book reviews

Research and development in information retrieval. Proceedings of third joint BCS and ACM Symposium

C J van Rijsbergen (Ed.)

Cambridge University Press, UK (1984) £25.00 ($49.50) ISBN 0–521–26865–6, 433 pp

The 28 papers of this symposium (held at King's College, Cambridge, UK on 2–6 July 1984) consist mainly of interim reports of projects linked to the classical IR model of document retrieval from mechanized databases operating on surrogate groups of index terms. Although that model has now been worked out to the point at which it would be difficult to justify further refinement in operational systems by practical results, the urge to seek further refinement of this classical model remains strong. There is also scope for tidying up aspects of its probabilistic theory in the search for logical completeness. Meanwhile, however, advances on other research fronts — in database management systems, in expert and office systems — make possible new approaches to the model which also steadily widen the connotation of the 'IR'.

The papers are published in the order in which they were presented to make an interesting mix for the symposium. For review purposes, they are grouped differently to display the trends they reveal more clearly.

Three papers critically examine aspects of the theory of the classical model. Wong and Raghaven (Regina) explore and clarify ambiguities in the vector space models; Peter Bollman (Technical University of Berlin) continues his meticulous critique, establishing two axioms which point to the need to define the *homogeneity* of a document collection; and Don Kraft and Duncan Buell (Baton Rouge) briefly review the stopping rules and outline a model from which they hope to derive retrieval performance measures related to the stopping point.

A larger set of papers considers various ways of improving IR performance. Bruce Croft and Roger Thompson (Amherst) explore the possibilities of an adaptive mechanism which selects the search strategy most appropriate to different kinds of user queries. K L Kwok (CUNY, Flushing) proposes to incorporate terms from the titles of documents cited by those documents regarded by the user to be most relevant. Metzler *et al.* (Pittsburgh) invoke the linguistic theory of Moulton and Robinson to enhance relations between keywords. Frakes (AT & T, Lincroft, NJ) describes a technique applied to CATALOGUE for truncating keywords automatically, while Niedermair *et al.* (Munich) outline the use of MARS, an aid designed to find previously unkown terms morphologically related to the users' search terms. Robert Korfhage (Dallas) presents a method of 'personalizing' queries by means of user profiles.

Database management systems and document retrieval systems begin to interact. Schek (Darmstadt) considers how best to put an IR system 'on top' of a DBMS. Jean Tague (W Ontario) considers the limitations for bibliographic retrieval of relational models developed in DBMS and extends Zaniolo's semantic model; then, by adapting the Wernier/Orr diagram from system theory, she establishes a notation enabling her to set out a logical model for IR in a wider but inclusive sense. Azmoodeh (Essex), Lavington and Standring (Manchester) explain how the Semantic Binary Relationship Model can be linked with semantic networks and logical rules. Omiecinski and Scheurmann (Northwestern) describe a method of record clustering and file reorganization in response to user enquiries which minimizes the number of pages accessed per query. Hollaar (Utah) updates previous reports on the message-based architecture of a text-retrieval system designed to operate with very large bodies of text. Estell and Smith (Queen's Belfast) study the costs and response times of distributing an IR system between a mainframe and a microcomputer acting as an intelligent terminal.

Automatic text analysis becomes increasingly relevant. The peculiarly difficult problem of searching a database of the texts of mathematical papers (some of 100 journal pages or more) densely packed with methods and theorems on the theory of finite groups has been tackled by Iofinova and Komissartschik (Institute for System Studies, Moscow) who describe their interactive search process which seeks elements which are 'near' to that expressed by the user. Hahn and Reimer (Konstanz) use programs that produce 'text graphs' of various depths as the basis for retrieval. Fuhr and Knorz (Darmstadt) describe their program AIR applied to the automatic indexing of papers in physics and display the recall/precision graphs obtained in tests. Kevin Jones and Colin Bell (Malaysian Rubber Producers' RA, Hertford) summarize their experience of ten years' work with their program MORPHS on documents of interest to the rubber industry.

Expert systems are also related to the IR model. French views are presented by Defude (IMAG) who considers how to accommodate thesauri within the architecture of an expert system and by Zarri (CNRS) who describes the inference techniques of RESEDA, an experimental database concerned with personages in the history of France.

