Indexes for Analysis and Diagnosis

Richard Joseph Hyman

Professor of Library Science and Director of the Graduate School of Library and Information Studies, Queens College of the City University of New York.

 Though indexing is more art than science, and textbook explanations of how and what to index are inconsistent or inadequate, indexes are essential for information retrieval. They can also be used for bibliometric analysis: citation counting; historical tracing of indexed concepts; correlation of terms in indexes to books on the same subject; evaluation of an individual book’s subject coverage. Diagnostic use of a preliminary index, ideally to be prepared by a book’s author, is not so familiar. Such an interim draft index can be a powerful tool in revising the structure and logic of a manuscript before it is set in type.

That indexing is more an art than a science is hardly news to indexers—the older the indexers, the older the news. Yet unequivocal definitions of an index are not hard to find. The 1943 ALA glossary of library terms still seems valid: ‘A list of topics, names, etc., treated in a book or a group of books, with references to pages where they occur’. (This definition applies also to indexes of serials.) A more abstract definition can be equally comprehensible: ‘An array of symbols (usually words) systematically arranged (usually alphabetical) together with a reference from the symbol to the physical location of the item symbolized’.

However, for so definable a product, explanations of its creation are surprisingly elusive. Textbooks do not agree even on specific mechanical procedures, e.g., whether See also references should come at the beginning or end of an entry. Even less consistent are their explanations not merely of how but of what to index. Neophytes are advised to index everything, but only everything crucial, significant or pertinent. These adjectives are left undefined, though used repeatedly, as in the University of Chicago Press A manual of style:

The subject matter and purpose of the book determine which statements are pertinent and which peripheral … The ideal indexer … must make certain that every pertinent statement in the book has been recorded in the index in such a way that the reader will be able to find without difficulty the information he seeks … An entry should be made for every proper name about which a pertinent statement is made.

Experienced indexers will have developed their own concepts of ‘pertinent’ as distinguished from ‘peripheral’. Beginners will seek authoritative guidance, only to be told that any explanation must depend upon the context, and to be warned that over-indexing is a cardinal sin. In lieu of defining ‘pertinent’ etc., standard indexing texts supply numerous examples of over-indexing. Wheeler warns against ‘the folly of increasing the size and expense of the work by useless entries’ with entries from the index to St George Mivart’s On the origin of human reason, ‘a notable example of absurd repetition and unwise choice of keywords’.

A manual of style cites instances of when not to index proper names, e.g., the first sentence of John M. Rosenfield’s Dynastic arts of the Kushans, which is categorized as ‘Fine scene setting, impressive cast of characters—but the indexer should simply pass it by without making a single entry’. Though the examples may not define, admittedly they may be useful to the inexperienced. About all the examples, however, hovers a suspicion of after-the-fact reasoning.

What to include

Authorities on indexing in the ‘hard sciences’ may be strict, even doctrinaire, on what should be indexed, though much subject knowledge would seem necessary for compliance—assuming that the specialist indexers agreed with the dicta. Thus, Borko and Bernier distinguish among subjects, concepts, topics and words. They prescribe that only subjects need all be indexed, because they are ‘the central themes toward which the attention and efforts of the author have been directed. They are the aspects of a work that contain novel ideas, explanations, or interpretations’. All other items are also-rans, at least in the scientific and technological ‘subject’ indexes which Borko and Bernier have in mind:

Users of subject indexes do not waste time consulting concepts, topics, and words in which they have no interest. This does not mean that indexes to concepts, topics and words are of no value; it simply means that such indexes are different in function from subject indexes.

Most of the technical, as distinguished from these ideological, problems of indexing derive from traditional descriptive cataloguing and subject analysis, with their numerous optional or alternative or ‘it all depends’ choices. Descriptive cataloguing and subject analysis, which in a broad sense may also be regarded as indexing, resemble indexing in the narrower sense of being more art than science. Indexers, though, without a thorough grounding in these traditions, will always be in danger of
reinventing the wheel or designing a square one. The most technologically advanced indexing techniques rest on foundations laid by old-fashioned cataloguers and subject analysts. Cutter and Dewey faced the same problems as today’s indexers of monographs and serials. Issues of choice and form of entry, of multiple and double entry, of syndetic structure—all have been wrestled with by predecessor librarians.

