Back to the Contents Page

Chapter One

A Conceptual View of Citation Indexing

The concept of citation indexing is simple. Almost all the papers, notes, reviews, corrections, and correspondence published in scientific journals contain citations. These cite—generally by title, author, and where and when published—documents that support, provide precedent for, illustrate, or elaborate on what the author has to say. Citations are the formal, explicit linkages between papers that have particular points in common. A citation index is built around these linkages. It lists publications that have been cited and identifies the sources of the citations. Anyone conducting a literature search can find from one to dozens of additional papers on a subject just by knowing one that has been cited. And every paper that is found provides a list of new citations with which to continue the search.

The simplicity of citation indexing is one of its main strengths. In compiling a traditional subject index, someone, usually with specialized knowledge of the subject covered, must read, or at least scan, each document and make a series of intellectual judgments. These involve the selection of subject terms that describe the contents of the document. The greater the depth of the indexing (the more terms used to describe the document), the more judgments the indexer must make. These judgments take time, which quickly runs up the cost of the indexing operation and may reduce the timeliness of the finished product. It is not unusual for subject indexes to be a year or more behind the literature, a situation that limits their effectiveness as search tools.

The cost of indexing depth is even more critical. Though most of the major subject indexes are either financed completely or subsidized by the federal government and, therefore, do not pass on full production costs to their subscribers, they all must make a tradeoff between indexing depth and cost.

As a way of getting around this tradeoff, some organizations have adopted in recent years a method called "title word indexing." In this approach to subject indexing, a document is indexed by the terms used by the author in the title of the docu-

ment. There is no need for an intellectual judgment by an indexer; all terms that appear in the title are considered acceptable indexing terms. A title-word index, therefore, can be compiled much faster, and sometimes cheaper, than a traditional subject index.

There are, however, some penalties associated with the speed and economy of title-word indexing. For one thing, titles tend to be limited in length, which limits the number of terms available for indexing. This produces a rather shallow index, which tends to focus only on the main subjects of the papers and overlooks all the material they contain that is ancillary to those subjects, but may be important to the user of the index. Another penalty is that the ability, and even inclination, of authors to compose titles that accurately describe what they are writing about is very uneven. Consequently, the quality of the terms used in a title-word index also is very uneven.

There are some ways of mitigating these shortcomings. In some title-word indexes, depth is increased by adding subject terms assigned by indexers. In others, the quality of the index is improved by showing each title term in the context of all the other terms with which it has been used.

Citation indexing solves the depth versus cost problem by substituting the authors' citations for the indexer's judgments. This approach has the advantage of eliminating the need for intellectual indexing without compromising either the depth of the index or the quality of its "terms." Since there normally are no limits imposed on the length of bibliographies, there are no artificial limits on the depth of a citation index. And though the quality of bibliographies is variable from author to author and from one type of paper to another (review papers, for example, generally have more comprehensive bibliographies than any other type), the standards of good science exposition and the practice of viewing a good bibliography as a sign of scholarship tend to make citation quality considerably and consistently higher for indexing purposes than title quality.

Another important strength of citation indexing is its search effectiveness. This quality has two components. One is search productivity, which is concerned with finding the largest possible number of relevant papers. The other is search efficiency, which is concerned with minimizing the number of irrelevant papers the searcher must check out to identify the relevant ones.

Indexing depth is the primary quantitative measure of search effectiveness. The more indexing statements used, the more detailed the description of the document. As indexing depth increases, so does the probability that the searcher will satisfy his or her needs. Since the average scientific article contains approximately 15 citations, a citation index has an average depth of 15 "terms." Most traditional subject indexes can't afford to match this depth.

There also is a qualitative side to search effectiveness that revolves around how precisely and comprehensively an individual indexing statement describes the pertinent literature.

The precision of the description is a matter of semantics, which poses a series of problems in a subject index. The basic problem is that word usage varies from person to person. It is patently impossible for an indexer, no matter how competent, to

reconcile these personal differences well enough to choose a series of subject terms that will unfailingly communicate the complicated concepts in a scientific document to anyone who is searching for it.

The job is made more difficult by the dynamic nature of language. New terms are introduced, old ones disappear, and new meanings are attached to old words.

Another part of the problem is the need to standardize on the terms used in a subject index in order to exercise some degree of control over the consistency and quality of the indexing. This reduces the richness and variety of language available to the indexer for dealing with the first two problems.

The result of all this is to make the search process more complicated, less productive, and less efficient than it can be.

Citations, used as indexing statements, provide these lost measures of search simplicity, productivity, and efficiency by avoiding the semantics problems. For example, suppose you want information on the physics of simple fluids. The simple citation "Fisher, M.E., *Math. Phys.*, **5**, 944, 1964" would lead the searcher directly to a list of papers that have cited this important paper on the subject. Experience has shown that a significant percentage of the citing papers are likely to be relevant. There is no need for the searcher to decide which subject terms an indexer would be most likely to use to describe the relevant papers. The language habits of the searcher would not affect the search results, nor would any changes in scientific terminology that took place since the Fisher paper was published.

In other words, the citation is a precise, unambiguous representation of a subject that requires no interpretation and is immune to changes in terminology. In addition, the citation will retain its precision over time. It also can be used in documents written in different languages. The importance of this semantic stability and precision to the search process is best demonstrated by a series of examples.