Office systems become more complex. The most ambitious project, part of an ESPRIT study, outlined by Christodoulakis (Toronto) is a system in which the messages may consist of any mix of attributes, text, images and voice; it therefore demands a formidable array of

hardware and software and is being developed to provide many kinds of query response in a dynamic office environment. A second paper related to an ESPRIT study of mixed mode message filing systems is presented by Rabitti (CNDR, Pisa) and Zizka (Brno) who describe experimental tests on various access techniques. Laender and Stocker (East Anglia) describe a flexible use of 'forms', applied to a database of gramophone recordings, in which the 'form content' specified by the user becomes the basis for searching the database.

Four papers about more general topics were presented. John Tolle (OCLC) has several extensive transaction log tapes from operational IR systems and is developing methods of analysing the abundant data thus acquired. Data from the NLM covered eight weeks and included 440 000 transactions, 11 000 user sessions and 1 600 hours of system use; but that is only one of the several systems providing log tapes. So Tolle has the complex problem of sorting out all this data to reveal in detail how systems are actually searched.

Barbara Allen (Leeds Polytechnic) assesses the requirements of open learning in educational contexts and considers the kinds of expert system needed. Zeng Minzu (Beijing Document Service) describes the setting up of modern information systems in a country that still lacks the infrastructure of communication systems and professional expertise needed for an integrated Chinese network. He wisely concludes (as Confucius might have done) that: 'Modernization is not purchasable. No hard work, no system available.'

W S Cooper (Berkeley) adumbrates two major areas of study: a general logico–linguistic theory to support the performance of inference-making systems within natural language, and the concept of 'information states' of an organism or machine and the changes of such states when new information is acquired. He points to the need for work to be done in both these areas.

The long, long history of logic — from Aristotle through Leibniz, Boole, Frege, Russell, Wittgenstein, Carnap and Bar-Hillel to Quine *et al.* — suggests that the natural languages are not wholly reducible to logical forms, and necessarily so. Advances in this direction, such as those reported at this symposium, might therefore proceed modest step by modest step towards increasing sophistication. These steps will be applied to databases operating with languages which, wittingly or not, consist of some controlled subset of natural language and which suffice for specific purposes. There seems to be no end to this step-by-step approach. It may be logically untidy, but it is both necessary and more psychologically satisfying than the more direct approach that Cooper appears to advocate.

But logical inferences are not all we need. Mechanized systems will, in due course, be expected to point to logical contradictions or to gaps in the information states of their databases, to suggest logically incomplete but plausible ideas — hypotheses — as well as firm conclusions. In short, this reviewer gives priority to Cooper's suggested study of information states because that study holds the key to our understanding of *all* the information processes with which the study of IR is concerned.

The editor and publishers are to be commended for the speed of publication of the proceedings, although it is regrettable that discussions evoked by its stimulating papers were not also reported. But the trend of IR research is evident: the development of IR is fast breaking free from the confining bounds of the classical model.

B C Brookes
London, UK

An introduction to computer-based library systems (2nd Ed)

L A Tedd

Wiley Heyden, UK (1984) £10.50, ISBN 0471 26285 4, 262 pp

The publication of the second edition of Lucy Tedd's textbook is long overdue and will be welcomed by students, teachers and practitioners of librarianship and information science. Because of its emphasis on state-of-the-art technology and the inclusion of details of specific systems, the first edition (1977) had dated very rapidly.

In the new edition, the general organization of the previous edition is preserved. The first chapter gives an overview of computer-based library systems, the next four chapters deal with hardware, software, storing and retrieving information and telecommunications, the sixth covers general aspects of setting up computer-based systems in libraries, and the remaining five treat the computerization of specific library systems e.g. circulation control systems. The chapter dealing with telecommunications is an appropriate indication of the growth of this area since 1977. Other changes from the first edition include the merging of separate chapters on acquisitions and cataloguing systems, which reflects current practice, while information retrieval systems have been subdivided into two chapters, local and external. Surprisingly, videotex systems are not treated under external information services but appear in the chapter on telecommunications aspects.

When one examines the coverage of this textbook, the emphasis is heavily on current practice in library/information centres. In reviewing any book of this type, it is tempting to identify specific software packages or systems that should have been included, but there are no glaring omissions that this reviewer could identify. However, the topic of integrated systems is very poorly

treated at both the conceptual and the practical level. A mention in the preface and cross-references at certain points in the text are insufficient for this important topic.

