3.4 *Semantics*

There are a large number of semantic problems in practical data bases. These are not peculiar to the relational model but nor are they solved by it. McLeod (1978) gives a comprehensive listing of these. However, again because of the simplicity of the environment, semantic problems do not seem to be severe in document retrieval. For example it can be seen that every relation we have given is naturally in fourth normal form.

4. Conclusions

Despite the possible problems noted above, we feel that it is apparent that any shortcomings of the relational model are far outweighed by its potential advantages and it certainly deserves consideration as a practical model for the design and implementation of document retrieval systems. One of the most serious problems in DRS research at the present time is the lack of a suitable framework within which to carry out experiments. Proponents of new ideas often find themselves in the position where they virtually have to build a new retrieval system from scratch in order to obtain some practical results. This has been, and continues to be, a major obstacle in the effective evaluation of new techniques and is probably a major reason why so few experimental ideas ever carry over into practical systems. Hopefully our examples above show how easily retrieval systems based on quite different concepts can be represented naturally within a relational model. One justification of a relational system therefore is as an experimental test bed within which to evaluate new techniques.

In a general system, response time for queries will always be somewhat slower than those of a special purpose document retrieval system. However, as we have shown above, retrieval in conventional systems is no different from basic retrieval in a relational system. Thus there is no fundamental reason why practical working systems cannot be developed using the relational model. In return for some performance sacrifice, access will be gained to a much more powerful system. Yet at the same time we have a very simple basic system. As our discussion on the representation of relations shows, a relational file organisation can be identical to that currently used in many retrieval systems, yet a flexibility and adaptability can be provided which we have not seen demonstrated on any existing conventional system.

Acknowledgement

This work was partially supported by a grant from the National Science and Engineering Research Council of Canada.

References

CHAMBERLIN D. D. *et al.* (1976). SEQUEL 2: unified approach to data definition, manipulation, and control, *IBM Journal of Research and Development*, Vol. 20, pp. 560–575.

CODD, E. F. (1970). A relational model of data for large shared data banks, *Communications of the ACM*, Vol. 13, pp. 377–387.

DATE, C. J. (1980). An introduction to the unified database language, *Proceedings of the Sixth International Conference on Very Large Data bases*, Montreal, pp. 15–29.

DATTOLA, R. T. (1979). FIRST: flexible information retrieval system for text, *Journal of the American Society for Information Science*, Vol. 30, pp. 10–14.

HELD, G. and STONEBRAKER, M. (1978). B-trees re-examined, *Communications of the ACM*, Vol. 21, pp. 139–143.

IBM (1972). *Storage and Information Retrieval System (STAIRS) Program Reference Manual*, Form No. 5734-XR3. IBM Corporation, New York.

KNUTH, D. E. (1973). *The Art of Computer Programming: Searching and Sorting*, Vol. 3, pp. 473–480, Addison-Wesley, Reading, Mass.

MACLEOD, I. A. (1979). SEQUEL as a language for document retrieval, *Journal of the American Society for Information Science*, Vol. 30, pp. 243–249.

MACLEOD, I. A. (1980). The Mistral/11 retrieval system, Technical Report, Department of Computing and Information Science, Queen's University, Ontario.

MCLEOD, D. (1978). A Semantic Data Base Model and its Associated Structured User Interface, Technical Report MIT/LCS/TR-214, Laboratory for Computer Science, Massachusetts Institute of Technology.

NATIONAL LIBRARY OF MEDICINE (undated). *MEDLINE Reference Manual.* Bibliographic Services Division, National Library of Medicine.

RIJSBERGEN, C. J. Van (1970). A clustering algorithm, *The Computer Journal*, Vol. 13, pp. 113–115.

SALTON, G. (1971). *The SMART Retrieval System—Experiment in Automatic Document Processing.* Prentice-Hall, Englewood Cliffs, NJ.

SEIDEN, H. R. (undated). *ORBIT System Information.* System Development Corporation, Santa Monica, California.

STONEBRAKER, M. (1980). Retrospection on a database system, *ACM Transactions on Database Systems*, Vol. 5, pp. 225–240.

UNIVERSITY OF STANFORD (1972). *SPIRES User's Manual.* Systems Documentation Library, Stanford Computation Center, University of Stanford, California.

Book review

Distributed Data Bases, by I. W. Draffan and F. Poole, 1981; 374 pages. (*CUP*, £15·00)

This book is based on an advanced course held at Sheffield in 1979 as part of the CREST series. The text consists of a collection of papers covering many aspects of distributed database systems. The papers illustrate very clearly that distributed database technology covers not just the now familiar database aspects, but a far wider area including such fields as communication networks and query analysis, decomposition and distributed processing. Furthermore, many of the problems and difficulties associated with non-distributed database systems become even more taxing in a distributed environment, for example, integrity, reliability, consistency, access control, performance and recovery mechanisms.

An obvious danger in presenting a series of papers within a single volume is the lack of continuity of discussion from one chapter/paper to the next. Nevertheless, the book covers a wide variety of topics beginning with a paper which places distributed database systems in perspective discussing such aspects as communications links, design philosophies and distribution costs. A later paper discusses the components of a distributed system and identifies many of the areas of database technology made more difficult by distribution. Other areas such as query decomposition, optimisation and distributed processing are also identified and form the subjects of later papers. The book also discusses concurrency control and processing synchronisation, update strategies when data is replicated and recovery procedures. Several papers cover aspects of distributed database administration and control from design considerations to granting access, testing and policing procedures. Finally, as an appendix, a case is made against the CODASYL proposals being adopted as an ANSI standard. The argument is not convincing, but it begs more fundamental questions relating to the usefulness, expected life-span and investments placed on national standards, especially in such a rapidly changing area as computer technology as a whole.

I found the book interesting despite areas of repetition among some of the papers. Those which covered distributed query analysis, optimisation and processing were of particular interest, as were those which discussed design philosophies and distributed architectures. This book should be of interest to those engaged in all aspects of database research as it serves a useful purpose in bringing together a series of papers covering an expanding, highly topical and intellectually demanding field.

D. M. R. BELL (Aberdeen)