



Book Reviews

Materialized Views: Techniques, Implementations, and Applications. Edited by: Ashish Gupta and Inderpal Singh Mumick. Cambridge, MA: The MIT Press; 1999; 616 pp. Price: \$50.00 (ISBN: 0-262-57122-6.)

In the forewords of this work, Jeffrey Ullman defines “materialized view, as follows: *In essence, any index on, or summary of, a database is a materialized view.* The assertion describes with crystalline clarity the otherwise complex matter that is topic of this book. It is well documented (Oracle SQL High Performance Tuning. Guy Harrison. Prentice Hall, 1997) that appropriate index design “makes possible” single-table queries that otherwise would not be exploitable (an observation shared by any database developer as well as by the reviewer). Similarly, materialized view is the tool of choice to allow interactive resolution of queries involving multiple base tables (of local or distributed databases). Thus, a well-selected collection of relevant papers covering several aspects of materialized view is a needed resource for any database developer and administrator. This book is of high impact in Information Retrieval [IR] since often the data used by IR is stored in Database Management Systems [DBMS] and since much work of IR is influenced by high performance data integration and mining. All of these applications require often very large “joins” and interactive performance. Shall not be ignored that the first implementation of the SMART information retrieval prototype (Salton 1983) was using INGRES DBMS to store its data (Fox, E. (1983) Some considerations for implementing the SMART information retrieval system under UNIX. Technical Report TR83-560, Cornell University). More recently, Grossmann and Frieder (Information Retrieval: Algorithms and Heuristic. David A. Grossmann and Ophir Frieder. Boston: Kluwer Academic Publishers, 1998) describe in thorough way their own implementation of a purely SQL-based IR system. Also some commercial IR systems (e.g., BasisPlus, Information Dimensions, Inc., Dublin, OH) are built on top of a DBMS and make use of its functionality (Information retrieval A Health care perspective. Hersh, W.R., Springer, 1996). Finally, the TIPSTER architecture (R. Grishman. TIPSTER Architecture Design Document Version 3.1. Technical report, DARPA, 1998) could as well be easily implemented by a DBMS for the storage and retrieval of the annotations (Natural Language Information Retrieval. Strzalkowski, T. Boston: Kluwer Academic Publishers, 1999). Nevertheless, it must be clarified that this book is not about impact of materialized view on IR, and that the reference (p. 126) to Salton’s IR text (Salton 1983) unfortunately only concerns natural language query systems for user interaction.

The editors, Ashish Gupta and Singh Mumick, are recognized authorities in the field and have impressive curriculum in the Academia as well as in the Industry. Although this being their first book, both Editors have published milestone papers and patented own algorithms. Not only theoretically inclined, they have been involved in standardization of SQL3 and founded own Companies. Their visions and experiences have influenced both the academic field as well as the development of IBM DB2 and Oracle 8 DBMS.

Following the authoritative foreword of Jeffrey Ullman, which presents the work and appropriately positions it from an historical perspective, the book articulates over 5 parts, covering 32 chapters, a rich literature reference and a concise term index. Each part is introduced by a summary, which allows to rapidly evaluate the content of the individual chapters and the role they play in the book. Twenty-six chapters are based upon publications from peer-reviewed conferences and journals, spanning a period from 1987 to 1997. The remaining six chapters are written specifically for this book and supplies: a useful framework to understand the materialized view (the whole part 1), a review on incremental maintenance of recursive views (chapter 12) and a “materialized view” perspective of Oracle7 snapshots (chapter 26).

Part I introduce to views, the problem of database warehouse maintenance and applications of materialized view. The concluding section stimulates the reader with open scientific questions.

Part II describes into depth the application of views and formulates problems such as self-maintenance, database integration, query optimization, query rewriting, dynamic queries.

Part III is entirely devoted to the cumbersome and partially still unresolved problem of maintenance of materialized view. It discusses incremental maintenance, efficiency of different maintenance strategies, handling of duplicates, deferred maintenance, Datalog queries, chronicle algebra, warehousing, derived relations, updates independent queries.

Efficiency in On Line Analytical Processing certainly is one of the most demanding task to be exploited by a DBMS. In Part IV, the role played by materialized view on data cubes and multidimensional aggregates is presented to the reader in a thorough and exciting way.

Performance, a primary requirement for any DBMS application, and the motivation for the creation and handling of materialized view, is further analyzed in the last part of this manuscript. Different implementations are discussed and novel materials presented for a complete coverage of the field. The previously unpublished material of chapter 26 presents Oracle 7 snapshots as “precursors” of materialized view. Unfortunately, however, this chapter is mostly of historical interest since the recent Oracle8i have raised the materialized view to “first class citizen” with the statement “CREATE MATERIALIZED VIEW”: a yet scarcely documented feature.