The author has succeeded admirably in drawing together a coherent text on computer technology and its application to libraries. Her ability to express technical concepts simply and concisely and to extract the significant points from a myriad of details make this a very valuable textbook. The level is such that it is suitable for readers without a technical background, be they undergraduate or postgraduate students or practitioners. But the author's desire to produce an introductory text may also have some undesirable effects. A textbook, even if its primary aim is to be an introduction, should wet the reader's appetite for more. Thus, in this reviewer's opinion, the lack of concern with future trends is a fault of this otherwise excellent volume. A few lines on electronic publishing in the introductory chapter and a section entitled 'Related Technological Developments' in the chapter on telecommunications are the closest the author comes to looking to the future. The latter section outlines the operation of electronic mail systems, telefacsimile, electronic document delivery projects and electronic network communication projects i.e. computer conferencing. This two-page unit is difficult to find because of its inclusion in the telecommunications chapter and because it is poorly indexed.

Another deficiency in the text is the lack of an overall analysis of the implications of current and future technological developments for libraries and information centres. For example, the reader can find out what a databank is and obtain details of some search services of this type but he/she gets no indication of the relative importance of this type of information source in the future. The same criticism would apply to the handling of many other current developments e.g. videotex systems, automatic document ordering, electronic journals, etc. The emphasis on 'documentary' records also results from the concern with current practice rather than future trends.

While the text of the chapters dealing with specific systems gives wide international coverage, the detailed examples of systems at the end of each chapter are all UK based. Although Ms Tedd justifies this emphasis in the preface, one cannot share her confidence that this feature will not detract from the international applicability for the text. In particular, it must be a disadvantage for the US and Australian markets. The value of including these examples, which are rather detailed descriptions of operational systems, is doubtful. It is unnecessary to give such detail in a textbook, when readers have access to reports of these (or similar) systems in the primary literature. This emphasis on details of present systems rather than principles or objectives will also result in the work dating more rapidly.

The very useful glossary of terms and the list of acronyms have been updated since the first edition. The glossary, apart from its value as part of the text, is a valuable reference item in its own right.

The index is rather patchy with some glaring omissions. There are no entries for terms such as 'optical discs' or 'intelligent terminals'. On examination of the previously mentioned section on Related Technological Developments, one finds entries in the index for 'facsimile transmission' and 'electronic journal' but none for the concepts of 'electronic mail' or 'electronic/automatic document delivery'. Although all four projects are referred to, Blend and Hermes appear in the index while Adonis and Artemis do not. Of a more serious nature is an error in page numbering of index entries for pages 219–237, where each entry appears to read three pages too high.

Overall, the second edition of this textbook by Lucy Tedd is an extremely useful publication, which is technically sound and gives a perceptive and lucid analysis of current developments in library automation. But, like the first edition, it will date very rapidly. It is also of value to members of the library and information profession, who need an introduction to the basics of computerization or an overview of current practice, particularly if they are interested in UK examples. Related professionals (computer scientists, systems analysts, etc.) interested in library systems would also find this item a useful addition to their reference bookshelf. The potential market should justify the publication of a paperback edition, which would certainly increase sales to the student population.

M Burke
Department of Library & Information Studies
University College Dublin, Eire

Computerized literature searching: research strategies and databases

L Gilreath

Boulder: Westview Press, USA (1984) £20.25, ISBN 0-86531-526-4, 177 pp

This is a well written and interesting introduction to making the best use of online bibliographic databases. It is suitable for use by 'end-users' without formal library training, although librarians and information officers new to online services may well find it useful as a starting point before going on to the more exhaustive sources.

Two preliminary chapters provide a brief introduction to the way in which online systems work and the role of the search intermediary. (Gilreath comes down fairly firmly on the need for 'search analyst' help during the preparation and performance of a search.) The remaining

six chapters develop 'for the end users of information' an account of six subject areas: agricultural and life sciences; social sciences and education; physical sciences and engineering; business literature; humanities, arts and architecture; and legal research. In each case, the main databases in the field are listed, with notes on their contents and sometimes on their idiosyncracies, and the reader is introduced to the indexing practices that have evolved for each field.

