EVENT NUMBER 103.2

American Society for Engineering Education

Annual Meeting, June 17-21, 1963

Philadelphia, Pennsylvania

CITATION INDEXES - CLOSING THE INFORMATION GAP BETWEEN THE PURE AND APPLIED SCIENCES

by

Eugene Garfield, Ph.D.
Director
Institute for Scientific Information
Philadelphia, Pennsylvania

Abstract

The engineer's information requirement is generally quite specific, though it may be vague and difficult to articulate. Citation indexes are capable of retrieving information on highly specific subjects provided the user has minimal knowledge of his search topic. In addition, citation indexes can open up new areas of product application in the same way that it enables one to trace the use of new theories or ideas. Search strategy in citation indexes frequently involves a technique called "cycling" which overcomes the failure of certain authors to provide pertinent citations. Use of this strategy from an experimental citation index to 1.4 million references will be illustrated by some "typical" engineering problems.

The scientist who has ever been involved in the complex process of research has, on one occasion or another, been stumped by the problems of how best to search the maze of accumulating scientific literature. Once upon a time he considered it adequate to depend upon classically organized indexing systems. This traditional method still works, but today has distinct disadvantages and limitations. In this modern era, modern techniques of documentation must naturally arise to lead the scientist out of his dilemma.

The Institute for Scientific Information believes it can with the Citation Index.

During the past two years, the Institute has been concerned with citation indexes to science. An intensive, original research program is compiling comprehensive citation index files, experimental computer printouts, studying and testing their practicality in the dissemination, retrieval and evaluation of scientific literature. Under joint sponsorship of the National Institutes of Health and the National Science Foundation, work began officially on January 1, 1961.

The Citation index is a dynamic, straightforward bibliographic system which not only makes up-to-date literature searching practical, but also stimulates the process of information discovery. It provides a variety of features beyond the scope and power of conventional indexing systems, at a time when these systems cannot cope with the increasing welter of data erupting from modern poolings of scientific talent. This disturbing mass, accentuated by the burgeoning of new specialties has paralyzed conventional techniques.

This same prodigious growth is the vital force of the citation index method.

In short, the Science Citation Index developed by the Institute is a self-organizing system which thrives on a growing literature. It is an orderly arrangement that improves qualitatively as the quantitative character of the literature changes, constantly adding, modifying and developing itself. A living, growing system is achieved that rescues much of what is destined otherwise to become debris (lost and forever useless) and turns it into meaningful, well-disseminated and retrievable knowledge.

To appreciate the citation index method, two major points must be emphasized and understood clearly: the basic concept and the basic indexing technique.

The citation index "idea" is founded upon pragmatic philosophy--"what has happened since?"...exactly <u>what</u> has a man published and precisely <u>who</u> or <u>what</u> from any field of endeavor has cited <u>any</u> of his works following its publication.

This concept will be familiar to the attorney, for the citation index is an indispensable research tool of the legal profession. Since 1873, attorneys have relied upon <u>Shepard's Citations,</u> a comprehensive system which indexes individual American court decisions giving a complete history: publications that have since referred to the case, other court decisions that have affected the case, and literature references in which the case is discussed.

In creating the Science Citation Index, a somewhat similar a <u>posteriori</u> theory of indexing is utilized. The 'logic' of all conventional scientific classifications has inevitable broken down with experience. Aristotelian logic has been a chain around the neck of the scientist and classifier alike.

If the old rigid classifications are eliminated, the problems of changing terminology, and the other inherent limitations of a priori indexing can be solved." E. Garfield, Chemical Bulletin, 43:4 (1956)

Thus, in citation indexing, the familiar hierarchial classifications, hopelessly non-specific to the scientist, are discarded. Because the scientist is usually aware of a particular work covering his specialized interests, more often than not, he would begin a search in the citation index with a citation for a specific paper. Under conventional indexing systems, he cannot do this. The citation index introduces a new technique of indexing and provides the scientist with an improved starting point: the specific work of a specific author, at a specific time and location - author, year, journal, volume and page.

The citation index lists the author and his work, and with it, groups together all those diverse authors and papers, from field to field, who referred to that same work since its publication. Each of the author's works (accompanied by its citing authors) are arranged chronologically. As rapidly as they become available, the citation index continues to add new works of an author or new citation listings to his previously recorded papers. The result of such an input is a comprehensive, interdisciplinary, elastic index which is always up-to-date and encompasses the entire body of scientific literature.

Let us take as an example, Dr. Joshua Lederberg's article, "Genes and Antibodies," published in Science (129:1649, 1959), listed in a preliminary sample citation index compiled from 326,000 citations appearing in 1960 journals. This citation index shows fourteen authors who have cited this specific paper. Three of these fourteen have been cited in turn by six other authors. Thus the scientist looking up Lederberg and this particular work in the Science Citation Index would be led, in a matter of minutes, to a core of twenty authors whose work is in some way related to Lederberg's original article. These twenty papers were published in journals ranging from immunology to biological chemistry and cancer research to general science publications.

