

REPRINTS AND REFLECTIONS

Citation indexes for science. A new dimension in documentation through association of ideas[†]

Eugene Garfield, Ph.D.

'The uncritical citation of disputed data by a writer, whether it be deliberate or not, is a serious matter. Of course, knowingly propagandizing unsubstantiated claims is particularly abhorrent, but just as many naive students may be swayed by unfounded assertions presented by a writer who is unaware of the criticisms. Buried in scholarly journals, critical notes are increasingly likely to be overlooked with the passage of time, while the studies to which they pertain, having been reported more widely, are apt to be rediscovered.'¹

In this paper, I propose a bibliographic system for science literature that can eliminate the uncritical citation of fraudulent, incomplete, or obsolete data by making it possible for the conscientious scholar to be aware of criticisms of earlier papers. It is too much to expect a research worker to spend an inordinate amount of time searching for the bibliographic descendants of antecedent papers. It would not be excessive to demand that the thorough scholar check all papers that have cited or criticized such papers, if they could be located quickly. The citation index makes this check practicable. Even if there was no other use for a citation index than that of minimizing the citation of poor data, the index would be well worth the effort required to compile it.

This paper considers the possible utility of a citation index that offers a new approach to subject control of the literature of science. By virtue of its different construction, it tends to bring together material that would never be collated by the usual subject indexing. It is best described as an association-of-ideas index, and it gives the reader as much leeway as he requires. Suggestiveness through association-of-ideas is offered by conventional subject indexes but only within the limits of a particular subject heading.

If one considers the book as the macro unit of thought and the periodical article the micro unit of thought, then the citation index in some respects deals in the submicro or molecular unit of thought. It is here that most indexes are inadequate, because the scientist is quite often concerned with a particular idea rather than with a complete concept. 'Thought' indexes can be extremely useful if they are properly conceived and developed.

In the literature-searching process, indexes play only a small, although significant, part. Those who seek comprehensive

indexes to the literature of science fail to point out that such indexes, although they may be desirable, will provide only a better *starting point* than the one provided in the selective indexes at present available. One of the basic difficulties is to build subject indexes that can anticipate the infinite number of possible approaches the scientist may require. Proponents of classified indexes may suggest that classification is the solution to this problem, but this is by no means the case. Classified indexes are also dependent upon a subject analysis of individual articles and, at best, offer us better consistency of indexing rather than greater specificity or multiplicity in the subject approach. Similarly, terminology is important, but even an ideal standardization of terminology and nomenclature will not solve the problem of subject analysis.

What seems to be needed, then, in addition to better and more comprehensive indexes, alphabetical and classified, are new types of bibliographic tools that can help to span the gap between the subject approach of those who create documents—that is, authors—and the subject approach of the scientist who seeks information.

Since 1873 the legal profession has been provided with an invaluable research tool known as *Shepard's Citations*, published by Shepard's Citations, Inc., Colorado Springs, Colo.² A citation index is published for court cases in the 48 states as well as for cases in Federal courts. Briefly, the Shepard citation system is a listing of individual American court cases, each case being followed by a complete history, written in a simple code. Under each case is given a record of the publications that have *referred* to the case, the other court decisions that have affected the case, and any other references that may be of value to the lawyer. This type of listing is particularly important to the lawyer, because, in law, much is based on precedent.

Citation indexes depend on a simple system of coding entries, one that requires minimum space and facilitates the gathering together of a great volume of material. However, a code is not absolutely necessary if one chooses to compile a systematic listing of individual cases or reports, with a complete bibliographic history of each of them. Thus it would be possible to list all pertinent references under each case with sufficient completeness to give the index more of the appearance of a bibliography. However, this would result in an extremely bulky volume.

There are analogies in bibliographic operations. For example, in cataloging books for booksellers' or library catalogues, an attempt is made to find references to each book in one or more authoritative bibliographic sources, such as the catalogues

Director of the Institute for Scientific Information, 33 South 17th Street, Philadelphia, PA, USA.

