Information Technology: Research and Development (1983), *2* (239–242)

BOOK REVIEWS

M. Loomis. *Data communications.* New Jersey: Prentice/Hall International, 1983. 212 pp. £22.45. ISBN: 13 196469 0.

Because of national differences, both technical and legal, it is difficult to cover the field of communications in a manner which is universally applicable. This book, well described in its preface as introductory, avoids many of these difficulties by dealing with the US environment. Thus much of the history, and some of the terminals described, are of less relevance to users on this side of the Atlantic. The remainder of the material forms a good introduction for the non-technical reader.

Many of the end of chapter exercises require access to material outside the scope of the chapter. Often, these rely on the ready availability of information from vendors; an availability more likely to be experienced in the US than elsewhere. Something might have been said about enciphering, which is mentioned in passing. Also, the description of bit patterns for character transmission could have led to a discussion of bit and byte ordering—sometimes a problem when interconnecting disparate devices.

It is reassuring to meet an author who, when discussing time division multiplexing, knows the difference between time slot occupancy and bandwidth. A useful introduction for the non-specialist, its utility for specialist undergraduates is reduced by the absence of references.

B. B. Byrne
Lecturer in Computer Science
University College Dublin

D. Longley and M. Shain. *Dictionary of information technology.* London: Macmillan Press, 1982. 381 pp. £20.00 hardcover. ISBN: 0 333 327624. £6.95 paperback. ISBN: 0 333 348060.

A technical dictionary is not intended to provide an expert with information about his own field: it should aim to explain concepts, describe devices and define terminology for those who are seeking a general understanding of some topic. In considering such a work, three particular questions need to be asked: does it really cover the field which its title implies? can the entries be understood by an intelligent non-expert? and are the entries accurate in the eyes of an expert? These question will be considered in turn.

In view of its title, this compendium should contain a great deal of information relevant to readers of the present journal. The trouble is, I could not find it. To be sure, there are substantial entries on videotex systems, computer networks, word processing, etc., but precious little on 'real' information retrieval. Starting from what seemed the obvious place, I found 'information retrieval' defined as '. . . techniques for storing and searching large quantities of data and making selected

data available. These techniques can include on line storage, KWIC indexes and database methods. See KWIC, on line storage, database.' Under 'KWIC' I found 'see keyword in context'; under 'on line storage' I found 'In computing, storage devices, and the media they contain, under the direct control of the computer system. See on line'. I was not encouraged. The entry under 'database' ran to 16 half-lines, and was mildly informative.

I next decided to take the direct approach and look up a host of likely topics. Nothing was found under 'fact retrieval' nor 'question answering system', but 'expert system' had a 27-word entry and a cross-reference to 'artificial intelligence'. There was an adequate though brief entry under 'inverted file', whilst the entries for 'search' and 'search key' together totalled 37 words. Nothing at all, however, could be found under 'information science', 'query', 'request', 'similarity', 'overlap', 'centroid vector', 'nearest neighbour', 'stemming', 'suffix stripping', 'conflation', 'relevance feedback', 'probability ranking principle', 'fuzzy set theory', 'index language', 'classification', 'Zipf's Law', 'full text searching', 'text searching', 'recall', nor 'utility'. The entries for 'clump', 'cluster', 'feedback', 'rank', 'index', 'Boolean algebra', 'user profile' and 'precision' had little or no relevance to information retrieval. The entry under 'facetted classification' (sic) said 'a system of identifying elements in a collection of information, such as a book, so that they relate to the requirements of a person seeking information, as distinct from less flexible systems such as the Dewey Decimal Classification'. The entry under Dewey contained no such note of disparagement, but I could find nothing on the BC, LC or UC schemes.

Thinking that perhaps the above topics were too specialized for a wide-appeal dictionary, I went on to try several of the well-known online systems and databases, and indeed found short entries for a dozen or so, though LEXIS, ORBIT, STATUS, STAIRS and (of course) SMART were missing.

What does all this mean? Do we conclude that the editors of this dictionary have been negligent in failing to acknowledge an important segment of their subject, or that the present journal is wrongly named? This question is highlighted by the six-page entry under 'information technology', which must surely be the focal point of the whole dictionary: I could find no occurrence there of the word 'retrieval', nor (except in a non-technical sense) of the word 'search'. There seemed to be no acknowledgement anywhere that the retrieval of pertinent information might actually be a problem. To be fair, a reading of the introduction does make it clear that IR will not loom large in the dictionary, but a couple of dozen entries, totalling two or three pages, does not seem much to ask for.

This book does contain a great deal of information, however, and it does not deserve to be dismissed in such a sweeping fashion. Instead, I propose to rename it 'A dictionary of modern computer technology and its applications', which enables it to be reviewed in a whole new light. Much of the material here is outside my own areas of expertise, enabling me to address my second question: that of intelligibility. I read through 20 or more individual pages, and felt that I understood just about all that I found there. I cannot say that it was interesting, but that can hardly be expected of a dictionary. I felt that (in the areas it covered properly) the book could provide a lot of useful information. I may add that the text is attractive to look at, with a two-column format, entry terms in bold face, and clear print and diagrams.

