Indexing Books--
Small and Large Loopholes

June 26, 1974   Number 26

The success of *Current* Book *Contents* has stirred up a number of librarians and scientists who bemoan the artificial separation of books from journals. This is especially true in the *Science Citation Index®* *(SCI®)*.

In 1974 the *SCI* staff will process approximately 5,000,000 reference citations. To prepare the *SCI*, every reference given in about 500,000 articles will be keyed. About, 80% of those references cite journal articles. About 20% cite books or other 'non-journal' items that are not clearly identifiable as references to journals.

For many years we used to identify these 'non-journal' citations with a special code, but we discontinued the coding because of its high cost, inconsistency, and its apparent lack of usefulness. Few users have ever commented upon our dropping the code. Nevertheless, it is now important to review the decision because of the advent of the *Social Sciences Citation Index™*. More than 50% of the references picked up from social science journals cite books.

More often than you would imagine, it is difficult to distinguish between a reference that cites a book and one that cites a journal. It is all the more difficult when it has to be done at breakneck speed while keying for input to the ISI data bank. Not infrequently the difficulty is compounded by a lack of generally accepted citation standards or a flouting of what standards may exist.

One may ask why all the concern. The user of *SCI & SSCI™* does not, presumably, need to be told that the citation he or she is consulting is a 'book'. However, one may have come across it, when looking up a journal article by the same author.

In addition, as we have previously pointed out, the *SCI* has become a tool for citation verification.[1] Thus many librarians would like to use the *Citation Index* section of *SCI* to check on the accuracy of information about a book that some client wishes to cite. One wonders why this should be the case when the client presumably consulted the book he wants to cite. But we all can recall citing a book or article from memory. The librarian is often asked to corroborate our memory. Unlike journal citations, book citations in *SCI* can often be less than adequate. This was brought home to me when we

82

recently reviewed our conventions for book citations in a study of citation counts for highly-cited 'classics'[2]. Discrepancies between information from our magnetic tapes and the printed *SCI* appeared.

A single book or article may be cited a dozen or more times by the same citing article. In the case of books, if the author of the citing article does not cite a page number, all of the repeated citations are essentially duplicates. The data-entry operator does not usually remember this repetition as the article is processed, and creates dozens of duplicate entries. During the final printout of the *SCI* the computer detects these duplicates and "unifies" them into a single line of information. However, the individual references remain in the data bank tapes. They may cause confusion in any statistical analysis which uses the printed *SCI* for verification.

If the citing author does supply a page number, we have another problem, and it must be carefully considered. If a dozen different pages of the same book have been cited, they are regarded as twelve separate citations. When the information is printed in the *SCI*, twelve separate *lines* will appear in the index. As far as the computer is concerned the book title could be a journal title. And therefore, without some other clue, it is not possible to unify what the human eye or brain knows is a book title. If you feel suddenly superior to our computer algorithms, consider the case when a book title is identical to a journal title. Today publishers relish the use of journal titles that do not clearly identify them as journals.

If I allow myself the luxury of unlimited central-processor time (not really possible when handling such enormous files) I could deal with the problem in the following way.

Given two entries under the same cited author and the same cited year, do the two contain the same cited titles, as for example, *Prostaglandins*. Though they may, they may still represent citations of a book or a journal. One feature of *most* journal citations is that they contain a volume number. One might distinguish on that basis. Unfortunately, such heavily cited journals as *J. Chem. Soc. London* do not. Thus, if we encounter a blank in the volume field we may reasonably take the time to examine a list of journals that do not use volume numbers.

Even if we should be successful, one way or another, in identifying all such citations as citations of a particular book, the question arises as to whether we should give the cited *page*. If the book in question is cited only a few times a year, the matter is trivial; the reader can examine the titles of the citing articles in the *Source Index*. But if it is a frequently cited book such as Pauling's *Nature of the Chemical Bond* or Freud's *Collected Works,* don't we do the searcher a disservice by eliminating these clues. If we eliminate such specific page citations, we would save an enormous amount of space in the *SCI*. Each cited book would have a single entry. Though

such citations now appear together in the *SCI*, the presence of the page number causes them to print out as different, although contiguous cited items.

