CITATION INDEXING:*

A Natural Science Literature Retrieval System for the Social Sciences

by Eugene Garfield

The director of the Institute for Scientific Information of Philadelphia discusses the Institute's approach to the retrieval problem and emphasizes the multidisciplinary, linguistic, and quantitative factors common to all fields of knowledge. Citation indexing is illustrated and compared with conventional indexing, and its advantages in resolving retrieval problems are clearly stated.

THE increased attention devoted to the information problem of the natural sciences[1] has inevitably focused as well on the social sciences, or to use the newer term, the behavioral sciences. Indeed, realignments in the classical disciplines and the growth of multidisciplinary research have helped create the so-called information crisis. These qualitative changes have added to the quantitative difficulties of the conventional, discipline-oriented abstracting and indexing services. The quantitative change, due to the voluminous growth of literature, is primarily a mechanical problem of production. However, the qualitative change from discipline-oriented to multidisciplinary research is the more significant problem, for it poses a fundamental intellectual challenge.

For these reasons, it would be pointless and artificial to separate the literature problems of the "social" sciences from the "natural" sciences. The behavioral sciences epitomize the meaninglessness of the separation. The literature problems of the social sciences —and those of the humanities—can be treated with equal success, I believe, with techniques similar to those employed successfully in the natural sciences. Special difficulties may arise insofar as each particular research field allows or requires different degrees of indexing specificity, depending upon the methodological precision of

that field. However, as an information scientist, I find that even the informational problems of Biblical research are comparable to those of chemical research. In some aspects of both fields one deals with highly specific informational or data elements which are clearly defined and prescribed. Where such elements are not clearly defined, whether in the social or natural sciences, one encounters difficulties.

Time-Space Continuum of Information

By now it is obvious to information scientists that the "spatial" or N-dimensional problem of classifying scientific knowledge cannot be treated "geographically"—as a map-maker handles the boundaries between countries. Knowledge is multi-dimensional and does not cluster in convenient

parcels like countries or homesites. No more rational is the tendency to aggregate indexes in yearly packets, artificially separating the past and present literature. Concomitantly, it is not really possible to separate the concepts of retrieval and dissemination. The artificiality of this separation is illustrated in the ambiguous uses of the term "information retrieval." IR has become a vague concept which covers all aspects of the information problem. Documentalists, however, generally speak of disseminating "current" information (recently published information), as in a newspaper or journal, whereas "past" information, stored in archives and libraries, is retrieved. However, one man's current awareness may be another man's retrieval. What is the alternative to the meaningless chronological and arbitrary subject distinctions of typical conventional abstracting (disseminating) and indexing (retrieving) services? I believe that these problems are largely overcome by citation indexing —an index methodology and philosophy which can be used interchangeably for retrieval and dissemination, and which does not entail artificial separation of the past from the present and future.

Citation Indexing

A citation index[2] is an ordered list of references (cited works) in which each reference is followed by a list of the sources (citing works) which cite it. The cited work may have been quoted, discussed, criticized (as in a book review), etc. In an explicit citation, the source or citing document will identify the cited works by use of

LINDNER RM	•1944•REBEL WITHOUT CAUSE–❑			
PAINTING DH	J ABN SOC PSYCH	1961	62	352
RANDOLPH MH	J CONS PSYCH	1961	25	293
	–1946–STONE WALLS AND MEN–❑			
PAINTING DH	J ABN SOC PSYCH	1961	62	352
	–1948–PSYCHOANAL QUARTERLY–❑	17	454	
DAVIS DR	BRIT J MED PSYC	1961	34	53
	–1952–PRESCR REBELLION––––❑			
BAILEY P	PERSP BIOL MED	1961	4	199
WINTHROP H	PSYCHOL RECORD	1961	11	119
	–1956–BRIT J MED HYPNOT–––❑	7	2	
MONTSERR. S	ACT PSYCHOTHER	1961	9	429
	–1956–MUST YOU CONFORM–––❑			
ADAMS ME	GERIATRICS	1961	16	177
WINTHROP H	PSYCHOL RECORD	1961	11	119

Figure 1: Representative entry from a citation index.

