

BOOK REVIEW / CRITIQUE DE LIVRE

Information retrieval: a health and biomedical perspective. 3rd ed. By William Hersh. New York: Springer, 2009. 486 pages (hard cover). ISBN 978-0-38778-702-2.

In his preface, William Hersh states that his work strives to inform its readers about the fundamentals of information retrieval (IR) systems in health and biomedicine: a lofty goal indeed. In addition to this stated goal, the author also attempts to address this complex subject to the uninitiated, rather than to established health information professionals.

The book itself is divided into three sections: basic concepts, retrieval systems, and research and development. Within these sections are chapters (laid out much in the same way as the *Dummies* series) that encompass a myriad of topics. Section 1, for example, has chapters on definitions and models, as well as an overview of health and biomedical information. Section 2 discusses contemporary content in retrieval systems; it also relates how IR systems are modified into digital libraries. Finally, section 3 is an overview of research and development into this area, research and user evaluation, and supplemental topics such as text mining and categorization.

The initial section, like the rest of the book, takes a literature review approach to its content. Here, Hersh reviews the scholarship surrounding the terms, models, resources, and evaluations of the health information and IR environment. Take, for example, his discussion about the current state of IR databases. He stipulates that the hard-and-fast world of the physical library (i.e., books, stacks) is slowly being replaced by its digital counterpart. Furthermore, Hersh also reviews the scholarship surrounding the mass digitization movement and how it thoroughly affects libraries, copyright, business models, and publishers in the health information field.

One of the middle sections, section 4, reviews the scholarship relating to indexing and its associations with IR. In subsection 4.2, Hersh briefly summarizes how specific factors, such as arithmetic measures of consistency or specificity, influence both database use and its users. Subsection 4.4 reviews assorted academic commentaries on manual indexing; under bibliographic, full-text, and Web subsections, Hersh again relates how manual indexing (e.g., thesauri) has undergone a paradigm shift from manual to automated means. In the same general section of the book, the author also highlights changes to the MEDLINE system — specifically, that its indexing details, which were once found in a bound manual, are now distributed through Web-based means.

The latter part of the book, namely section 9, discusses ancillary issues and is entitled “Related Topics”. An example of said related topics is found in subsection 9.2: “Text Categorization”. Here, Hersh summarizes the scholarship of this specialized field via formulae and literature review. Some of the points he establishes are that this process can be used to filter e-mail spam, prepare biomedical documents for

database extraction, and identify the author’s gender, albeit controversially.

The work concludes in an unusual manner. The ultimate section (9.4), which imparts information and scholarship about an information distillation process called text summarization, concludes without review or concluding analyses of previous sections or subsections. Simply put, the monograph ends without any form of synopsis.

What, then, of Hersh’s argument or thesis? Ostensibly, as stated in the introductory paragraph of this review, Hersh designed his work to be an introductory overview, not a critical appraisal of health information IR scholarship. The absence of such an argumentative stance, however, is often lamentable as this reviewer was often left wondering what Hersh himself thought of the various controversies and issues as his discourse unfolded.

While it is true one can deduce what Hersh thinks about the various and sundry aspects of IR from his own cited papers, some lingering issues remain. One such issue noted by this reviewer concerns readability. Take, for example, one of the early sections on “Health and Biomedical Information”. In this early section, Hersh attempts to relate how information is perceived (and one can conclude retrieved) by different individuals. Although this declaration seems clear, the subsequent information he uses to further his review is cloudy. Later on, in his meticulous, but nevertheless dense, subsection on “Theories of Information”, the reader is provided with a fair number of mathematical equations that form the underpinnings of information theory. While some may consider the logarithmic equation in subsection 2.2 to be elementary, information specialists who are unfamiliar with such calculations may find themselves at a loss as to what, exactly, these quantitative relationships have to do with the day-to-day realities of IR.

Another one of Hersh’s subsections, this time dealing with indexing, also preoccupies readers with its precise, mathematician-like detail. As in the example mentioned above, Hersh relates yet another equation, this time relating to index consistency. Although it is understandable that the author wants to acquaint IR novices with foundational algebra, he, as he does infrequently in other sections of his work, overwhelms readers with complicated formulae.

To conclude, it is noted in the inner jacket of the book that Hersh is currently Chair of Medical Informatics & Clinical Epidemiology at the Oregon Health & Science University. Evidently, his work reflects this knowledge and attention to detail. As a result, this book, which claims it is accessible for the uninitiated, is for the more experienced professional. I would suggest that readers look elsewhere if they wish to gain a practical knowledge of biomedical information retrieval systems.

Jason William Clark

Canadian College of Dental Health – Trillium College
Burlington, ON L7R 4B7, Canada
E-mail: jason.clark@trilliumcollege.ca