Book Reviews

Edited by Michèle Hudon

Book Review Editor

Knowledge Discovery in Bibliographic Databases. Edited by Jian Qin and M. Jay Norton. *Library Trends* 48, no 1. Champaign, IL : University of Illinois Graduate School of Library and Information Science, 1999. 281 p.

Knowledge discovery in databases (KDD) uses various means of "investigating data to extract knowledge" (p. 1), and more precisely, knowledge that is useful. As noted by the editors of this issue of *Library Trends*, research in library and information science is implicitly related to KDD by virtue of methods and techniques, although the characteristic language of discourse typically does not use this explicit term to describe what it does. This collection of thirteen articles addresses "aspects of KDD that are relevant or reflective of the field of library and information science" (p. 1), offering evidence to support the notion that no small part of what we do (and in fact have long done) can be defined in terms of knowledge discovery in bibliographic databases.

Appropriately, the introductory and concluding chapters address broad issues affecting KDD. In the former case, M. Jay Norton's review of the implications for KDD of database quality, along with underlying factors of database design and data collection and representation, focuses on the more technical issues. In contrast, the latter has a more psycho-social thrust in Herbert S. White's argument that in providing end users with direct access to increasingly more information that is increasingly more complex, information specialists have abdicated both responsibilities and privileges unique to their role as knowledge workers. The intervening chapters address various aspects of the structure and substance of the "data" that facilitate knowledge discovery, tools and techniques, and specific applications in KDD.

The implications for KDD of different characteristics of the nature and quality of data and their representations are discussed in three chapters. Maria Pinto and F.W. Lancaster contrast the requirements for abstracts read by humans in making relevance decisions and those searched by computers in document retrieval. LixinYu describes GeoMatch, an interactive system for retrieving spatial cartographic information through end-user specification of coordinate data and their relationships. Barbara Kwasnik examines the nature of the relationship between classification and knowledge, and explores the potential of different systems of classification to provide a conceptual infrastructure that can either permit or prevent discovery of the knowledge inherent in them.

Four of the chapters present tools and techniques used in or potentially useful for KDD. Qin He reviews the development of co-word analysis as a technique for KDD, weighs its advantages and disadvantages, and discusses some research issues. Helena Ahonen presents the methodology for computing maximal frequent word sentences and their potential use in information retrieval. Gobinda G. Chowdhury advocates the use of standardized templates to facilitate data mining. Bipin C. Desai, Raijan Shingal, Nader R. Shayan, and Youquan Zhou describe CINDI, a system for indexing and retrieval based on semantic and syntactic metadata in Web documents.

Among the specific applications of knowledge discovery in bibliographic databases presented, Don R. Swanson and Neil R. Smalheiser's Arrowsmith is a system using expert knowledge to bring to light implicit complementary relationships among medical texts. Kenneth A. Cory reports on his efforts to exploit the application of Swanson's methodology in discovering previously unrecognized analogies in humanities databases. Two chapters present the incorporation of KDD tools and applications into the methodology for a research paradigm increasingly more important in today's context of large distributed data resources but traditionally the domain of library and information science, namely bibliometric analyses of text and document linkages. Henry Small demon-

strates that strong co-citation links can be used to construct cross-disciplinary pathways through the scientific literature, which in turn can support interdisciplinary knowledge discovery. Jian Qin presents a methodology for semantic pattern analysis of keywords from bibliographically coupled documents and discusses the potential to incorporate such techniques for knowledge discovery into information retrieval tools.

Somewhat ironically, the overall value of this collection is weakened by its exclusive focus on the contributions of library and information science to current progress in KDD. Limiting the scope in this way diminishes the importance of multi-disciplinary collaborations for the future success of automated approaches to true knowledge discovery, and may have led to the inclusion of papers somewhat peripheral to the central topic. Further, the individual contributions to this collection would have been better served by a more explicit sense of their relationships both to the broader topic in the context of other domains and in association to one another. Even so, these papers provide thought-provoking insights from a traditional perspective on an emerging field. The volume would be a useful addition to the professional library and an excellent resource for supplemental course readings in library and information science.

Carol A. Bean

Dr. Carol A. Bean, National Library of Medicine, Division of Extramural Programs, 6705 Rockledge Drive, Bethesda, MD, 20892 USA, e-mail: beanc@mail.nlm.nih.gov

The Web of Knowledge : Festschrift in honor of Eugene Garfield. Edited by Blaise Cronin & Helen Barsky Atkins. Medford, NJ: Information Today, Inc., 2000. 564 p. ISBN 1-57387-099-4.

Eugene Garfield's 75th birthday is the occasion for this book honouring his contribution to the field of information science. This work is a compilation of papers written by outstanding researchers in information science and other specialties. All share the belief that the citation indexes that were produced by the Institute for Scientific Information (ISI) as a result of Garfield's ideas have brought about a paradigm shift similar to that in the sociology of science.

The content of the book is not merely a compilation of papers; on the contrary, its goal is to provide a comprehensive description of the main research themes for which the Science Citation Index (SCI) and related products play a remarkable role as information resources. The book contains five sections devoted to the most relevant research topics based on SCI. Considering the difficulties of maintaining a unified structure in a work written by thirty seven different authors, we can say that the result is an interesting hybrid. It is a cross between an introductory work to many of the important research issues in the areas of bibliometrics, the science of sciences, and related topics; and a work of the type "perspectives on..." these fields that could be used as a manual for university students. It is also an indispensable source for learning the history and the origins of ISI.

It is important to point out, even though we do find in here highly specialized topics, that the book is suitable for an audience wider than that of the equally highly specialized researcher. Its narrative, essay style and limited mathematical content make it easy to read. *The Web of Knowledge* is an adequate guide to the topics it covers for any professional engaged in scientific information or any researcher who has worked with citation indexes or wants to know more about their potential uses. The editors obviously had a wide and varied audience in mind when they outlined the contents, a potential audience made of the millions of scientists and information professionals who use citation indexes nowadays.

The work is divided into five broad sections: 1. Historical perspectives. 2. The scientific literature. 3. International issues. 4. Evaluative bibliometrics. 5. Social networks analysis.

Section one is especially significant. Various researchers, including some, such as Joshua Lederberg, who were close to Eugene Garfield and his ideas, reveal the man to us. They show us Garfield's enterprising, persevering, and enthusiastic temperament and help us to understand his ultimate success in creating a scientific citation index, despite important difficulties and high costs. In a sense, it could be said that Garfield is like the robot picture of a scientist whose qualities allow him to change the history of a discipline. The essays included in the first chapter show us how Garfield overcame adversities in the early years of his work on the index and went on to transmit his strong faith in his simple and elegant "idea" first to researchers and finally to partners willing to finance its realization. These papers also show the innovative aspects of SCI, such as its multi-disciplinary character and the possibility of tracking scientific ideas through