58

formal reference citations, which enable the reader to locate the document in question. In an *implicit* citation, one recognizes that some other work has been drawn upon or alluded to, but the citing author does not consider it important enough for a formal citation. Figure 1 illustrates the information to be found in a citation index. The sample is based on data taken from the *Science Citation Index —1961*, an index to approximately one million articles, books, etc.[•] In this sample, a half-dozen different books and articles by R. M. Lindner, published between 1944 and 1956, have all been cited in 1961—the source year. For example, a book by Lindner, published in 1952, was cited in an article by Percival Bailey in the review journal *Perspectives in Biology and Medicine* 4, pp. 199-265 (1961). It was also cited by Winthrop in the *Psychological Record* 11, p. 119 (1961).

Perhaps the most dramatic facet of this example is that the *Science Citation Index* was not designed to cover psychiatry or psychoanalysis *per se*. If anything, the bias in our 1961 experiments with citation indexing was toward the field of genetics.[•] The interdisciplinary nature of the information-searching problem faced by behavioral scientists is typified by this example. Pertinent information on Lindner's *Prescription for Rebellion* has been found in a non-psychiatric journal source. Even if Bailey's article had been selected by one of the discipline-oriented abstracting or indexing services, its title, "A Rigged Radio Interview—with Illustrations of Various Ego-Ideals," would surely have been inadequate for conventional word or subject indexing. Not only is the word "radio" irrelevant, but even the term "ego-ideals" is far from adequate to describe the ground covered in this article, an excellent critique of psychoanalysis containing 161 references. However, any one of these references, through citation indexing, would lead to Bailey's paper.

Automatic Citation Alerting Service

Similarly, using any *one* of the 161 papers and books quoted by Bailey,

one could have been alerted automatically to his article using the *Automatic Subject Citation Alert* (ASCA), a new current-awareness service designed by the Institute for Scientific Information to complement the retrieval uses of the printed *Science Citation Index*. Each scientist using ASCA will prepare a "profile" consisting of the question or subject citations *for specific works* which reflect his "current" interests. Each week he will receive a report showing which current source journal articles have cited any of the citations in his profile. Thus, as in the Lindner example, a scientist can solve a retrieval problem by referring to the printed, cumulated citation index which shows past citations to key works. On the other hand, using one or more of the key books or articles as a question citation, the user can be kept currently posted on newer work "related" to the particular "subjects" (citations) involved.[•]

Citation Indexes Compared to Conventional Indexes

Scientific nomenclature and terminology change rapidly. Since conventional indexing systems rely on nomenclature, they quickly obsolesce. To overcome this characteristic is impractical since it would require constant re-indexing of older literature by trained subject specialists. However, such re-indexing is precisely what is achieved routinely in a citation index. Each new reference to an "old" paper re-indexes it. Thus, the citation index is a self-organizing system since each new citation modifies the previously existing store of information.

In conventional indexes, the indexing or encoding must be done by trained indexers. The user must also be extremely well-trained. He must learn to use each particular index and its terminology. Therefore, the best designed word index must be highly cross-referenced. In sharp contrast, a citation index does not require indexers trained in special nomenclature. This is not to say a citation index does not use experts. The most expert subject specialists—the authors—provide the citations, and by so doing, index the literature.

Another important difference be-

tween conventional and citation indexing is the time dimension. In the conventional index the source literature is the new or current year indexed. The chronological coverage of a conventional index corresponds to the coverage of the *Source Index* which accompanies the *Science Citation Index* proper. On the other hand the citation index proper not only indexes the current source literature but merges the new with the older, reindexed data. Thus, the citation index, as a self-organizing system, is constantly being upgraded by the feedback of more current information. Advances in science are quickly reflected in the index, thereby bringing old information up-to-date.