Now, for the first time in the indexing of science literature, the researcher is brought forward in time to the most recent published work referring to a particular author. Conventional indexing systems carry the user back to the past only, showing who has preceded, rather than followed an author. Still, in a citation index, the searcher can, if he chooses, go similarly backward in time, and compile additional material by consulting bibliographies appearing in the various citing works. This technique of "cycling" produces a literature network which satisfies a need not only for covering a limited area in depth, for browsing, but also for keeping pace with developments in a chosen area of interest, reducing unintentional and unnecessary duplication of research.

Further analyses define the interdisciplinary character of the index and show that seemingly unrelated material is pulled together, cutting across the apparent disciplinary lines implied by the journal title. Because the papers are collated on the basis of common citations, works that might never be brought together, by even the most sophisticated high level indexing, are routinely pulled together by the citation index. An interesting example is the unexpected relationship ISI's Citation Index sampling uncovered when it collated the paper of C. H. Whitnah, J. Dairy Science 42,27 (1959), and a paper by Albert Einstein, Ann. Physik 19,289 (1906). The apparently incongruous combination proved to be a legitimate reference to an equation

used in calculating molecular dimensions and was being applied in a study of the physical properties of milk. This same article by Einstein was cited in an article by V. V. Varadaiah, Polymer Science, 46,528 (1960) in which Einstein's equation was used as the basis for calculations relating to the Flory universal constant, in the equation for intrinsic viscosity. In yet another paper, by Elworthy, J. Chem. Soc. (1959), 1951 the Einstein equation was cited in a discussion of the size and shape of lecithin micelles. The Einstein article well illustrates the utility of Citation Index in crossing scientific disciplines, in picking up isolated, seemingly unassociated bits of information buried in the literature.

What is current? In the citation index method there are no artificial time barriers. Most conventional indexing services are limited to a relatively short and specific time period; the past six months or the past year. In the citation index, older, but still currently important information and more recent information come together naturally, as illustrated above in the citations to Einstein's 1906 article by three contemporary writers.

The citation index is based on the supposition that every author is "his own indexer" -- that is, in effect, each time he adds a citation to his paper, he is re-indexing the scientific literature. The natural interplay of old and new upon each other are an indissoluable part of citation indexing. These constant reshufflings and modifications in the citation index are its source of dynamic power.

By arranging literature in its natural chronological order, the citation index offers a simplified practical solution for easy, rapid tracing of the history or current evaluation of an idea, method or discovery. Unfortunately, many authors do not know what becomes of their work after it appears in print. Many are totally unaware of new insights, meanings and implications of their works revealed by their contemporaries. Through the citation index, the scientist can see which scientists are citing his works -- using his methods or that of colleagues -- modifing them, improving, criticizing, etc. Communication is thereby fostered, especially between researchers in different fields.

By eliminating arbitrary subject classifications, the citation index eliminates intellectual indexing decisions by catalogers. Compiled by a mixed machine and clerical process, there are none of the errors, inconsistencies, ambiguities or semantic confusions of language-oriented indexing.

Although output can be as comprehensive as input, computer sorts can also yield a selective citation index service suitable to any specific field, such as genetics. The criteria of selection include quantitative as well as qualitative factors. For example, the Institute is using the citation index to define the subfield of genetics and locate active researchers in that field. From the 1961 file of 1.4 million citations all papers which meet a set of "genetics" criteria were extracted. Computer programs have been designed to examine each individual citation for the purpose of determining whether the cited author, citing author or journal meets one of the "genetics" criteria set forth. The resulting printouts reveal a selective group of geneticists and journals considered to be relevant to the field of genetics.

The file of over 1.4 million citations from the 1961 literature reveals an estimated 750,000 different papers have been cited. Every citation appearing in 600 journals was included, making available for the first time a large reservoir of interdisciplinary data for testing the power and need of citation indexes. A number of additional studies designed to learn more of the characteristics of the literature are suggested. Some of these experiments involve using citation networks for measuring the effects and "impact" ratios of authors and journals, sociological and historical research, utilization of journals, "cycling" experiments, definition of scientific disciplines and statistical information. All of these possibilities remain to be completed experimentally.

This is intended to introduce the reader to the citation index concept. Copies of several earlier papers are available upon request. Your comments are invited.

Garfield, E. "Citation Index for Science" Science 122:3159 108-111 (1955)
Garfield, E. "Citation Indexes - New Paths to Scientific Knowledge" Chemical Bulletin 43:4, 11-12 (1956)
Tukey, John W. "Keeping Research in Contact with Literature: Citation Indices and Beyond" J. Chem. Doc. 2:34-7 (1962)