Turning now to the question of accuracy, I enlisted the help of several colleagues from a wide range of backgrounds. Reactions varied from 'all in all a useful little book' to 'would be better in a computer databank so it would not clutter the

shelves'! The quality of the entries is evidently variable, being adequate in many areas, but flawed or even incorrect in others (e.g., 'p-n-p transistor'). The same applies to the 11 extended entries on major topics: these are mostly good, but one or two are written in a style which is distinctly euphoric and in places quite purple; referring to the essay on 'microelectronics', one of my colleagues was moved to observe that 'Section (d) is journalese and means nothing'. In the essay on programming, the 28 lines of illustrative Pascal contained, on my reckoning, nine errors. Incidentally, since the blurb states that these essays were written by 'specialists in the field', it is puzzling that there is no indication of who these specialists might be.

But it was not so much the accuracy of the entries which provoked comment as their completeness and presentation: there is a general feeling of haphazardness. Many expected entries are just not there (CAD/CAM seems an extraordinary omission). Diagrams are generally not referred to from the relevant entries, so that it is possible to look up 'footprint', 'network topology', 'shadowmask' or 'traffic' without ever realizing that there is a diagram overleaf. There is even a diagram captioned 'cassette' which is 16 pages away from the entries for 'cassette' and 'cartridge', and overleaf from 'compact cassette'. Cross-references are not always well chosen: thus 'see KWIC' leads to 'see keyword in context'.

Of course, the editors faced an impossible task in attempting to cover a field which is both broad and in a state of headlong development. My impression is that they would have succeeded better by leaving out most of the material on photography and printing technology, and instead paying closer attention to the completeness and orderliness of what remained. I believe that many people will find this book useful (if unpredictable) for reference, but I doubt if many will think it worthwhile to spend £20 on the hardback, especially since much of the material will be quite dated in a couple of years time. Personally, I cannot see myself referring to it many more than a dozen times within that period.

I should like to conclude by expressing my thanks to Ron Anderton, Steve Cliff, Colin Hannaford, Brian Jones and Tony Llewellyn for their helpful and thoughtful comments.

C. D. Paice
Department of Computer Studies
University of Lancaster

G. Salton and H. J. Schneider (eds.) 'Research and Development in Information Retrieval', *Lecture Notes in Computer Science 146* (Springer-Verlag, 1983). ISBN 0 387 11978 7.

This book represents the proceedings of a conference jointly organized by the Gesellschaft für Informatik and the information retrieval groups of the Association for Computing Machinery and the British Computer Society, and held at the Technische Universitat Berlin in May of 1982. The aim of the conference was to present basic research ideas in information storage and retrieval, as well as interesting new applications, and thus complements the similar conference held at Cambridge in 1980.

On the basis of the papers presented here two areas are currently of primary importance to research workers in IR. The first of these areas is a continuation of

the work on mathematical models of information retrieval, as exemplified by the work of Salton *et al.* (An evaluation of term dependence models in information retrieval), who compares two methods of encompassing term dependency information within relevance feedback environments, and of Bookstein (Explanation and generalization of vector models in information retrieval), who studies the use of the vector inner product as a matching function under a variety of assumptions regarding the distribution of term frequencies. A particularly interesting paper in this category is that by Robertson *et al.* (The unified probabilistic model for IR) which attempts a synthesis of two approaches to probabilistic retrieval that have hereto been quite disparate in character.

A frequent criticism of much research work in IR is that it has been carried out in the quite unrealistic framework imposed by the standard document test collections, and it is hence noteworthy that the second main area of interest is in the implementation of operational retrieval systems that exhibit features different from those in conventional interactive retrieval services. The main components of such a system are by now well-established, and include the automatic indexing of natural language documents and queries and the provision of ranked output and relevance feedback facilities at search time. Doszkocs describes a system (From research to application: the CITE natural language information retrieval system) in which provision is made for the searching of a file of 600,000 MEDLINE documents, while an alternative approach is suggested by Morrissey (An intelligent terminal for implementing feedback on large operational retrieval systems) who has constructed an elegant front-end to EURONET; it is interesting to note that both of these projects use what is basically the same search algorithm to obtain ranked output from inherently Boolean systems, although the actual implementations are very different. A recent development in the design of experimental IR systems has been their integration with DBMS software and this attractive approach is discussed by Buell and Kraft (LIARS: a software environment for testing query processing strategies) and Biller (On the architecture of a system integrating database management and information retrieval).

Of the other papers, that by Tague and Nelson (Simulation of bibliographical retrieval databases using hyperterms) illustrates the great difficulties inherent in the accurate modelling of the frequency distributions that characterize IR systems while Smith and Lïnggard (Information retrieval by voice input and output) discuss some small-scale experiments with voice synthesis and recognition packages: this paper is of some value not because of what was achieved in the tests, which was rather little, but because of the possibilities for the design of interfaces in the future as the technology improves.

Conference proceedings are generally a bit of a hotch-potch but this one is better than most, and I would imagine that all readers of this journal would find much of interest.

P. Willett
Department of Information Studies
University of Sheffield

Vol 2 No 4 October 1983

Information Technology: Research and Development

Contents

Printed in Great Britain by J. W. Arrowsmith Ltd., Bristol BS3 2NT

ISSN 0144-817X