Consider now the criteria for selecting articles to be abstracted and indexed. Discipline-oriented services are selective. According to the interpretation of editors or abstractors, individual articles may or may not be included. In the *Science Citation Index* complete coverage is guaranteed. Every source article, in every journal covered, is indexed, and every reference citation in every article is included in the index. This comprehensive approach is vital because it gives a significance and degree of certainty to a literature search not hitherto possible—especially when the result is negative. In a conventional index, a negative result is uncertain both because the user cannot be sure whether a particular article was included in the particular index searched, and also because his terminological approach may be different than that of the indexer. In a citation index, however, a negative search result is fully prescribed and terminology is eliminated.

After retrieving a document with the aid of a conventional index, one can search back in time through the bibliography. The starting point, therefore, is very critical. If it is not the most recent paper, it is impossible to find all current papers from its bibliography. In a citation index the user can not only go backwards in time through the bibliographies of documents, but can also come forward in time, retrieving documents even more current than the starting point. This added dimension enables one to

build up a bibliography by alternately examining the citation index and the items in the bibliographies of the papers retrieved.

Citation Networks and Search Strategies

It is instructive to compare conventional indexes to citation indexes by constructing a model of the literature. Figure 2 shows a network diagram for
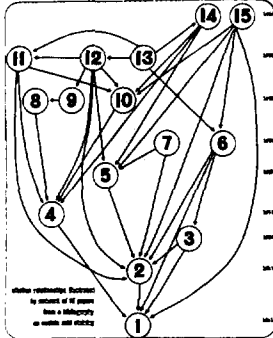
a bibliography of fifteen citations on the subject of nucleic acid staining.' The diagram, or historical map, is arranged chronologically from top to bottom. Each circle or node represents a published paper and each arrow indicates a reference citation. Thus, paper #1 is cited by #2, 3, 4, 6 and 15, whereas paper #15 cites #1, 2, 5, 6, and 10.

Using a typical conventional index one conducts a search for each year. If I examine the indexes for 1959, and presuming I use correct terminology, I might identify papers #11, 12, 13—all published in 1959. The common practice is then to retrieve these papers from the library and examine their bibliographies for footnotes to identify the *earlier* papers on the subject. Thus, paper #11 cites #2, 4, and 10. If one retrieves these papers both #2 and #4 cite #1. By this process one gradually builds up a bibliography. Note however that paper #15 cannot

be located unless an index is available covering 1960 as well. Paper #7 would not be found until one had searched each annual index back to 1956.

In using a citation index, however, the scholar does not have to ensure that his starting point is the most recently published paper on the subject. He could begin with paper #1 or with any other paper, and a cumulated citation index would immediately inform him that this paper had been cited by #2, 3, 4, 6, and 15. By again checking the citation index he would determine that paper #2 has been cited by #3, 5, 6, 7, 11, 12, and 15. By continuing this process he could rapidly build up a bibliography while simultaneously following the natural, historical development of the field. Obviously, one can and does combine the use of the citation index with use of footnotes in the papers retrieved to make the search both rapid and highly selective.

Citation Indexes in Historical and Sociological Research

Figure #2 also suggests the great potential value of citation indexing for sociological and historical research, *per se.*⁴ At the present time the Institute for Scientific Information is engaged in research along these lines. One project concerns the determination of the feasibility of computer-produced historical network diagrams.⁵

The author arrangement of present citation indexing should not obscure the high degree of specificity achieved in this index, in that particular words, sentences or paragraphs may in fact be the "subject" indexed. A highly specialized citation index to the Bible,⁶ among other possibilities, has illustrated that citation indexing, far from being limited to use in the experimental sciences can be equally powerful when applied to humanistic and literary subjects. For example, book reviews are indexed in the 1964 SCI, as are other types of literary material. References to private communications, letters, and manuscripts are also included.

Another reason why citation indexing will probably find great popularity in the social sciences and humanities is that conventional language-oriented indexes are even more difficult to use

than they are in fields like chemistry and physics. Ambiguous terminology is rampant in the humanities and each term has many different meanings depending upon the context involved.