These specific technical problems of indexing are neither negligible nor often easy to resolve. Some perhaps never will be—not an auspicious omen for regional, national or international standardization! Many of these problems, e.g., the above-mentioned controversy on the placement of *see also* references, could be solved if we did not have to choose: we could place the references both before and after the related headings. Unfortunately, economic constraints always intrude, so solutions usually become compromises. Sometimes the Gordian knot can be cut. It is common sense that the index to a microfilm run of a journal or newspaper is more helpful at the beginning than at the end of the reel. Reversing the printed custom solves the problem. (Ideally, the index should be repeated throughout the reel.) Computerized search capabilities, i.e., free-text and full-text, promise simplified solutions to many such difficulties.

**The Index as Guide**

Other technical problems also have to be resolved. For example, in 1978 Borko and Bernier described index entries as ‘guides rather than surrogates . . . designed to guide searchers to, rather than inform them about the contents of a work’. Wheeler, who had published her *Indexing: principles, rules and examples* in 1905 and left little to be changed in later editions, had similarly cautioned: ‘The indexer should bear in mind always that the index entry is only a “pointer” showing where certain information is to be found’. Immediately thereafter, she mitigates her stern decree:

If in a work the gist of the information can be indicated, sometimes saving the cataloger the trouble of turning to the text, it may well be given, but there should be no attempt to make the index entry a digest.

How does one heed Wheeler’s suggestion to indicate the ‘gist’ but without offering some type of ‘digest’? How can a selector select from the many appearances of a personal name unless the index gives their contexts? If a context can be given in a brief phrase, should the indexer—and this is doubtless what troubled Wheeler—conceal the context by *not* providing what in effect would be a ‘digest’? Fortunately for the reader, practicality will often outweigh purism. Perhaps the *New York Times index* is no index according to the definition of Borko, Bernier or Wheeler, but it keeps its title through priding itself that besides being ‘a carefully made index . . . the brief synopses of articles answer some questions without need to refer to the paper itself’.

Hyman has suggested one approach to analytical access to information in monographs: requiring an informative abstract in every nonfiction monograph and its bibliographic record. The abstract would be indexable for manual or computerized access through controlled vocabulary and/or free-text searching. It must be acknowledged, though, that preparing a reasonably short yet comprehensive and informative abstract of a monograph is, like indexing, more art than science. *Precis* advocates are not far from claiming such an achievement, even though *Precis* purports to be an indexing, not an abstracting, system. In any case, the concept of a comprehensive informative abstract as a back-up for a detailed index deserves a trial. Such a feature could be built into the new register-index format for the *National union catalog*.

Despite all these strategic and tactical problems, macro- and micro-, some of which appear insoluble, serial and monograph indexing continue. No professionally responsible author or publisher would offer a nonfiction book without an index.

**Some Uses of an Index**

Once an index is completed—assuming most of the problems have been dealt with adequately—it can be used for bibliometric purposes. Historical analysis may be applied to the occurrences of indexed terms and concepts. Pads can be traced by the frequency or disappearance in indexes of serials, e.g., the varying popularity of ‘browsing rooms’ in academic libraries. We recognize the analytical potential of citation indexes, in which the number of times a work is cited may be a measure of its importance.

Bibliometric analysis of book indexes can yield insights into the nature of a new or changing discipline. (Also, practitioners are aware that examination of the index is frequently needed when classifying or subject-cataloguing a monograph). Collating subjects or concepts—allowing for synonymy—in the indexes to books which are ‘introductions to information science’ can identify the consensus or disagreement of the experts on the legitimate components of that still mysterious discipline. Although the merging in 1976 by Kilgour of the indexes to ninety-six books in library and information science was unsuccessful as a retrieval device, an analysis of the terms in that compilation can reveal the definition or self-definition of the field at the time.