The approach is very practical, and examples are provided illustrating the recommended strategies. For instance, the role of *CAS Registry Numbers* in making the complexity of chemical naming less intractable is illustrated by a *Chemical Abstracts* citation and then discussed in the context of patents and toxicology literature. The problems of geographic location are recognized and suggestions made for developing effective search profiles. Terminological drift in the social sciences and difficulties in the searching of value ranges in physics are clarified, and the ways in which searches may fail to deliver all relevant material are explained.

The use of thesauri in developing search strategies is taken up in each of the subject chapters, and the principal systems are described. Printed sources such as *Enzyme Nomenclature* are noted as well as the MeSH group of search aids, and, in general, a combined approach using both online and printed sources is encouraged.

The author is head of the Automated Information Retrieval Service for Texas A & M University Library. His practical experience of searching in a wide range of sources is evident throughout the book, and not only in the chapters on life sciences where he is a specialist. On the other hand, the choice of examples and the lists of relevant databases do show some bias towards North American material, and will need to be supplemented for European readers. (To be fair, BHRA Fluid Engineering, Chemical Engineering Abstracts, GEOARCHIVE, INSPEC, PIRA Abstracts and RAPRA *are* cited, even if Eurolex and World Reporter are missing.) *Online bibliographic databases* by Hall and Brown (Aslib, 1983) is much more detailed, and gives better coverage of European material; Gilreath is perhaps complementary in providing 'how to go about it' guidance.

At half the price, this would have been a good book to lend prospective users of online services before they approached the library to run a search; at £20.25 one copy chained to the reference desk seems more likely.

J H Ashford
Maidenhead, UK

Microelectronics and office jobs: the impact of the chip on women's employment

D Werneke

International Labour Office, Switzerland (1983) ISBN 92 2 103278, 102 pp

Recent advances in microelectronics technology have led to much faster and more powerful computing equipment being available at substantially lower costs than in the past. Much of this equipment is being designed for office work and, since the majority of office workers are women, the effects of the technology will be seen most clearly in women's employment. A World Employment Programme project, sponsored by the International Labour Organization, looked into this subject, and this book is a presentation of their findings. The object of the study was 'to provide an overview of the issues and an analytical survey of the recent literature on the subject; in so doing, it is hoped that in-depth and penetrating work on the subject will be forthcoming'.

The author begins by describing the items of office equipment that have been developed from the advances in microelectronics. Word processors, facsimile equipment, optical character recognition equipment, computer output microfilm, minicomputers and the integrated electronic office are all described at the first chapter.

The book then goes on to discuss the influence that technical changes in the past have had on employment and contrasts this with the current economic situation. The likely effects on female workers of a rapidly increasing spread of new technology into the office are also discussed.

The chapter on 'Technical change and labour displacement: North American experience' summarizes the results of detailed sudies in the USA and Canada. Banking, insurance and general office work have been studied, with the general conclusion that the growth rate of clerical staff has been slower than the business growth rate. In most cases, improved information services have been achieved also.

The North American situation is also reviewed in the chapter on 'New technology and work content and organisation in North America', which also discusses job satisfaction, the changed working environment and health and safety considerations.

The European situation is reviewed in a chapter looking at the 'OECD experience', and the conclusion is that the pattern is similar to the US experience, but that the rate is slower.

This book provides a very useful summary of the effect of the chip on women's employment at the present time. The work is well supported by detailed references at the end of each chapter, and the stated objective of the study appears to have been well satisfied.

C L S Stainton
School of Studies in Computing
University of Bradford, UK

book reviews

The application of mini- and microcomputers in information, documentation and libraries

C Karen and L Perlmutter (Eds)

North Holland, The Netherlands (1983) DFI 200/S 85, ISBN 0-444-86767-8, 801 pp

Information processing and management has been undergoing a continuing evolution along several paths over the past ten years, particularly with the advent of mini- and microcomputers. Also, the availability of various software packages has given a new dimension to the problems of information storage and retrieval and has obviated the need for that brand of professionals, 'the information providers', who provide metainformation rather than solutions to individual problems.

There is an ever-increasing need for educational and awareness courses, as well as suitable publications to help individuals understand the various approaches to intelligent information handling and documentation. What is desperately needed is suitable 'overview' material to help the person who is planning a new system by means of existing software, or by in-house programming. This book fulfils this need to a large extent and also offers information on what has been achieved in practice by various practitioners and institutes of learning.

