledge and Data Engineering* starting in January 2005. He is the winner of the 2004 ACM SIGKDD Service Award. Contact him at the Univ. of Vermont, Dept. of Computer Science, 33 Colchester Ave., Burlington, VT 05405; xwu@emba.uvm.edu.