

Information Retrieval. By C. J. VAN RIJSBERGEN. Reading, Mass., and London: Butterworths, 1975. Pp. 152. \$14.95. ISBN 0-408-70717-8.

The emphasis in Van Rijsbergen's book is the use of automatic clustering and classification techniques in experimental information retrieval systems. While designed as a general introduction to information retrieval for undergraduates in computer science, graduates in library science, or researchers, I view the book as most suited to advanced graduate students or researchers in the field.

The chapter on automatic text analysis concentrates on statistical methods for content analysis (such as the ideas of H. P. Luhn) and excludes syntactic or semantic analysis. Automatic classification is dealt with extensively in chapter 3, and in varying depths Van Rijsbergen describes how specific clustering algorithms work. Guidelines are given for the type of data to which clustering techniques can be applied, and the mathematical properties of various coefficients of association are explored. The chapter synthesizes many of the ideas that have appeared in the research on clustering and provides a unified treatment of the problem.

The chapter on file structure gives descriptions of all the standard techniques for the organization of records in machine form (sequential, index sequential, inverted, list, and multilist), and the advantages and disadvantages of each are sketched. There is scant attention paid to the distinction between logical and physical storage of records, and little is said about the use of research in data-base management systems as applied to information retrieval. Search strategies are dealt with in chapter 5. Boolean searching is introduced, and the concept of matching using association measures is discussed. Here again the emphasis is on clustering and retrieval using clustering. Feedback techniques, such as those developed by Rocchio, are also presented. The chapter on evaluation covers the standard measures of retrieval performance (recall and precision) as well as the more important alternatives that have been suggested (such as those by Swets, W. S. Cooper, and Salton). Also included in the chapter is some of Van Rijsbergen's own work on a measure of search performance.

The book cannot be considered as a comprehensive text in information retrieval. It is, however, useful in its emphasis on automatic classification and clustering. Its major asset is as a research tool. The references to the information retrieval literature are comprehensive in the areas that are covered and selective in the citations provided.

Michael D. Cooper, *University of California, Berkeley*