At first sight, the book gives the impression of being just another publication resulting from an international gathering of sun-loving academics. It does, in fact, present the proceedings of a conference held in Israel in March 1983, but this book is exceptional in that it places emphasis on practical approaches to the subject concerned. It is also unusually uniform in its presentation (this is not always the case with conference proceedings) and is readable, since most of the papers are short and very synoptic, considering the complexity of the problems and the solutions they present.

The need for controlled change in libraries and information organizations is greater than ever before. The individuals that have been charged with providing access to printed information and knowledge must adapt to providing access to electronic media or they will be replaced as sources of information. Change to electronic formats will provide increasingly convenient, rapid, rich and cost-effective access to information. The model for the library of the future must therefore be committed to two basic principles: (a) the widest possible access to information, and (b) the use of computers to increase and manage information resources.

The problem is not so much to capture and store knowledge as to manipulate it to satisfy specific user requests. To be able to maintain intelligent systems, it is necessary to adopt data structures which can be traversed as and when required in the minimum possible time. The larger the database, the more crucial the choice of the structure will be and the more machine resources it will require. Although the title of this book suggests that the work is particularly relevant to micro- and minicomputers, most of the problems and solutions presented are equally applicable to large mainframes. What does not come out clearly is material on documentation and its relationship to the actual information service environment.

There are over 80 papers, which present work on the characteristics of information systems, data management, text acquisition and retrieval, relational database systems, indexing, archiving, language translation, library systems, hardware characteristics and others. As is the case with most conference proceedings, the papers include substantial duplicate material, particularly when each is put into its right perspective, and introductions to the concepts and methodologies involved. With such a multiplicity of topics as well as substantial overlap among them, it is very difficult to establish appropriate sessions or headings for classification. The editors have however managed to group them under 16 appropriate headings:

- hardware and peripherals,
- description of existing systems,
- application software and requirements (sessions 3 & 10),
- minis and micros for library and information science education,
- applications in chemistry; graphic retrieval systems,
- improving man–machine interaction,
- applications in agricultural libraries and information centres,
- applications in education,
- problems of developing countries,
- applications in libraries and information centres; general,
- system design considerations,
- applications in libraries and information centres; technical processing,
- minis and micros for online searching,
- applications in numerical and factual information,
- networking aspects of information services and libraries.

There is such a wealth of material on so many different aspects of information handling that the reader may find it difficult to trace specialized information, particularly as the index provided is not adequate and does not do justice to the book. Also, cross-referencing proves very difficult as a result of this.

In conclusion, the book will be of great help to all concerned with the issues it tackles, and this includes librarians, information system designers and students at both undergraduate and postgraduate levels.