[†] First published *Science* 1955;122:108–11. Copyright AAAS. Reprinted with permission.

of the British Museum (BM), Bibliothèque Nationale (BN), or the Library of Congress (LC). The 'authority' card used in cataloging sometimes looks like a Shepard entry.

Another example is a book-review digest, in which one finds for each book title a series of references and selections from published reviews, critical and otherwise. Certain indexing publications perform a similar function.

Some time ago I became concerned with the problem of developing a citation code for science. This was necessary for the efficient manipulation by mechanical devices of entries to scientific indexes. In the course of this research I developed a very simple system for identifying an individual scientific article that had appeared in the periodical press. The resulting numerical code consisted of two parts. The first part was a serial number, used instead of an abbreviation, to identify each periodical; it was similar to the serial numbers employed in the *World List of Scientific Periodicals*, by no means a new idea. For example, *Die Bibliographie der fremdsprachigen Zeitschriften Literatur* has for many years used such a system to save space.

The second part of the code number was also a serial number, assigned to each article in a particular publication, starting with 1 and continuing throughout all volumes. The code thus gives no indication of year or volume number, a serious shortcoming. The article number is also not unique, having been used by the *Proceedings of the Society for Experimental Biology and Medicine* since its inception. These two serial numbers taken together, it can be seen, can identify any published periodical article. It soon became apparent, after such codes had been utilized on an experimental basis, that the use of the codes would facilitate the compilation of a citation index. (Other coding systems would be equally applicable.)

A citation index to science would have the following main characteristics. First, there would be a complete alphabetic listing of all periodicals covered, in addition to the code numbers for each periodical. The list would be similar to the *World List*, but without the library holdings information. The main portion of the citation index would list in straight numerical order the code numbers for all the articles covered. Under each code number, for example, 3001-6789, there would be listed other code numbers representing articles that had referred to the article in question, together with an indication of whether the citing source was an original article, review, abstract, review article, patent, or translation, and so forth. In effect, the system would provide a complete listing, for the publications covered, of all the original articles that had referred to the article in question. This would clearly be particularly useful in historical research, when one is trying to evaluate the significance of a particular work and its impact on the literature and thinking of the period. Such an 'impact factor' may be much more indicative than an absolute count of the number of a scientist's publications, which was used by Lehman³ and Dennis⁴. The 'impact factor' is similar to the quantitative measure obtained by Gross,⁵ in evaluating the relative importance of scientific journals, a method later criticized by Brodman⁶ but used again by Fussler.⁷

Other advantages would also be obtained. In a way such listings would provide each scientist with an individual clipping service. By referring to the listings for his article, an author could readily determine the other scientists who were

making reference to his work, thus increasing communication possibilities between scientists. It is also possible that the individual scientist, thus, might become aware of implications in his studies that he was not aware of before.

Most authors like to see how their works are received. Bringing together all book reviews and abstracts is very important, for it is not possible for an author to keep up with the thousands of publications in which his contribution might be reviewed. This applies equally to publishers. It would not be impossible to include books in the citation index. Indeed, as a first suggestion, the use of Library of Congress card numbers as the identifying code for books would seem appropriate.

It is necessary next to discuss some realistic questions concerned with the realization of such an index. Bitner⁸ has estimated that 30 000 cases are covered by *Shepard's Citations* in 1 year, the cases and articles appearing in not more than a few hundred publications. In 1953 about 1 million citations were added—close to 40 citations per case.

What is the prospect in scientific literature? The last published edition of the *World List of Scientific Periodicals* contained more than 50 000 titles in science and technology. It is variously estimated that between 1 and 3 million new scientific articles are published each year. The *Journal of the American Chemical Society* alone publishes more than 3000 per year, including ~2000 original articles. The order of magnitude is, therefore, potentially from 50 to 100 times as great as it is for *Shepard's Citations*.