The precision directly affects search efficiency. If, for example, someone is searching the 1971 *Index Medicus* subject index for information on the effect of aspirin on prostaglandin production, a reasonable starting point would be the term "prostaglandins." That term would lead to a list of hundreds of papers organized under a variety of subheadings. The subheadigs most pertinent to the subject of the search would be ANTAGONISTS and INHIBITORS, and BIOSYNTHESIS. The first lists 10 articles, the second, nine. Two articles under each of the subheadings appear to deal with the aspirin-prostaglandin relationship. But the seacher would have to read the titles of all 19 to identify the four relevant ones. In contrast, a search of a 1971 citation index on "Vane, J.R., *Nature New Biology*, **231**, 232, 1971" would lead, in a single step, to a list of 15 papers. Because they had all cited a paper on "Inhibition of Prostaglandin Synthesis as a Mechanism of Action for Aspirin-like Drugs," the precise subject of the search, this list of papers is likely to contain a high percentage of relevant material.

Precision also can affect search productivity. In 1963 an important paper was published on the topic of "seasonal variations in birth." Anyone searching the 1965 edition of *Index Medicus* for information on the subject would be unable to find the key paper unless he or she looked under the unlikely heading of "periodicity."

The importance of semantic stability to search productivity is demonstrated by a search on the subject of "euphenics." This word was first coined by Professor J. Lederberg in a paper published in *Nature* in 1963. Until then, the subject was described by the term "engineering human development," a term that is still used. A search of a subject index on the term "euphenics," while it was still a new word, probably would have identified papers that used that term, but not ones that used "engineering human development." Conversely, a subject-index search on the term "engineering human development" would have been likely to miss papers that used "euphenics." In a citation index, however, all papers that cited the key Lederberg paper would be listed together, regardless of the terms the authors used in describing the subject. Even if the subject index were searched on both terms, there would be a significant difference between its search effectiveness and that of a citation index. A complete search of the subject index would require two lookups, whereas a complete search of the citation index would require only one. As the terminology of the subject has continued to evolve to include "genetic engineering," "plasmid engineering," and "recombinant DNA," the semantic stability of key papers in the field (1) has made the effectiveness of citation searching even more pronounced.

How comprehensively an individual indexing statement describes the pertinent literature is a matter of the disciplinary and time constraints built into it. A subject index covers the literature of a specific discipline or group of disciplines, such as chemistry, medicine, or biology, within a specified time frame, and it imposes these limits on the scope of its indexing terms. This is not the case with a citation index.

Even if a citation index were compiled from the literature of a limited number of disciplines published within a given time frame, the citations it would use as indexing statements would not be bound by these limits at all. Authors consistently cite papers outside their discipline, and the citations range over the entire time spectrum of twentieth-century science. The use of these citations as indexing statements enables a citation index to provide a trail of information that follows the convoluted process of scientific development as it crosses disciplinary lines and moves back and forth in time. This characteristic greatly increases the search productivity of a citation index.

One example demonstrates, rather spectacularly, the cross-disciplinary reach of a citation index. From 1961 to 1969 a citation for one of the classic papers published by Albert Einstein in *Annallen der Physik* in 1906 is linked in a citation index, called *Science Citation Index,* to papers from the *Journal of Dairy Sciences, Journal of the Chemical Society, Journal of Polymer Science, Journal of Pharmacy and Pharmacology, Comparative Biochemistry and Physiology, Journal of General Physiology, International Journal of Engineering Science, Journal of Materials, Journal of the Water Pollution Control Federation, American Ceramic Society Bulletin, Journal of the Acoustical Society of America, Chemical Engineering Science, Industrial and Engineering Chemistry Process Design and Development, Journal of Colloid and Interface Science, Journal of Fluid Mechanics, Journal of Lubrication Technology, Journal of Molecular Biology, Journal of Food Science, Journal of Biological Chemistry, Journal of Sedimentary Petrology, Review of Scientific Instruments,* and the *Journal of the Electrochemical Society.*

Equally as important to search productivity is the time reach of a citation index. As the example of the 1906 Einstein paper showed, a citation-index search can start at any point in time at which a pertinent paper on the subject was published. If that paper was cited during the year covered by the index, the searcher will be brought back to the most recent information. Then the recent papers can be used to cycle back in time to the publication dates of their citations, which again will bring the searcher back to the current literature. In this way a searcher can often use a single edition of a citation index to obtain a view of the historical development of a subject that would require a methodical search of many editions of a subject index.

The ability to search back and forth in time from the past literature to the current literature, to identify cross-disciplinary developments, to eliminate the search restrictions and complexity imposed by semantic problems, and to provide an in-depth index to the literature within a practical time and cost framework have all proved to be as significant in practice as they appear to be in theory. Within 10 years, citation indexing had gained acceptance, despite its newness and departure from traditional indexing methods, as an important method of conducting retrospective searches of the scientific literature.

And its impact has gone beyond even that. Citation indexing has come to play an important role in current-awareness services, library management, and studies of the policies, history, and literature of science.

REFERENCE

1. **Garfield, E.** "Genetic Engineering—Too Dangerous to Continue or Too Important to Discontinue?" In *Essays of an Information Scientist,* Vol. 2 (Philadelphia, ISI Press, 1977). Pp. 335–341.

To Contents Page

To Chapter 2