The idea of assembling back-of-book indexes for discrete subject areas was revived by Superindex, Inc., whose data base was accepted in 1982 for online access by *DIALOG Information Services*. The base consists of the indexes to about one thousand science, engineering, and medical reference books from nine publishers. *DIALOG looks to the day when “it will be standard procedure to include the index of every new reference book in the Superindex program”*. This computerized index

*The Indexer* Vol. 13 No. 3 April 1983
compilation already promises extraordinary analytical possibilities.

Readers and reviewers may study the index of a book to see if the 'great names' and latest terms have been included: an evaluative or litmus-test use. Authors hope that critics will also peruse the text—and frequently complain about this not having been done.

Author as indexer

Indexing and the study of indexes as a diagnostic procedure is not so familiar. 'Diagnostic' and 'diagnosis' will be defined later. We must begin with a consideration of the classic controversy: Is the best indexer the author, or a professional indexer? The pros and cons are well known, and have been summarized again in the University of Chicago Press A manual of style: the subjective [sic] involvement of the author producing a lack of perspective and scale versus the detachment and objectivity of the outside expert with the knowledge of what and how. The ideal is, naturally, a combination of the best qualities of both:

'Invariably, the best scholarly indexes are made by authors who have the ability to be objective about their work, who understand what a good index is, and who have mastered the mechanics of the indexing craft.'

The ideal, we all know, is an extremely rare bird. Usually, the author cedes the responsibility to a freelance indexer whose services are obtained through the publisher or by the author directly. The rule in publishing is that the author be responsible for supplying the index, not for compiling it.

Yet there are cogent reasons for the author to attempt at least a preliminary index, before the book is set in type. McColvin gives four purposes of an index, to which Borko and Bernier add three. None of the seven describes the diagnostic value of a preliminary index for the writer of the book.

Diagnosis presumes possible treatment. As the author, probably an indexing neophyte, indexes the manuscript, a diagnosis emerges: faulty structure of the work as a whole, over-emphasized unimportant topics, unnecessary synonyms, details in inappropriate places or better combined in a general treatment. Wide scattering of topics, self-indulgent or far-fetched allusions, illogical sequence of ideas—all can be revealed by a diagnostic preliminary index, even if not professionally done. Indeed, the unsolved macro- and microproblems of indexing can enhance the value of such a preliminary index; nothing will sharpen the skill of a conscientious beginner indexer—and revising author—more than struggling with the issues reflected by these problems.

A professionally prepared 'honest' index would unearth much of the same, but by then it is too late to revise the text. Professional indexers know from painful experience that publishers' deadlines usually preclude meticulous index preparation and any but the most inconsequential changes of text. Add to these the probability of index reference errors as pagination is changed from that on the indexed proofs. Of course, infallible authors, if there be such, may cease reading here.

The interim index

Optimally, an interim index is compiled by the author since it is for his or her benefit. If necessary, however, it can be done by a professional (though few authors could afford the double expense). The interim index is a powerful tool for revising the manuscript. Often the author works from an outline or prospectus, which can also be tested against the draft index. Authors expect to revise a draft; the draft index exposes necessary revisions in the text and inevitably in the index, as an X-ray reveals hidden pathological conditions. Indexing by its nature reorganizes in outline the full text, and thus supplies a workable miniature model for analysis and correction.

Though desirable, is such an approach feasible? It can be if the author allows for the preparation of a draft index within the schedule for submitting the completed manuscript. The time allotted would depend, of course, on the length of the manuscript and the experience of the indexer. However, the deadline pressure would be much less than for the final index. Also, this early index could shorten the time for compilation of the final, either by author or professional indexer.

Indexing dissects the contents and identifies major topics and relationships as well as their logical and physical positioning. Surely this kind of analysis should be done by the author in composing any text. In fact, authors do much of this quite naturally during their writing, but are seldom certain they have entirely succeeded. A careful reading of the text by a skilled editor or by an 'objective' author could reveal the same problems. As a practical matter, however, authors become so involved in composition and scissors-and-paste-work that they tend to overlook the forest for the trees. Preliminary indexing can lead to a more precise, better balanced and more logical presentation of ideas. It can almost automatically reveal structural flaws and save the time of a skilled editor.