However, not all these 50 000 publications are being covered in our present indexing activities, and yet this has not prevented us from continuing indexes of standard type or from starting new ones. Lack of complete coverage is not necessarily an argument against a citation index. It is in fact an argument in its favour. Coverage could perhaps be limited to the list of periodicals covered by one of the leading indexing services. This approach would, of course, have an immediate disadvantage. Such a *subject* selection would mean that less directly related subjects of interest would be excluded, and these are the publications that the individual is least likely to cover in his own research. It would be necessary to consider all the pros and cons in a selective approach and then to determine the possible utility of such a tool. For example, would a citation index to the 1500 periodicals covered by the *Current List of Medical Literature* be of real value, or, similarly, a citation index to the 5000 periodicals covered by *Chemical Abstracts*? The *Current List* would, in fact, offer a good starting point, since it already provides a unique code for the 100 000 items indexed by it each year. Presumably these are the most significant contributions in the covered fields for the year. If 10 is the number of references in the average article, then about 1 million citations would be involved. The preparation of that number annually is not unreasonable. Shepard's has already used well over 50 million citations in its publishing activities.

The ultimate success of a citation index would depend on many factors. For example, if each periodical would assign unique code numbers to the articles published, it would be possible for authors to list these numbers in their bibliographies and, thus, to save the work of coding on the part of the citation index staff. It is unlikely that such a development could take place in <5 or 10 years, but it is comparable to the problem of

getting publishers to include Library of Congress card numbers in their publications.

When such a large volume of data is to be handled, mechanical devices of high speed and versatility could be used to great advantage and would probably determine success or failure. Once the coding is done, compilation itself is quite mechanical. This could be done by means of conventional filing slips; the Shepard organization itself has used them successfully for 80 years. However, it would be facilitated by a mechanical approach using punched cards.

The utility of a citation index in any field must also be considered from the point of view of the transmission of ideas. A thorough scientist cannot be satisfied merely with searching the literature through indexes and bibliographies if he is going to establish the history of an idea. He must obviously do a great deal of organized, as well as eclectic, reading. The latter is necessary, because it is impossible for any one person (the indexer) to anticipate all the thought processes of a user. Conventional subject indexes are, thereby, limited in their attempt to provide an ideal key to the literature. The same may be said of classification schemes. In tracking down the origins of an idea, the citation index can be of real help. This is well illustrated by an example from my own experience.

Many years ago the Radio Corporation of America developed a reading-aid for the blind.⁹ This device had an electronic system for converting printed letters into recognizable sound patterns. Using the device, a blind man could scan a printed page; in a set of headphones he could hear a series of sound patterns, each letter having its own recognizable sound pattern. In effect, the words were spelled out, letter by letter, in code. I was particularly interested in this device, because I had been independently working on a device that would *copy* print, letter by letter, and reproduce it for bibliographic and other purposes. The two devices had something in common in that they both employed scanning devices. I then wanted to learn whether anyone had ever suggested that the RCA reading-aid could be used for this purpose. It will be apparent that if anyone had known of the RCA device and had thought of adapting it for copying purposes, a reference to the article might have been made. This reference could easily have been included in an article or patent that was not at all related to the problem of reading devices. A citation index would have given me just what I was after. Nothing could substitute for extensive reading, but a great deal of time could have been saved by bringing the appropriate works to my attention.

In the course of my reading, I did find a few references to this device, one in a book,¹⁰ and several others in periodical articles, one of which was a German article on the mechanization of philological analyses and concordance building. The latter article¹¹ did not discuss my own special interest in copying devices, but it did show the similarity between the author's and my own thinking from the point of view of letter-recognition devices, which is what the RCA device attempts to be. In other words, both of us were interested in this device as a letter-recognition device for the analysis of text.

In another instance the RCA article was unexpectedly cited in the journal *Electronic Engineering* in an article on information theory¹² that I was reading because of an entirely different interest. No subject indexer could have anticipated this cross-

breeding of interests. Perhaps there are many other articles and books unknown to me that have made similar references to this device. How can they be located when the main subject matter of the article is, on the surface, so unrelated in nature?

