of publication. It may be that journals in the sciences are published more frequently than those in the social sciences. He does not examine the numbers of submissions received by the editors. These factors may influence the sizes of editorial staffs. If a science journal is published weekly and receives a thousand submissions a year, that journal will require a larger staff and budget than a social science journal published quarterly with one hundred submissions per year.

These shortcomings, while serious, are not the most important. McGinty includes many quotations from his interviews with the editors. At times the editors' words are illuminating and provide insight into the workings of the journals. At other times, the statements are repetitive and intrusive. Many opinions and experiences could have been summarized so that the reader could better understand the challenges faced by the editors. Without the numerous quotations this book would have been even shorter—this is the most troubling aspect of all. McGinty's examination of journal editing could have been an interesting and informative article, but it has been unnecessarily expanded to justify publication as a high-priced book.—*John M. Budd (buddj@missouri.edu), School of Information Science and Learning Technologies, University of Missouri-Columbia*

Knowledge Discovery in Bibliographic Databases. Ed. Jian Qin and M. Jay Norton. Library Trends 48, no. 1 (Summer 1999). Champaign: University of Illinois at Urbana–Champaign, Graduate School of Library and Information Science, 1999. 281p. single copy, $18.50 (ISSN 0024-2594).

Knowledge discovery in databases (KDD) is one of those arcane information science topics that seem both mysterious and inviting to most librarians, bearing an aura of the future of librarianship. While being discussed in the major information science journals (e.g., Trybula 1997; Vickery 1997; Raghavan et al. 1998), it has not found its way into mainstream library science literature. If for no other reason then, the appearance of this issue of *Library Trends* is a welcome development, especially because of its focus on using KDD in bibliographic databases.

The papers comprising this book have been artfully assembled. The introduction and a useful overview of KDD are followed by an assessment of classification schemes, from the standpoint of knowledge discovery, as devices of knowledge representation. This link to bibliographic organizational practice yields in turn to two accounts of finding new knowledge by discovering connections, through common citations, between sets of articles in the biomedical and philosophical literatures. Next is a demonstration of using cocitation links to forge a pathway of relationships through the literatures of several subject areas from economics to astrophysics. There follow three articles on different aspects of discovering knowledge in word-occurrence patterns, another four on automated knowledge discovery using various kinds of document surrogates (search-engine templates, metadata headers, abstracts, MARC-encoded geospatial data), and a concluding essay on the significance of automated information retrieval for librarians. Each article takes on a distinct subtopic, complementing its neighbors and contributing to a largely satisfying whole.

At the same time, the collection suffers somewhat from not sufficiently tailoring its presentation to its primary audience. The authors are all well versed in the information science concepts underlying KDD, but unfortunately most working librarians lack such familiarity. Each article seems intended to introduce a particular aspect of KDD to the nonspecialist; only a few report new research. It is therefore doubly frustrating when bibliometric jargon and obscure statistical formulas are employed without explanation, as they frequently are in this volume. Such explanation would of course slow down a presentation and annoy information scientists, but by writing as if for *JASIS (Journal of the American Society for Information Science)*, most of the authors have squandered an excellent chance to educate the working librarian and drive home the relevance of their topics.

A recurring theme in this volume is KDD's function of revealing the broader intellectual context of a scholarly work by using computer-aided association techniques to uncover links between two apparently unrelated articles. This process can have dramatic results. For instance, Don Swanson and Neil Smalheiser present a classic example of bibliographic KDD: linking articles through common citations to produce a promising but unsuspected idea for treating migraine headaches. In the following chapter, Kenneth Cory recounts how humanities researchers adapted Swanson and Smalheiser's methods and discovered an undocumented intellectual link between Robert Frost and the Greek philosopher Carneades. Henry Small's 331-article path from economics to physics is a spectacular demonstration of both the power of bibliographic association and the interrelatedness of knowledge. Jian Qin's study of using bibliographic coupling (through common citations) to discover semantic patterns in the literature shows how frequency distributions of keywords can delineate "core" and "marginal" literatures in any subject, and identify interdisciplinary regions; Qin shows how analyzing the co-occurrence of words and phrases in various documents can also perform these functions.

Bibliographic KDD by its very nature depends on probabilistic techniques of text processing. It rests on the assumption that the frequency of certain words or the citation of certain documents provides a reliable clue to

the "aboutness" of a document. The need for such techniques is perhaps obvious when one is dealing with very large databases. This is particularly true when the data are relatively unstructured, as with full texts, abstracts, uncontrolled keywords, or even titles. Another recurring theme of this book, however, is the role of human judgment in KDD and the continuing need for human processing, whether at the front end, providing structured descriptions; at the back end, interpreting search results; or in the middle, normalizing text data to facilitate machine analysis. Swanson and Smalheiser, in describing the process of discovery in medical databases using the Arrowsmith program, emphasize the necessity of professional judgment at all three stages, especially in the analysis of results.

Several other papers promote a less promising mode of human input, having the authors of documents provide a structured surrogate of their work to facilitate database access. While the point is well taken that authors know better than anyone else what their works are about, it does not necessarily follow that they have any special expertise in the organization of knowledge or the larger intellectual context in which their works exist. Nor would the author—unless he or she is also a librarian—likely have a commitment to calling attention to the document only when it meets a searcher's expressed information need. While author-generated surrogates can add value to documents by providing additional data for "mining," they are no substitute for professional indexing or cataloging.

Although he does not address this point directly, Herbert White seems to recognize it at least by implication in his concluding essay. After briefly acknowledging the topic of the book, he discusses on a more general level the limitations of automated information access, the continuing and increasing need for trained information intermediaries (a.k.a. reference librarians), and the forces that prevent people and institutions from recognizing this need. Both his style and his arguments will be familiar to readers of his longtime *Library Journal* column. Though many readers of this volume may see White's contribution as jarringly out of place, it reinforces the recurring theme of human judgment and its role in providing meaning to the induced serendipity of machine-assisted data analysis, and in the process helps the reader make the connection between KDD and the work of librarianship.—*Gregory Wool (gwool@iastate.edu), Iowa State University Library, Ames*

Works Cited

Raghavan, Vijay V., Jitender S. Deogun, and Hayri Sever, ed. 1998. Special topic issue: Knowledge discovery and data mining. *Journal of the American Society for Information Science* 49: 397–470.

Trybula, Walter J. 1997. Data mining and knowledge discovery. *Annual Review of Information Science and Technology* 32: 197–229.

Vickery, Brian. 1997. Knowledge discovery from databases: An introductory review. *Journal of Documentation* 53: 107–22.

Correction

A book review in the April 2000 issue of *LRTS* referred to the "now-defunct Getty Information Institute's 'Introduction to' series" (106). Though the Getty Information Institute closed in 1999, all of the volumes in its "Introduction to" series are still in print; an updated edition of *Introduction to Metadata* is available at www.getty.edu/gri/standard/intrometadata.