Some examples from this writer's latest book may clarify the purposes and results of such diagnostic indexing. The first index revealed, not unexpectedly, problems of synonymy in quoted texts—the easy solution being a see reference in the final index—but it also revealed that the author, striving for elegance, was complicating the task by his synonyms. Deathless prose was not endangered by reducing these synonyms to one standard form. This also made very clear that the author was not referring to different topics.

The diagnostic index also showed that some paragraphs reappeared verbatim although their sense was still
appropriate. Accordingly, paraphrase replaced repetition. Since most authors reread their manuscripts, why not prepare a preliminary index while doing so? This would make undeniably obvious the sequence and repetition of ideas.

Treatment after diagnosis

An extensive structural change suggested by the preliminary indexing was to recast at least two chapters. Although the basic plan was to treat the same problems in chapters devoted to different types of libraries, the index showed that in two types of libraries the same pattern was followed in respect of one particular problem. The author decided to discuss the problem for both types of libraries in one chapter and in a subsequent chapter to refer back to the previous. Thus, the diagnosis and treatment reduced rigidity in structure.

Perhaps authors unconsciously avoid such an analytical task from a painful anticipation of the defects to be exposed. Many authors, too, share the mistaken concept of the index as an unimportant mechanical appendage to the vital intellectual burden of the text. How many authors ever read the indexes to their books? Authors may feel also that the skills required for indexing are unrelated to those for writing, much as humanities students assume that mathematics calls for aptitudes alien to them. But any author not content with the first version of a test—and how many authors are?—must exhibit at least a mild form of split personality: writing the draft in the heat of inspiration only to return to it, after a cooling-off, to evaluate and alter, almost like a stranger: ‘Did I really write that?’ Few authors agonize over the disparity of these roles to the extent of becoming acute neurotics.

In the millennium, all authors will have learned to index, either through formal classes—a good opportunity for library schools—or self-instruction, each combined with unremitting practice. Meanwhile, a preliminary index by the author or by an outside specialist is a potent diagnostic instrument in preparing a text for its final index.

References

6. Ibid.
7. Ibid., p. 11.
9. Ibid.
15. Borko and Bernier, p. 4.

Food for thought

The knitted woolly 5" x 3" index award for the most egregious index entry spotted to date goes to the following in Equality by Keith Joseph and Jonathan Sumption (Murray, 1979, £4.95):

Menus:
- apples, 83
- bread, 49
- cakes, 12, 50, 84-88, 105, 106, 112
- carrots, 24
- manna, 86
- omelettes, 44
- salt, 112

The book treats of economics and politics. The above references are to: the principle of distribution, as for instance of a thousand apples among a hundred people; whether lack of bread implies lack of freedom; ‘a cake-sharing mentality’, fixed and divisible cakes, ‘Cakes: their use and abuse’ (i.e. sharing out of), ‘What goes into the cake?’, ‘the value of the national cake’; the income tax rating analogy of ‘donkeys in children’s cartoons trotting forever onwards towards the carrot suspended from a stick attached to their neck’; the fruits of the productive process ‘ likened to the distribution of heavenly manna’; Beatrice Webb quoted as remarking of trucks of political prisoners, ‘you can’t make an omelette without breaking eggs’; and, finally, the page reference for the entry ‘salt’ leads one, incredulous, to, ‘Some egalitarians have now begun to concede that figures compiled on that basis should be taken with a pinch of . . . ’ yes.

A variegated and abstract menu indeed. Can it possibly be intended as an indexer’s joke?

I for my part venerate the inventor of Indexes; and I know not to whom to yield the preference, either to Hippocrates, who was the great anatomiser of the human body, or to that unknown labourer in literature who first laid open the nerves and arteries of a book.

—Isaac D'Israeli, Literary Miscellanies (1801)