One might say that it would be possible to index articles more thoroughly to achieve the same results. For example, the article on information theory, if thoroughly indexed, might have included an entry under reading devices for the blind. Yet if this were done, our periodical indexing services would clearly become hopelessly overloaded with material that is not necessary to lead us to the micro unit—the entire article or one of its major sections. Although it might be said that no scientist interested in the greater comprehensiveness to be found in a citation index would object to having such a great mass of references in a subject index, this is impracticable. It would require an army of indexers to read the articles and identify the exact subject matter of every paragraph or sentence. Yet this would be necessary. To illustrate, it is only in the very last paragraph of the article on information theory that one would find a reference to reading devices for the blind.

Even if an army of indexers was available, it is still doubtful that the proper subject indexing could be made. Over the years changes in terminology take place, which vitiate the usefulness of a standard subject index. To a certain extent, this is overcome through the citation approach, for the author who has made reference to a paper 40 or 50 years old has interpreted the terminology for us. By using authors' references in compiling the citation index, we are in reality utilizing an army of indexers, for every time an author makes a reference he is in effect indexing that work from his point of view. This is especially true of review articles where each statement, with the following reference, resembles an index entry, superimposed upon which is the function of critical appraisal and interpretation. To the indexer this has its advantages as well as its disadvantages.¹³

To determine in a practical way what the citation index could offer, it was decided to track down the citations made in one journal to a single significant article, in order to compile a sample entry for the citation index. At the suggestion of Erich Meyerhoff, I selected Hans Selye's famous article on the general adaptation syndrome.¹⁴ A systematic search was then made of all papers that were published in the *Journal of Clinical Endocrinology* subsequent to Selye's paper up to 1951—a period of 5 years, including well over 500 articles. Every bibliography in each of the 500 articles was checked for a reference to Selye's article. Twenty-three articles were found to make such a reference; each of them was then checked for the character of the information provided.

Examination of the citation list (Table 1) shows the great variety of subject matter included. One thing became quite clear, even to the uninitiated—that is, the influence of Selye's article has been quite pronounced. Such evidence is extremely valuable to the historian.

It is interesting to note that, although all the articles cited were indexed in *Quarterly Cumulative Index Medicus*, not one is to be found there under the heading 'Adaptation'. In fact, it is surprising not to find any articles from this journal under this subject heading.

It also becomes quite obvious that many references to Selye's paper were general and contribute little or nothing to

Table 1 Index sample based on article by Hans Selye, 'General adaptation syndrome' [*J Clin Endocrinol* 1946;6:117]

1. Williams RH. Thyroid & adrenal interrelations. 1947;7:52-57.
2. Venning EH. Glycogenic corticoids. 1947;7:79-101.
3. Forbes *et al.* 17-Ketosteroids in trauma and disease. 1947;7:264-288.
4. Talbot *et al.* Excretion of 11-oxy corticosteroids. 1947;7:331-350.
5. del Castillo EB *et al.* Syndrome of rudimentary ovaries. 1947;7:385-422.
6. Forsham PH *et al.* Pituitary adrenocorticotropin. 1948;8:15-66.
7. Pincus G *et al.* Rhythm in biped excretion. 1948;8:221-226.
8. LeCompte PM. Width of adrenal cortex in lymphatic leukemia. 1949;9:158-162.
9. Wolfson WQ. 17-Ketosteroids in gout. 1949;9:497-513.
10. Stein HJ *et al.* Hormonal response to heat and cold. 1949;9:529-547.
11. Davis ME. Eosinophils in pregnancy & labor. 1949;9:714-724.
12. Conn JW. Na and Cl of sweat as cortical index. 1950;10:12-23.
13. Recant L *et al.* Effect of epinephrine on eosinophils. 1950;10:187-229.
14. McArthur JW *et al.* Urinary excretion of corticosteroids in diabetic acidosis. 1950;10:307-312.
15. Bors E. Fertility in paraplegic males. 1950;10:381-398.
16. Grossman S *et al.* Idiopathic lactation following thoracoplasty. 1950;10:729-734.
17. Cooper JS *et al.* Metabolic consequences of spinal cord injury. 1950;10:858-870.
18. Hioco D. Adrenal metabolites in bronchial asthma. 1950;10:1570-1578.
19. Jailer JW. Pituitary-adrenal system in infants. 1951;11:186-192.
20. Deane HW. The adrenals in experimental hypertension. 1951;11:193-208.
21. Hioco D *et al.* Epinephrine and ACTH in bronchial asthma. 1951;11:395-407.
22. Schaffenberg CA *et al.* p-Hydroxypropiophenone (PHP) and other so-called pituitary inhibitors. 1951;11:1215-1223.
23. Talbot NB *et al.* Urinary water-soluble corticosteroids. 1951;11:1223-1236.