E J Yannakoudakis
Computer Centre
University of Bradford
UK

calendar

Date	*Title*	*Organizers*	*Place*	*Other details*
5–7 March	Hardware and Software Components and Architectures for the 5th Generation	AFCET, 156 Boulevard Péreire, F.75017 Paris, France. Tel: (1) 766 24 19, Telex: 290 163 Code 235	Paris, France	
18–20 March	Computer Assisted Information Retrieval	Contact: E Chouraqui, C.I.D., 36 bis, rue Ballu, 75009 Paris, France	Grenoble, France	
25–27 March	Fourth Symposium on Principles of Database Systems	Contact: S J Ginsburg, Dept. of Computer Science, University of Southern California, University Park, Los Angeles, CA 90089–0782, USA, Tel: (213) 743 5501	Portland, OR USA	
26–28 March	Eighth Information Technology and Office Automation Exhibition	BED Exhibitions Ltd, 44 Wallington Square, Wallington, Surrey SM8 8RG, UK. Tel: 01–647 1001, Telex: 893069 BEDATA	London, UK	
16–18 April	Informatics 8: Intelligent Knowledge Based Systems	Contact: S Jespersen, Manager, Professional Development Group, Aslib, 3 Belgrave Square, London SW1X 8PL, UK. Tel: 01 235 5050, Telex: 23667	Oxford, UK	Includes automatic text analysis, pattern recognition, intelligent front ends, natural language processing, retrieval algorithms and integration of IR systems
23–25 April	Communications — London	Contact: T Brandon or P Mead, Communications — London, Industrial and Trade Fairs Ltd., Radcliffe House, Blenheim Court, Solihull, West Midlands B91 2BG, UK. Tel: 021–705–6707, Telex: 337073	London, UK	Exhibition and conference covering electronic communications, IT equipment, business services
22–24 May	International Conference on Foundations of Data Organization	Contact: Prof Yahiko Kambayashi, Dept of Information Science, Kyoto University, Kyoto, 606 Japan. Tel: 75–751–2111 Ext. 5382	Kyoto, Japan	
28–31 May	ACM SIGMOD: International Conference on Management of Data	Contact: S Navathe, Computer and Information Sciences Dept., 512 Weil Hall, University of Florida, Gainesville, FL 32611, USA. Tel: (904) 392 7440	Austin, TX USA	
5–7 June	ACM SIGIR Conference on Research and Development in Information Retrieval	Contact: Prof. V Raghaven, Dept. of Computer Science, University of Regina, Regina, Saskatchewan S4S OA2, Canada	Montreal, Canada	
21–23 August	Eleventh International Conference on Very Large Databases	Contact: A Pirotte, Philips Research Laboratory, 2 Avenue Van Becelaere, 1170 Brussels, Belgium or Y Vassiliou, New York University, Graduate School of Business Administration, 100 Trinity Place, New York, NY 10006, USA	Stockholm, Sweden	
2–5 September	HCC 3, Third Conference on Human Choice & Computers	Contact: Sven Jonasson, HCC 3, Local Conference Chairman, Swedish Society for Information Processing, Rindogatan 8, S-11536 Stockholm, Sweden	Stockholm, Sweden	Authoritative national reports on emerging national policy and social impacts of computers
15–18 September	IRFIS 6 Intelligent Information Systems for the Information Society	Contact: Prof. A Serrai, Chair, IRFIS 6 Organizing Committee, Via Stazione S Pietro, 22, 00165 Rome, Italy	Frascati, Italy	Intelligent systems in information retrieval, human/computer interface

INTERNATIONAL ELECTRONIC PUBLISHING RESEARCH CENTRE

If you are involved in publishing and distributing information you have probably considered some form of electronic publishing.

- You may have tried a few alternatives already
- You may have found it does just what you want
- You may have found it cost you a lot of money
- You may have found that it is a lot more difficult to match markets and technologies than anticipated
- You may not have tried it at all
- You will have realised that there is a lot of potential, and a lot of problems

This is why nearly 60 of the major companies worldwide have joined the IEPRC.

Activities are governed by the members and are directly related to their needs. Sharing the costs of solving common problems and sharing the experiences of people in similar situations makes sound commercial sense.

Recent and Current Activities

Study tour to Japan
Conferences
Technical, market and economic scenarios for electronic publishing
Study tour to USA
Regular bulletin on electronic publishing— EP journal
Electronic network for rapid and informal interchange between members
Electronic Publishing Abstracts

The International Electronic Publishing Research Centre is a non-profit institute dedicated to applying electronic technologies to the publishing process.
It provides an international forum for exchange of research ideas and information between publishers, users and suppliers concerned with electronic publishing.
If you are a publisher of any sort, or a systems supplier, service supplier, printing company, or research organisation, find out more about the IEPRC and its work. *For more details contact:*

THE INTERNATIONAL ELECTRONIC PUBLISHING RESEARCH CENTRE
Pira House, Randalls Road, Leatherhead, Surrey KT22 7RU, England
Telephone: Leatherhead (Code UK+44) (Code Area 0372) 376161 Telex: 929810

APPLYING THE 68000 FAMILY

More and more systems with a 32-bit architecture are hitting the market. Many are based on the 68000 family of chips. For people contemplating or already using such a system, *Microprocessors and Microsystems* presents the state of the art in 68000-based computing.

Articles cover applications, VMEbus, Xenix and Pick, and MODULA-2, POPLOG and LISP. The issue also includes a complete introduction to the 68020 processor.

The collection is essential reading for systems designers, systems engineers and OEMs.

For further details please contact:

Anne Browne, Editor, Butterworth Scientific Ltd.,
PO Box 63, Westbury House, Bury Street, Guildford, Surrey GU2 5BH, UK.
Telephone: (0483) 31261 Telex: 859556 SCITEC G