

Information Representation and Retrieval in the Digital Age (ASIST Monograph Series) by Heting Chu, Medford, NJ: Information Today; 2003. 248 p. ISBN 1-57387-172-9

© Springer Science + Business Media, Inc. 2006

I was looking forward to an up-to-date introduction to information retrieval accessible to a wide audience; unfortunately, this book is not it. It lacks the clear elaboration of the underlying conceptual structure that is essential for lucid presentation. It is full of oversimplifications, uncritical inclusion of oft-repeated misconceptions, and statements that are just wrong. It is in serious need of editing. A 200-page book on such a complex topic is almost by necessity doomed to be superficial.

Chapter 1 starts out with a historical overview. For Calvin Mooers it omits his most important contribution, indexing by concepts derived from user-oriented analysis (*descriptors* in the original sense of Mooers when he coined the term). This is followed by a confused elaboration of key concepts and a brief outline of major system components. To give a sample of the writing, “The **word** *digital*, as opposed to *analog*, is a relatively new **concept**” (bold added). Chapters 2 and 3 deal with information representation, basic approaches, and other related topics. Here we are treated to statements such as “In no case can information retrieval be performed with the original documents themselves.” It sounds though the author has never heard about full-text retrieval.

Chapter 4, *Language in Information Representation and Retrieval (IRR)*, gives a superficial and oversimplified description of types of IRR languages. When one reads statements such as “Neither training nor maintenance is required for doing IRR in natural language.” or “... because natural language should always be compatible with itself.”, one wonders why the author thought she could enlighten others on these issues. When one reads statements such as “In natural language, there are synonyms that use different terms referring to the same entity.” or “Homographs, also known as polysemy, seem to be another argument ...”, one wonders where the editor has been. Chapter 5, *Retrieval Techniques and Query Representation*, talks about specific techniques of formulating queries, while Chapter 6, *Retrieval Approaches*, gives the general context of query-based searching versus browsing. In Chapter 5 the author fails to make the obvious distinctions between techniques that deal with individual query elements, such as truncation or case sensitivity, and techniques that deal with the combination of elements (and their relation within a combination), such as proximity searching. Chapter 7, *Information Retrieval Models*, discusses the Boolean, vector space, and probabilistic models; it does not elevate the general level of discourse in the book.

Chapter 8, *Information Retrieval Systems*, discusses traditional online systems, CD-ROM systems, OPACS, and Internet retrieval systems. Internet retrieval systems and their features are discussed at some length; much of this discussion applies to all types of retrieval systems, and a broader discussion would have been more useful, pointing out a true convergence of features rather than the superficial convergence (such as accessing CD-ROMs through OPACs or online databases through the Web) discussed in Section 8.5. Chapter 9 *Retrieval of Information Unique in Content or Format* deals with multilingual information (in which the author mistakenly includes monolingual retrieval in languages other than English), multimedia information, and Hypertext and hypermedia information. The section on still image retrieval is not bad, except that the relationship between Panofsky's levels of interpretation and Eakin's levels of image attributes is poorly analyzed. Sound retrieval is a poor category since speech retrieval and music retrieval have quite different problems.

Chapter 10, *The User Dimension in Information Representation and Retrieval*, deals very briefly with users and user needs and the cognitive model – topics that should have been treated much earlier in the book – and more extensively with human-computer interaction in IRR systems. It is not made clear that the cognitive model is not an alternative to the information retrieval models discussed in Chapter 7 but rather deals with a different aspect of the total information retrieval problem. The following paragraph illustrates the muddled thinking and writing so prevalent in this book:

The cognitive model has many different implementations. For example, Belkin, Oddy, and Brooks (1982) designed an interactive IR system (e.g., question-answer) to find out the user's information need. The analysis of thinking-aloud protocols is another technique applied in understanding the cognitive activities of user when they search for information (Ingwersen, 1982).

Chapter 11, *Evaluation of Information Representation and Retrieval*, is the longest and best chapter in the book. It does give a fairly broad discussion of evaluation measures (but it would have been better to put this part, not just the brief preview in Section 5.2.2, early in the book so that various retrieval techniques could be discussed in terms of their plausible effects on these measures). Section 11.1 introduces standard measures of retrieval performance, including Swets' E measure, but strangely not the now—prevalent (for whatever reason) F measure. While the oft-repeated but nevertheless oversimplified assertion of a universal inverse relationship between recall and precision is repeated here as well, at least one voice of reason challenging that statement (Robert Fugman) is cited. Section 11.2 gives a broader discussion of evaluation criteria for IR systems, but unfortunately the discussion is balkanized according to the types of IR systems distinguished in Chapter 8 although most of the criteria apply across systems. Section 11.3 gives an overview of major IR evaluation projects with a focus on TREC.

Chapter 12 *Artificial Intelligence in Information Representation and Retrieval* is just seven pages long; it focuses on Natural Language Processing (NLP), mentioning its major components, and mentions a potpourri of topics under the heading *Intelligent agents*. The treatment is so brief that the reader can hardly gain appreciation and certainly not understanding.

In sum, the book lacks clarity, is often misleading if not just wrong, and is poorly written.

Dagobert Soergel

College of Information Studies

University of Maryland

College Park, MD 20742-4345

dsoergel@umd.edu