Citation index entry

11123s-687_____

464-9789 (R)

869-3366 (R)

1105-9876 (A)

1123-4432 (R)

a11,123-0752 (O)

-0779 (O)

-7264 (O)

-7331 (O)

-7385 (O)

-0866 (O)

-8221 (O)

-9158 (O)

-9497 (O)

-9529 (O)

The code number for this journal in the *World List* is 11,123a; the article number is arbitrarily taken as 687; and the code number for the article is 11123a-687. The 23 articles that cited Selye's article are listed, followed by *A* hypothetical citation index entry for Selye's article: *R.*, review article; *A*, abstract; *O*, original article.

the readers' enlightenment, since exact page references are not provided. In several cases the Selye article is even cited but not referred to in the text. Selye's influence on all these authors is quite apparent. In particular instances the citations are of value in locating confirmatory evidence of some of Selye's claims.

Thus, in the case of a highly significant article, the citation index has a quantitative value, for it may help the historian to measure the influence of the article—that is, its 'impact factor'. With regard to a less significant work, one would suspect that the bibliographic advantages might be increased, because the scientist or librarian would be provided with references not to be found in conventional indexes. The preliminary evidence presented indicates that the citation index offers interesting possibilities for another approach to bibliographic control.

The next step in compiling the index for the Selye article would be to seek out additional references to it in more peripheral journals, but obviously the farther away you get from the immediate subject area of the main article, the fewer the references to it you will locate. Yet these may well be the most useful references of all, for the cross-fertilization of subject fields is one of our most important problems in science literature.

It will be well to close with a brief description of how the citation index might be compiled. The first step would be the selection of the particular group of periodicals to be covered; next, the period to be covered, say, only that since 1900.

The problem actually has two facets: the selection of periodicals to be covered in order to obtain citations and the selection of those articles for which we want a citation record. For example, all *articles* in journals in the *Current List of Medical Literature* that have remained in continuous publication since 1900 might be coded, in which case the *Journal of Clinical Endocrinology* would not be included. However, we might include as citation sources all *journals* covered by the *Current List*. Thus, the bibliographies appearing in articles in the *Journal of Clinical Endocrinology* would supply references to the basic group of articles.

Each coder would be assigned a group of articles in a particular journal. The first step would be to number each article in the journal in ascending order, by utilizing a complete table of contents of that journal from its inception.

Once a code number has been assigned to each article, the proper codes may then be assigned to each periodical. This might be the number given in the *World List*, with new numbers for any periodicals not to be found there.

Actual coding starts with the first article in a particular periodical. The coder prepares a 3 by 5 inch card for each citation made in the article. Each card should give (i) the code number for the citing article, (ii) the code number for the article cited, and (iii) a classification of the citing article as an original contribution, review article, abstract, and so forth.

Many references will be excluded by the limits of coverage set up. Thus all references to articles not in the prescribed list of journals would be excluded. All books would be excluded unless otherwise specified, in which case the reference card would carry the code for the citing article and the code for the book (its LC card number).

After all the articles had been coded, it would next be necessary to sort the cards by the code numbers for the items

cited. This would yield a group of cards for each cited article. These would then be sorted by code numbers for the citing articles. This completes the coding and sorting. The next step would be preparation for the printer.