Difficulties of Citation Syntax and Custom

Nevertheless, it will be more difficult, and thereby somewhat more costly, to process more scholarly works in the *Science Citation Index*, because citation practices in scholarly journals have "evolved" to the point where it frequently requires a scholar, instead of a clerk, to unravel the particular work cited in a footnote or bibliography.

However, none of the problems discussed here are insuperable. It merely remains for the scholarly community to determine whether the obviously useful results of sophisticated information retrieval systems, such as the *Science Citation Index*, justify the dedication of time, money, and energy recommended by the President's Science Advisory Committee.¹ I believe that it does, and that the rewards to the scientific community and society will far exceed the effort involved.

Current Contents

Although the Institute for Scientific Information is now concerned principally with indexing the literature of the natural sciences, it began its publication activities in the social sciences. In 1955 the Institute first issued *Current Contents of Social and Management Sciences,* a semi-monthly publication which consisted entirely of facsimile reproductions of contents pages of some 200 journals. Through *Current Contents* it was possible for the scholar or executive to scan quickly through the current journals, check off articles of interest, and then locate them in the library or obtain tear sheets directly from the ISI library. The publication was extremely popular with those who did use it,¹⁰ but it never gained wide support and was suspended in 1962.*

Essentially the same approach has caught on splendidly with bio-medical scientists and those in allied fields.¹¹ Another *Current Contents* edition is now also gaining popularity among physicists and other research

THE AMERICAN BEHAVIORAL SCIENTIST

workers in the physical and space sciences.[13] As research support in the social sciences increases, it may be appropriate to reintroduce *Current Contents of Social Sciences*, though in far more comprehensive form.

Conclusion

This article has attempted to explain the approach of the Institute for Scientific Information to the diverse problems of retrieving and disseminating information. I have emphasized the multidisciplinary, linguistic, and quantitative factors common to all research fields today, and suggested the ways in which citation indexing resolves the problems inherent in these factors. I have compared conventional indexing to citation indexing. The past, present, and future use of contents page or title listing

publications has been discussed in terms of the browsing problem. Both types of services are significant factors in the information discovery process and ought to be adequately tested in the social and behavioral sciences. The expectation, in spite of special obstacles presented by social science materials, is that they can be at least as effective as in the natural sciences.*

References

1 A. M. Weinberg, et al., President's Science Advisory Committee, *Science, Government, and Information* (The Responsibilities of the Technical Community in the Transfer of Information). Washington, D.C.: U.S. Government Printing Office, 1963, 52 pp.

2 For a more detailed description of the nature and uses of the citation index see E. Garfield, "The Science Citation Index—New Dimension in Indexing," *Science*, 144 (1964), 649-54.

3 E. Garfield and I. H. Sher, *Science Citation Index*. Institute for Scientific Information, Philadelphia 6, Pa., 1963, 3,872 pp.

4 E. Garfield and I. H. Sher, *Genetics Citation Index*, Institute for Scientific Information, Philadelphia 6, Pa., 1963, 864 pp.

5 Prepared by Gordon Allen, National Institute of Mental Health, June 2, 1960, whose assistance is deeply appreciated.

6 E. Garfield, "Citation Indexes in Sociological and Historical Research," *Amer. Documentation*, 14 (1963), 289-91.

7 E. Garfield and I. H. Sher, "New Factors in the Evaluation of Scientific Literature through Citation Indexing," *Amer. Documentation*, 14 (1963), 195-201.

8 Air Force Contract # AF 49(638)1256.

9 E. Garfield, "Citation Indexes to the Old Testament," American Documentation Institute, November 1955.

10 L. E. Albright and J. R. Glennon, "Journal Tables of Contents," *Amer. Psychologist*, 18 (1963), 109-10.

11 *Current Contents of Chemical, Pharmaco Medical and Life Sciences*, Institute for Scientific Information, Philadelphia 6, Pa., weekly.

12 *Current Contents of Space, Electronic and Physical Sciences*, Institute for Scientific Information, Philadelphia 6, Pa., weekly.

502