The 401 publications, listed in the sixth part of the book, are a rich set of valuable references to the field from 1962 until 1997.

The choice of only 246 index terms clearly reveals the very focussed domain upon which the Editors so diligently have annotated this work. It is worth to note the terms more frequently indexed: Deferred view maintenance (10); Immediate maintenance (9); Advanced Database Management System, Materialized views (8); On Line Analytical Processing, Data warehousing, Incremental Access Method, Incremental view maintenance (7); Data Integration, Data visualization, Hypothetical relation; Query modification(5). These are certainly the topics of highest focus in the whole manuscript.

Full address information is given for each of the 52 contributors, this constitutes a useful resource to promote communication between readers and Authors and thus stimulate scientific and technological advancements in this exciting field.

The editors had the dual objectives of collecting important work done on views as well as provide a reference for students and implementers. These goals are not easy to achieve since most of the existing work done on views is presented at conferences, seminars and workshops for which proceedings are not always available. Finding appropriate literature

on the materialized view is difficult also since these are often focus of industrial applications and as such either encoded in patents hard to find, or kept in the proprietary knowledge pools. Materialized view follows the rigor of the underlying relational model and thus not only “practice” but as well “theory” papers must be appropriately selected, in order to accomplish the purposes of the editors. The goals are however well achieved, by a careful selection of prominent authors, identification of most relevant aspects, co-authorship in 13 of the 32 chapters, unification of the literature, choice of specific terms for indexing.

This text is the first one specifically addressing the topic of materialized view. Written for administrators, scientists and developers of DBMS, its special blend of authors and contents (which satisfies the logic and the mathematically inclined reader as well as the algorithm designer and the database administrator) certainly makes it a textbook for advanced audience. One of the greatest values of this book is its clarity, as well as its depth, of a topic that if superficially could be thought as a “minor”, constitutes a bottleneck and often a very crucial limiting aspect of IR.

This book contributes to the field, being a “milestone” for any further study and research, and by conveying both the state of the art of the scientific discovery as well as algorithms for implementations. The unresolved theoretical and practical issues it raises provide a rich stimulus for further research and development. It is therefore expected that its “shelf-life” will last for several years.

The topics covered by this monumental book are well represented in their complexity and preciseness by the “physical properties” of the book itself: large, thick, heavy, with big fonts, almost no pictures, strictly black and white and with almost no miss-typing. The work is available as hardcopy only. An electronic version would be very much appreciated, especially by the mobile reader.

Luca I.G. Toldo

Merck KGaA

Darmstadt, Germany

E-mail: Luca.toldo@merck.de

Digital Libraries. William Y. Arms. Cambridge, MA: The MIT Press; 2000; 344 pp. Price: \$45.00 (ISBN 0-262-01880-8.)

The term “digital library” has been widely used since 1993. It describes the “hot topic” and “grand challenge” area that deals with the entire Information Life Cycle for digital content. Other related terms, such as “electronic library”, are less commonly used. By the end of 2000, there will be roughly 2000 reports and other publications in the digital library area, spread across a variety of conference proceedings and journal/magazine issues—but only a few good books.

Digital Libraries is the first in a new book series on this topic from MIT Press, written by the editor of that series, a Professor of Computer Science at Cornell University. In 1965, MIT Press published the precursor work “Libraries of the Future” by JCR Licklider, now out of print (see Panel 1.1 in the book by Arms). For a current view of the field, Arms’ book is clearly more appropriate. Indeed, in my two Fall 2000 classes entitled Digital Libraries,

one for honors students in their second or third year of undergraduate study and the other for advanced graduate students in computer science, students almost universally found this book to be more up-to-date than the other book they were required to read (i.e., the 1997 book by Michael Lesk). Clearly there has been rapid progress in the last three years, and a great deal only predicted by Licklider has become reality over the intervening decades.

William Arms has been an active participant in this unfolding of the field of digital libraries. Since 1993 he has served as Vice President at Carnegie Mellon University (CMU, in Pittsburgh, Pennsylvania, USA) and the Corporation for National Research Initiatives (CNRI, in Reston, Virginia, USA). At CNRI he launched and served as editor for *D-Lib Magazine*, the most commonly referred-to publication in the field. *Digital Libraries* draws extensively on that magazine and on Arms' experience with digital library proposals, projects, publishing, publications, meetings, workshops, and conferences.

