References

BARNARD, T. J. (1969). A new rule mask technique for interpreting decision tables, *Comp. Bull.* **13**, pp. 153–154.
BJORK, H. (1968). Decision Tables in ALGOL 60, *BIT*, **8**, pp. 147–153.
CHAPIN, N. (1967). An Introduction to Decision Tables, *DPMA Quarterly*, April 1967, pp. 2–23.
HARRIS, F. T. C. (1967). The partial automation of systems analysis, contribution to Datafair '67.
KING, P. J. H. (1967). Decision Tables, *Comp. J.*, **10**, pp. 135–142.
KING, P. J. H. (1968). Ambiguity in Limited Entry Decision Tables, *CACM*, **11**, pp. 680–684.
POLLACK, S. L. (1967). CR 12948, *Comp. Revs.*, **8**, pp. 501–502.
POLLACK, S. L. (1963). Analysis of the Decision Rules in Decision Tables, Memo RM-3669, Rand Corp., Santa Monica, California.
SPRAGUE, V. G. (1966). Letter to Editor (On Storage Space of Decision Tables), *CACM*, **9**, p. 319.
STOLL, R. R. (1963). *Introduction to Set Theory and Logic* (Chapter 4), W. H. Freeman and Co., San Francisco and London.
WIRTH, N. (1966). A note on Program Structures for Parallel Processing, *CACM*, **9**, pp. 320–321.

Book Review

Automatic Information Organization and Retrieval, by Gerard Salton. 1968. 514 pages. (*McGraw-Hill*, 126s.)

The main theme here is reference retrieval. Straight-forward mechanical aids to essentially manual methods are not really considered, however—the emphasis is on the sophisticated techniques needed to carry out the whole operation by computer, both the initial analysis of each incoming document leading to storage of a summary, and the subsequent retrieval of references to those documents relevant to a particular request. The main problem is to obtain an accurate picture of the meaning or content of each document and request; each is summarised as 'a vector of content identifiers', which vectors can then be matched. Thus the techniques involved include the construction and use of synonym dictionaries, syntactic analysis of English text, cluster analysis, statistical phrase analysis and structure matching techniques. Further, most importantly, ways of measuring the effectiveness of the various techniques are derived.

All these topics are explained, starting from a level to make them comprehensible to a reader with a general understanding of computer processing, but without specialist knowledge: representation of tree structures and the binary search technique, for example, are clearly explained. Rather more mathematical background is necessary: matrix manipulations appear without prior explanation, though the section on mathematical retrieval models is prefaced by an introduction to the necessary set theory.

Professor Salton being one of the pioneers in this research area, these are topics on which he is well qualified to write. The treatment is comprehensive, the material very well organised and the explanations clear. Many of the results tabulated in support of the theories have been obtained on the Smart experimental automatic retrieval system built up in the last few years by Professor Salton and co-workers. This system is indeed described, in a 65-page appendix, in the detail needed for a user to punch the cards specifying the options to be exercised on a particular computer run; do such details really belong in a text of this kind?

Results quoted all seem to relate to quite small collections of documents, which are adequate for comparison of various automatic techniques But little attention is paid to a comparison of either the effectiveness or the economics of these automatic methods with manual systems. In addition to the principal theme of reference retrieval, a chapter is devoted to systems for storing and retrieving facts themselves, rather than references. A number of auxiliary operations—text editing, indexing, selective dissemination of information—are touched upon in a further chapter. It is unfortunate that the sample of output from a syntactic analyser program on page 167 should show an incorrect analysis of the sentence chosen, but this is an isolated flaw in a book where a great mass of detail has been carefully assembled.

This book, then, gathers together the main theories and results of recent work in automatic I.R.; over 300 references are given, as well as a selective bibliography of another 300 items. In addition to those directly involved with I.R., programmers in other applications areas may well find some parts of the text relevant, for many of the techniques explained have wider application.

A. J. MITCHELL (Leeds)