Book Reviews

LANCASTER, W.F.: Indexing and Abstracting in Theory and Practice. Urbana/Champaign,IL: University of Illinois Press 1991. 328 p., 98 ill. ISBN 0-87845-083-1.
The book originates in a series of lectures given by the author and is intended to serve as a text for the teaching of indexing and abstracting. It is divided into the two main parts "Theory and Description" and "Practice". It emphasizes what indexing and abstracting have in common, with the borderline between both even beginning to be blurred from the author's point of view (p.5). Vocabulary control as an omnipresent subtopic in any book of indexing is treated at several points, including a reference by the author to one of his earlier book publications (p.14) which, in the reviewer's opinion, dealt with it less successfully than this present work (cf. Int. Classif. 14(1987)No.3, p.164-166). Regrettably, the indexing of books has not been treated here either.
Based partially on the literature cited, Lancaster stresses that indexing and abstracting are activities which presuppose a thorough knowledge of the subject field concerned, as well as much experience and knowledge of the needs of users of indexes and abstracts. Much the same is said of the retrieval problem. Here, among others, M. Bates is cited (p.220), who regards this problem as "indeterminate" (ipso facto obviously causing it to be non-programmable, if the quality of meticulous intellectual work is to be reached). In this connection, Kuhlen's sentence: "Abstracting is an intellectual art and as such not transferrable to automatic procedures..." is also quoted (p.245). In an epoch which wishes to make everything mechanizable or machine-operable, such statements are not trivial and should help every serious reader to realistically assess the possibilities and limits of the algorithmization of abstracting and indexing.
In trying to describe the current *theory* of indexing and abstracting, Lancaster set himself a difficult task. All too glaring are the deficiencies of the traditional and contemporary theoretical approaches we find compiled here in fairly great detail. In their experiments on the basic questions of indexing the various researchers arrive at fundamentally different and even inexplicably conflicting results. Even conclusions readily recognizable as wrong by the experienced practicioner can be quoted from the literature, such as the statement that "those searches in which both controlled terms and natural language were used performed less well on both recall and precision than the searches involving natural language alone" (p.207).
As a model example of fruitless and conflicting theorizing Lancaster cites literature on the discussion, still going on, on the value of indexing consistency as a quality criterion for any form of indexing. However: "Quality and consistency are not the same" (p. 81). The quoted work by Cooper from 1969 alone (an indexing method

can be "consistently bad instead of consistently good") should have sufficed to clarify the situation. With regard to the allegedly lawful inverse precision recall relationship derived from the Cranfield experiments and conjured up time and time again Lancaster likewise duly displays the reserve called for. On page 4 we read: "... that an improvment in recall will *usually* (my italics) cause a deterioration in precision and vice versa" (with a lawful relation being called into question). Much fuzziness also prevails in the discussion by the widespread failure to distinguish between "pertinence" and "relevance" (p.3), which Lancaster himself does not always do, even though in the glossary a clear-cut difference is made between both concepts (p. 291, 292).
Varying with the experimental conditions and accompanying circumstances chosen, the literature quoted sometimes shows the indexing procedure using a pre-established vocabulary as the superior one, and sometimes the uncontrolled storage of natural language expressions. While recognizing, with numerous other authors, that only the *combination* of both can yield an optimal result, Lancaster's assessment of the specific strengths and weaknesses of either approaches is contradictory at times. On the one hand, the length of the record is considered a factor favoring recall (which should make free, full text storage superior), on the other hand, in free text "the diversity of the way concepts are represented, however, makes it very difficult to achieve high recall in broad "conceptual" searches" (p.202, exhibit 91). It is also not clearly appreciated, for example, that the *paraphrasing* manner of expressing concepts cannot be recognized at all through free, extraction indexing. Also, uncontrolled natural-language indexing is by no means usually superior - as to specificity - to controlled vocabulary indexing (e.g. p.198, 202), but only there where the natural language mode of expression is at the same time also predictable.
Regarding the question, so important in practice, what measure of pre-coordination is advisable and still justifiable in the various types of indexes, one would have wished for somewhat more detailedness on pre-coordinate indexes in chapter 4.
Lancaster has some trouble in rejecting some evidently wrong, though still widespread assumptions, having cooperated himself, in part, in their elaboration within the framework of the ASLIB Cranfield experiments. This is true e.g. for his bias in favor of uncontrolled natural-language storage, as evidenced e.g. by the somewhat onesided literature selection on this matter and by his concluding forecasts, as well as by his noticeable agreement with statements asserting the equivalence or superiority of purely natural-language storage (e.g. p.204). On the one hand: "Full text should give greater recall..." (p.213) and "... a really comprehensive search on a topic ...will favor full text..." (p.201) and: "The full text of a document is likely to provide considerable redundancy, increasing the chance that it will include an expression used by a searcher and thus improving recall" (p.199). On the other hand it is recognized that ".. this very redundan-

cy creates great problems when the text of many documents is combined to form a large database - there are many ways in which a topic can be expressed in full text and, in some cases, the topic is represented implicitely rather explicitely (O'Connor, 1965), making high recall difficult to achieve" (p.200). Also, he perceives great possibilities for the improvement of natural-language systems (p.218). That such progress possibilities exist likewise for the systems with controlled vocabularies is not taken into consideration. His argumentation is not always free of polemics here.