Arms, as author, has a light touch, making the work easy to read. *Digital Libraries* has some of the flavor of a magazine. Bright second year college students majoring in the humanities had no trouble reading it nor did they have difficulty with the quizzes I prepared (to ensure that they thoroughly comprehended the content). Doctoral students specializing in digital library research also enjoyed the book, though they wished it had more technical details and provided more guidance on where to find further information on topics of interest. All enjoyed the "panels", grey regions of up to several pages that have analogous function to sidebars in a magazine. These panels cover case studies and examples, and illustrate the many important concepts that Arms so clearly and easily explains.

Digital Libraries has 14 chapters. There are 9 line-drawing type figures, 67 panels, 2 tables (for the ASCII characters and for Dublin Core metatags), 11 pages of glossary, $2\frac{1}{2}$ pages of index, and a table of contents covering less than a page. There are no references. There is no use of color. There are no tables of contents for figures, panels, or tables. There is no separate index of names. The single footnote explains a bit about the title of chapter 7. The figures are unevenly distributed, with 1 each in chapters 1 and 11, and the remaining 7 spread over chapters 7–9. The half page of Acknowledgements credits 34 people, two places where earlier versions of chapters (4, 8) appeared, plus 3 sources for that many panels. Readers who want more detail on topics mentioned should use a digital library or WWW search service to find such details, based on the descriptions given in the book. *D-Lib Magazine*, available from www.dlib.org, is a particularly good place to look.

Some of my students voiced valid complaints. There are a moderate number of typos and other minor editorial problems that should be fixed in a next edition. Since I gave them open book quizzes, the students were unhappy that the index is not very detailed and that the table of contents only goes down to the chapter level. They found the organization of chapters slightly confusing, wishing for decimal numbers (e.g., 6.2.4 for the fourth part of the second section of the sixth chapter) instead of the use of bold or italic headings to indicate structure. Since many of the chapters are in the range of 20–25 pages, such lack of clarity on structure makes the work hard to use for reference purposes.

For those with little background, chapter 2 introduces the Internet and WWW. It is well done, but in my course, I covered all but chapter 2 since my students had heard that before. Soon after, in chapter 4, on Innovation and Research, is an overview of current research in the field. It might make more sense to have this near the end of the book. Indeed, there are a

number of oddities in the organization. Metadata is split over chapters 10 and 12. Electronic publishing is split over chapters 9, 12, and 14.

In my courses, I cover the chapters in the following order: 1, 3, 9–12, 14, 4, 13, 8, and 5–7. I had to draw on other sources for my units on Multimedia and Representations and on Database Management. The strongest coverage is of Economic, Legal, and Social issues (in chapters 5–7). Next largest is the discussion in chapters 9–11 on text, information retrieval, and resource discovery. Since most people consider metadata to be one of the key topics in the field, it is not surprising that Arms discusses it in a number of places, but that makes it difficult to teach about. My solution was to cover it as the 8th of 9 units in my syllabus, referring back to portions of the book read as part of my earlier units on electronic publishing and retrieval. I had a similar though less severe problem with my unit on Architecture, which relates most closely to chapter 13 by Arms. Since this topic is one the author has explored extensively, there are related discussions scattered throughout the book, too. Finding them is left as an exercise to the reader!

Though I may be biased in my view of the important issues related to interfaces for digital libraries, I expect that most experts on human-computer interaction would find the coverage of that area, found in chapter 8, to be idiosyncratic. How can one cover this with only 3 figures that illustrate functional levels or where interface code is executed? Though Arms can explain with words what most others have trouble presenting even with the aid of numerous diagrams and screen dumps, in this case he seems limited by his approach, and so perhaps deals more with systems as opposed to user-oriented issues.

Some other groups may be upset with the coverage, though Arms does give fair warning to them in the Preface. There is a decided focus on a limited number of American universities. There is little discussion of content not dealt with in academic/scholarly environments. Maybe that is necessary, with so much of the cutting edge research led by university teams. However, as a digital library industry emerges, offering both software and services, it will be necessary to have new books written for the industrial sector.

Similarly, for the academic community, there is need for more coverage on applications. These are discussed along the way, but not treated as a separate topic. I particularly miss coverage on how digital libraries can support learning and teaching. Arms is now involved in National Science Foundation efforts to build a digital library to support undergraduate learning about science, mathematics, engineering and technology (see www.smete.org), so certainly there will be discussion of that in his future writings.

Nevertheless, in 2000, Arms' *Digital Libraries* is an important book for those interested in the field. It gives highlights of the knowledge and experience of one of the most active workers in the area. Read it once or twice so the many ideas sink in. You will remember the varied examples, and they will remind you of the important concepts. Then you can use digital libraries to discover the rest of the details.