Consider Pauling's book, which is cited hundreds of times every year. I have always assumed that is would be useful for the user of the *SCI* to enter the *Citation Index* from a specific page or chapter of such a work. The user's interest may be quite specific. That specific subject may be discussed in certain chapters or on specific pages. Such a user will *not* want to find all references to the book, but *only* those related to a specific chapter or page. Some members of our Editorial Advisory Boards think it is the rare user who will start a search with a specific page number. If they are right, our present process of including cited page numbers is wasteful. It uses extra space, adds to printing and other costs, and separates entries that otherwise would be arranged in alphabetical order by citing author--with all duplicates eliminated.

Obviously a book that is cited only a few times a year presents no such problem, even in a five-year cumulation. But a heavily-cited book, and there are many of them, if cited 20 times a year, will appear 100 times in a five-year cumulation. Will the user be able to distinguish the subject interest of these citing works without some indication of page or chapter? Will the title of the citing article or the name of the citing journal tell him what he really wants to know? Would he, if we ignored cited pages and listed authors alphabetically, know enough about workers in his field of interest to pick them from a straight alphabetic list?

How often have you used the page of a book as a starting point in an *SCI* search? Let me point out that the problem is much less severe when we are dealing with multi-authored books, with chapters or sections by different authors. Most people cite the chapter author first, then the book title, pagination, and then (perhaps) the editor. We *assume* that searchers will begin with the name of the chapter author. In any case there usually is no volume number. The page number is usually known by anyone who is looking for papers that cite that particular chapter. If we dropped the page number, librarians and other reference workers would lose the information which they frequently find useful in verifying reference sources for customers. Undoubtedly, on occasion, the omitted editor's name is a slight loss. For example, I have always assumed that referring to a book by Merton would turn up all citations. It turns out that the chapter authors have also been cited. Their names must also be added to my "Merton" *ASCA*® profile

Let me point out that the omission of page numbers from most references to books is so flagrant that one wonders why we should have attempted to deal with the matter at all. Perhaps you will now understand why I sometimes feel like King Solomon. As we say, out of the frying pan into the fire.

I suspect the compromise will be to drop page numbers for single au-

thored works and retain page numbers in the case of multi-authored volumes. The problem is how to identify them in the first place. If we stick with our present policy it will not be the first time we have done so after considerable soul searching.

All of the above leaves open a very important question. While the *SCI* has done much to integrate the book literature with the journal literature it has been an *indirect* assault on a major problem in bibliographic control. Every librarian and many scientists realize that books, especially multi-authored works of all kinds, are not adequately indexed. What needs to be done now is to take a big leap forward by processing such works as source materials—in the same way we handle journal articles. Until quite recently I was hopeful this could be done in the context of the existing *SCI*. But the continued growth of the scientific literature, in size and scope of interests, makes this economically difficult. It is a major problem to keep ahead of inflation in labor and paper costs.

Consequently I can foresee in the near future that ISI may embark upon a separate *SCI* Book Supplement which would handle all sorts of non-journals—monographs, conference proceedings, etc. Librarians will not find this separation difficult to accept since it is traditional in the library world. This separation would of course be necessary for printed indexes. But is should be remembered that the growth of our on-line services means that the book and journal material will be treated as one. The same could eventually be true for our tape and *ASCA* users.

1 Garfield, E. Precise bibliographical verification with the *Science Citation Index*. *Current Contents*® No. 35, 2 September 1970, p. 4-5.

2 .... A core research library for developing graduate schools--the 100 books most-cited by researchers. *CC*® No. 1. 2 January 1974, p. 5-9.