The book also shows clues to the solution of inconsistencies not yet clarified by contemporary theory. This is the case e.g. where attention is called to the necessity of predictability of the representation of concepts (p.196) and where the problems caused by the multiplicity of the paraphrasing manner of expression for a concept are briefly touched upon (p.200). For example, the table on p.217 with its listing of the great variety of imaginable natural-language manner of expression of a concept proves to be glaringly incomplete. In the view of the reviewer, from all this the hopelessness of any attempt to algorithmically recognize paraphrasing expressions of concepts and to machine-translate them into subject headings should be evident when more than just modest requirements are to be met.

In concordance with contemporary terminoloy, for Lancaster, too, an indexing language is still practically identical with its *vocabulary* (except the brief remarks on p.171 ff), with the desirable *grammar* largely remaining out of consideration, an omission which constitutes the source of many problems.

As to the *practice* of indexing and abstracting, this is, by the nature of things, discussed at various points in the theoretical part as well, and the abundant material compiled will prove most useful to the reader. Several variants of printed indexes and types of abstracts of varying detailedness are presented in informative examples, together with illustrative remarks on the technique of their elaboration. In particular, one learns several interesting and instructive things about the construction of databases, also in the early and pioneer stages of this art.

Lancaster rightly criticizes the fact that the value of the empirical material collected so far often suffers from its having been elaborated on too small a scale (p.213). The results collected under such circumstances do not permit conclusions for the practice of larger information systems. This is something which all those experimenting with small-scale systems should take to heart, and the lack of probative value of the results from such small-scale experiments is by no means compensated for by the sheer quantity of these experiments in the literature, which they dominate (the book reviewed here included), for they can be rapidly and cheaply conducted (as e.g. the experiment described in Tenopir's dissertation, amazingly termed exemplary on p.212).

Surprisingly, all these observations find only little reflection in the forecasts at the end of the book (p.252ff). Here, among other things, great confidence is expressed that the need for indexing and abstracting will disappear once the sources are available in machine-readable form and a (non-specified) kind of "filters" will be created (p.259). In the view of the reviewer, this confidence is not justified by the experience reported in the serious literature Lancaster quotes.

Illusory, too, are the expectations placed in the idea of post-controlled vocabulary (p.215), in which practically uncontrolled storage of natural-language text words is permitted and in which the different forms of expression that have occurred for a concept are to be compiled only afterwards so that they can be given attention in formulating the query. The reviewer has experience with such an approach and can only advise against it. For, the disambiguation of natural-language expressions after their storage proves hardly realizable, requiring as it does the renewed perusal of practically every original document. Also, the filling-up of the ellipses and the omission of clearly dispensible material from the index, tasks traditionally performed by the indexer, are dispensed with here. Furthermore, those concepts which occur in the paraphrasing mode of expression would wholly escape translation into subject headings and inclusion in the index.

The theory deficit presently prevailing in information science reminds the reviewer of the epoch of Galileo shortly after the Middle ages. Depending on the experimental conditions chosen, experiments of some sort seemed to prove that the velocity of freely falling bodies *varies with their weight* (in accordance with the opinion then held), while experiments of another nature seemed to *prove the contrary* (Galileo's revolutionary idea). The vacuum state, which could have cleared up all contradictions, had not yet been discovered then. Today's information scientific community is still awaiting a breakthrough to a similarly fundamental insight.

Thus, the chapters on indexing and abstracting theory present a faithful image of the currently still unsatisfactory state of our theory.

For the practitioner, the book, owing to the vast amount of material collected, will prove quite helpful for an expedient organization of his daily routines. For the same reason, the attentive reader already familiar with the field can draw from the book many a stimulation for the reformulation of several theoretical statements.

The book has been very carefully designed and lucidly organized and is free of printing errors. The index is a good guide to the book's contents. The reviewer recommends the book as reading material for the students attending his classes on indexing and abstracting.

Robert Fugmann

Dr. R. Fugmann, Alte Poststr. 13, D-6270 Idstein

WILLIAMSON, Nancy J., HUDON, Michèle (Eds.): **Classification Research for Knowledge Representation and Organization.** Proceedings of the 5th International Study Conference on Classification Research, Toronto, Canada, June 24-28, 1991. Amsterdam, NL,