

# Book Reviews

Edited by Michèle Hudon

Book Review Editor

Michele.Hudon@UMontreal.CA

ANDERSON, James D.: **Guidelines for Indexes and Related Information Retrieval Devices**. NISO Technical Report 2; NISO-TR02-1997. Bethesda, Maryland: NISO, 1997. 52 p. ISSN 1081-8006.

The intention of the book is to address all types of indexes, electronic and print, displayed and non-displayed, human and automatic: this is an ambitious undertaking in view of the complexity of such a task. Although the book is published by NISO Press, it wisely does not claim to constitute a standard. The foreword reveals that, after several years of efforts, none of the drafts elaborated by a committee of six experts and intended to replace ANSI Z39.4-1984 met with "sufficient agreement in the NISO membership". Therefore, the NISO committee dispensed with issuing a *standard*. Instead, the work is published as a *technical report*, the treatise of a recognized expert in the field with an emphasis on closed indexes, such as those for books.

Any standard with such a wide scope of coverage and in a field of so much ongoing experimentation would constitute (and has constituted) an obstacle to progress. Such a standard would be obsolete at the time of its appearance. In view of the great diversity of situations in which indexing has to be done, no standard can claim validity for each of the countless constellations of individual requirements and resources. It could only give advice that would be at least partly wrong or obsolete for an individual situation. It would expose everyone whose research and experience exceeds the standard's level of advancement to the reproach of ignoring existing standards. Standards make no sense in the scientific field.

That the 1984 standard has been withdrawn without having been replaced by another standard is, in the view of the reviewer, another beneficial outcome of the reasoning of the committee during its work. The decision of the NISO committee should constitute a warning signal to all those who are still working on index and indexing standards.

The abstract to this report states that it deals with the "principles of indexing regardless of the type of material indexed, the indexing method used (intellectual analysis, machine algorithm, or both), the medium of the index, or the method of representation for searching". One wonders how a report of 44 pages (double columns) can cover such a comprehensive topic in sufficient depth. The explanation is that the book restricts itself to mentioning *several principles* of indexing, is content with providing an overview, and directs the user to the enumeration of prominent books listed at the end of the book for details.

Some of the terms in the book receive uncommon usage, but in the comprehensive glossary of key terms their meaning is specified. For example, in this work, the term "role indicator" is used to mean not only an encoded, lexical (vocabulary) unit, agreed upon and contained in an indexing vocabulary to "identify the role of a topic...", but also (natural language) phrases and prepositions. The glossary is a valuable source of definitions.

In such a condensed text, many topics are only implied or had to be omitted, for example the problems caused by excessive precombination in the vocabulary. The discussion of the topic of syntax is limited to the type of syntax based on natural language (the use of which is unpredictable and which is therefore not suitable as a source of search statements in *mechanized* index search). However, the type of syntax inherent in a heading and its subheadings in a printed index is dealt with in detail. The capabilities and limitations of automatic indexing are not discussed, but the author states that what has been produced by automatic indexing so far resembles a concordance more than a genuine index. The problems incurred through the unhesitant admission of pre-combinations to the index language vocabulary remain undiscussed, as well as the problems caused by the nonlexical, paraphrasing presentation of concepts. The same holds for the various problems in informa-

tion system evaluation, and for the assessment of the survival power of an electronic index.

The book wisely advises against merely "controlled" indexing, (any term from the vocabulary is *permitted*) and recommends instead the use of the *most specific* term provided by the vocabulary. Vocabulary categorization is mentioned, and plenty of practical advice for book indexing is submitted. Software producers should recognize the desiderata raised for indexing software. The book also wisely excludes faceted formulae from "pre-coordination" (although they are also phrased *before* the search). Unfortunately, the glossary defines pre-coordination differently. (This represents the confusion prevailing in the literature. It is due to the lack of vocabulary categorization that would create clarity here).

The layout of the index is excellent, but access to the book's content can still be improved. For example, concordances were not indexed and a reference from compound terms to pre-coordinated headings would be desirable, as well as a reference from the latter ones to string indexing.

Through its conciseness and transparency, the book informs the novice in the field, the manager and the publisher of the complexity of the task of index preparation. Providing such an overview is the strength of the book. It destroys the still widely encountered opinion that indexing is a mechanical process that can be performed by clerical personal or by computer programs. Thus, the book, if widely distributed, can contribute much to a better recognition of index preparation as a work requiring indexing expertise and subject knowledge which deserves an adequate esteem by users and publishers.

Dr. Robert Fugmann

Dr. Robert Fugmann, Alte Poststraße 13, 65510 Idstein

MEISS, Brigitte: **Information Retrieval und Dokumentmanagement im Multimedia-Zeitalter**. Deutsche Gesellschaft für Dokumentation, Frankfurt/Main 1997 ISBN 3-925474-32-3, 118 pages, 8 figures.

The introduction to the book addresses the ever-continuing and increasing problems encountered in present-day information retrieval. The author suggests that competent indexing and skillful searching could counteract a lot of these problems.

With this goal in mind the conceptual foundations of information supply are discussed first. The importance and usefulness of classification schemes and thesauri are explained. Such devices are considered indispensable to subject searches and if proper names are not the exclusive search goals. The role of the degree of order to be attained in a collection through search

and desirable in the prevailing circumstances are unfolded, as well as the role of the definability of the search topic and the contributions of human memory, which also play an important part in the quality of the searches. The borderline between "relevance" and "pertinence" in the search results is clarified and thus much of the appertaining confusion is avoided which prevails in the literature.

The problems which originate from the variations in the orthography of names for concepts are explained. (But authors and searchers often choose the definition-like, paraphrasing and, hence, unpredictable mode of expression for concepts instead of their concise names, and the resulting problems also deserve being mentioned.) It is also argued in favor of writing one's own abstracts for the documents of potential interest.

The reader is informed of the limitations of the storage of the full texts of originals and abstracts, but also of the merits of full texts as complements to the indexing and classifying of the sources. The type of *noise* which inherently occurs in full text searches is discussed. But the information *loss*, from which this type of searches inherently also suffers, deserves mention, too, because it may well remain hidden for a long time in lack of experience and attention on the part of the user. We are also warned against simply storing full texts and trying to rely on tricks in the search for the text words.

Thus, the book sets out to counteract the ignorance which prevails on the part of many database users with respect to the conceptual devices for information supply, and it advises newcomers against beginning to collect knowledge on an insufficient basis.

The search strategies discussed in the book include the use of classifications and of thesauri (the latter with a well structured vocabulary) and the use of Boolean and proximity operators. In addition to the *noise* of irrelevant responses, the information *loss* deserves mentioning, which also inevitably occurs when these operators are employed in full text searches. The author states that the advance in information technology cannot substantially improve the quality of the searches if there is lack of a sound conceptual foundation. This precisely addresses the core problem of present-day information retrieval.

In a special chapter the properties of commercial mass media (where subject analysis is mostly completely dispensed with) are outlined. What is revealed here may well give the incentive to a reader to establish early a database of one's own for the knowledge in one's field of activity, and not to wait until any overview has been lost. Practical advice is given for the construction and maintenance of private databases if those of a merely ephemeral nature are to be avoided.