Edward A. Fox

Professor, Department of Computer Science
Director, Digital Library Research Laboratory
Virginia Tech, Blacksburg, VA 24061 USA
E-mail: fox@vt.edu

Ink into Bits: A Web of Converging Media. Charles T. Meadow. Lanham, MD: Scarecrow Press Inc.; 1998; 304 pp. Price: \$24.50 (ISBN: 0-8108-3508-8.)

This book presents a broad historical perspective on the media and communication issues in the movement from print to electronic publishing. It is not intended to be a scholarly work, but, rather, as described by the publisher, it is an introductory text for "... students in courses on communication or technology in society, for students of library and information science, for librarians, for writers, and for book people of all kinds." The following quote from the book's preface gives a good summary of one of the author's central themes (p. xv).

Books have had a tremendous role in transmitting culture. Changing books too fast may have consequences we cannot foresee now. On the other hand, there was a world before books and if books as we know them today were to disappear, there will still be a world. Better? Worse? I don't know. *Neither does anyone else.* Make up your own mind.

Realizing that the current movement towards electronic media is changing very rapidly, the author does not try to focus on current details of that development, but rather places the present in a broad historical context and speculates, very generally, about the more distant future.

The book contains seventeen chapters, an appendix of statistical charts, and a section on recommended further reading. The opening chapters introduce the notions of changing media and a changing world; define media and information; and give an historical account of the evolution of media, including the origins and development of the book, telephone, radio and television, and computers and computer networks. These chapters are followed by a chapter describing the special place that books and writing hold in our culture. The next several chapters discuss: ways of representing information, such as through language and images; linear text and hypertext; interactive computing; and multimedia. Then come a series of chapters on: modern telecommunications, or the information highway; distribution of information; comprehension of information, when received in various formats; technology adoption; markets; and consumer protection. The book concludes with reflective chapters on change, the future, and general conclusions.

The author sees the gradual replacement of the book as we know it by electronic media as inevitable. Because books, and earlier print media have played such a central role in our culture for thousands of years, we tend to take them as a given (p. 198).

One of my main contentions is that most of the attributes of current books, as objects, have been with us for so long that we tend to forget that they are merely technological devices: they replaced some earlier form, and they were resisted upon initial development. They are not inherent in nature.

The author goes further with his speculations on the evolution of print and electronic media, and asks if even the electronic form of print might disappear (p. 235).

I cannot imagine anyone suddenly deciding to end print. What I can imagine is a *very* gradual increase in people's use of voice input and output and VR-like computer recognition

of gesture. And I can imagine teachers and school boards, *eventually* realizing that the effort to teach use of the written word is no longer worth it.

In discussing the pros and cons of print vs. electronic media the author notes that while electronic media still cannot be read as comfortably as hardcopy print media, electronic information access, using information retrieval and computer human interaction, is made much easier. While on the topic of information retrieval the author takes issue with Bill Gates, who said in one of his books, “The highway’s software platform will have to make it almost infallibly easy to find information, *even if users don’t know what they’re looking for*” (Gates et al. 1995) (emphasis added by the reviewed book’s author). The author’s point is that if the user does not know, the computer cannot either, but the Gate’s quote seems to reflect somewhat the same sentiment as that of Belkin et al.’s concept of a user’s anomalous state of knowledge, or ASK (Belkin et al. 1982a, b).

The author makes some points about the future of publishers and libraries. He observes that the people who made decisions in motion picture studios used to be the owners of the studios, but that now the decision makers were employees, which means that they are less likely to take risks which might threaten their own job security. The author suggests that the publishing industry, whether of books or of newer multimedia, may face the same sort of stagnation. He also observes that publishers will be less concerned with finding new content, than with thinking of new ways to repackaging existing content. As for libraries, he sees their position in our society as secure, even though their paper collections will likely disappear.

In conclusion, this book serves well the purpose and target audience for which it was created. It gives a thoughtful, balanced overview of the historical development of print and more recent electronic media and the gradual replacement of the former by the latter. This book provides a good introduction of these topics, suitable for students of communications and library and information science, as well as the general reader.

References

- Gates B, Myhrvold N and Rinearson P (1995). *The Road Ahead*. Viking Penguin, New York.
- Belkin NJ, Oddy RN and Brooks HM (1982a). ASK for information retrieval: Part 1. Background and theory. *Journal of Documentation*, 38(2):61–71.
- Belkin NJ, Oddy RN and Brooks HM (1982b). ASK for information retrieval: Part 2. Results of a design study. *Journal of Documentation*, 38(3):145–164.

Paul Thompson

Lead Research Scientist
West Group; D3-66G
Eagan, MN
E-mail: Paul.Thompson@westgroup.com