From this description it will be apparent that, although a great volume of material is to be covered, relatively unskilled persons can perform the necessary coding and filing. Professional supervision would still be required, because certain decisions require skilled judgement, for example, when *ibid.* or *loc cit.* must be carefully interpreted. Footnotes tend to make coding somewhat cumbersome. The code I have described is merely an example used to illustrate the method in principle. If the system were adopted, then in the future every author ought to be required to include the serial number of each item he referred to, so as to facilitate not only the compilation of citation indexes but also other operations such as requests for reprints.^{15,16}

In a certain sense a citation index is not very different from a compendium like *Beilstein*, which gives a rather complete record of a compound, compiled by a similar method. A citation index for the literature of chemistry would undoubtedly make the preparation of such works as *Beilstein* much easier than it is at present. The new bibliographic tool, like others that already exist, is just a starting point in literature research. It will help in many ways, but one should not expect it to solve all our problems.

References

- ¹ Thomasson P, Stanley JC. *Science* 1955;**121**:610. Thomasson and Stanley were commenting on C. Zirkle's discussion of the use of fraudulent data [*Science* 120, 189 (1954)].
- ² Adair WC. *Am Document* 1955;**8**:31.
- ³ Lehman HC. *Sci Monthly* 1954;**78**:321.
- ⁴ Dennis W. *ibid.* 1954;**79**:180.
- ⁵ Gross PLK, Gross EM. *Science* 1927;**86**:385.
- ⁶ Brodman E. *Med. Library Assoc Bull* 1944;**32**:479.
- ⁷ Fussler HH. *Library Quart* 1949;**19**:19.
- ⁸ Bitner H. Personal communication, April 1954.
- ⁹ Zworykin VK, Flory LE. *Proc Am Phil Soc* 1947;**91**:139.
- ¹⁰ Shaw RR. *Machines and the Bibliographical Problems of the Twentieth Century*. Urbana: University of Illinois Press, 1951, p. 19. Reprinted from *Bibliography in an Age of Science* (University of Illinois Press, Urbana, 1951).
- ¹¹ Busa R. *Nachr Dokumentation* 1952;**3**:14.
- ¹² Andrew AM. *Electonic Eng* 1953;**25**:471.
- ¹³ Garfield E. Review literature as a source of critical entries for scientific indexes, unpublished paper, December 1952.
- ¹⁴ Selye H. *J Clin Endocrinol* 1946;**6**:117.
- ¹⁵ Garfield E. *Science* 1954;**120**:1059.
- ¹⁶ Schuke JA. *Science* 1954;**120**:1055.

Commentary: Fifty years of citation indexing

Eugene Garfield

About 30 years ago we began the series called 'This week's Citation Classic'. Over the next 15 years, we asked thousands of authors to write commentaries on these highly cited papers (www.citationclassics.org). So I was surprised to be asked to do the same for my 1955 paper in *Science*.¹ As a confirmed citationist, I must point out that it is not my most cited work. It is my 1972 paper in *Science*,² on using citation analysis to evaluate journals, which has attracted much more attention, although the 1955 paper is far more significant. In that sense, I am like many other authors who feel that their most-cited work is not necessarily their best. My most-cited work is in fact my 1979 book *Citation Indexing*.³

Tracing the genealogy of citations to this paper reveals the evolution of the concept of citation indexing from a system

for information retrieval to a tool for research evaluation. In a paper that I had prepared at the request of the then editor of *Science* in 1995, but rejected by his successor, I suggested that the tail was now wagging the dog.⁴

In the first few decades after the appearance of this 1955 paper, and its 1964 successor,⁵ most of the citing papers concerned the pros and cons of citation indexing for information retrieval. In those days there was a preoccupation with controlled vocabulary-based indexing. So we created the Permuterm Subject Index as a natural language supplement to the Citation Index. Henry Small much later would formalize the role of citations as concept symbols.⁶

An early portent of the use of *SCI* for evaluating science was the 1967 paper by Margolis.⁷ By that time Irving Sher and I had already done the simplistic exercise of sorting the *Science Citation Index* to produce a list of the 50 most cited authors. About one-third of these proved to be Nobel Prize winners and almost all were authors 'of Nobel Class'.⁸