Essays of an Information Scientist, Vol:1, p.188-195, 1962-73   Current Contents, #12, June 23, 1971
Reprinted from: Encyclopedia of Library and Information Science, Vol:5, p.16-40, 1971 - Marcel Decker Inc. NY.

CITATION INDEXES

For Introduction to this essay: "ISI announces a Five-Year Cumulation of the Science Citation Index" p.186, Essays, Vol:1

By

Melvin Weinstock, Senior Information Scientist
Institute for Scientific Information

PART I

Also See Citation Indexes - Part II

Citation Indexes - Part III

Introduction

Citation indexing is a relatively new method of organizing the contents of a collection of documents in a way that overcomes many of the shortcomings of the more traditional indexing methods. The primary advantage of citation indexing is that it identifies relationships between documents that are often overlooked in a subject index. An important secondary advantage is that the compilation of citation indexes is especially well suited to the use of man-machine indexing methods that do not require indexers who are subject specialists. This helps to make citation indexes more current than most subject indexes. Furthermore, citations, which are bibliographic descriptions of documents, are not vulnerable to scientific and technological obsolescence as are the terms used in subject indexes.

Citation indexing is based on the simple concept that an author's references to previously recorded information identify much of the earlier work that is pertinent to the subject of his present document. These references are commonly called citations, and a citation index is a structured list of all the citations in a given collection of documents. Such lists are usually arranged so that the cited document is followed by the citing documents.

The first practical application of this concept was *Shepard's Citations*, a legal reference tool that has been in use since 1873. *Shepard's Citations* owes its existence to the fact that American law, like English law, operates under the doctrine of *Stare Decisis*. *Stare Decisis* means that all courts must follow their own precedents as well as those established by higher courts. The precedents are the decisions handed down in previous cases.

To try a case under *Stare Decisis*, a lawyer must base his argument on previous decisions regarding a similar point of law. Before presenting the previous decision as a precedent, however, the lawyer must make sure that the decision has not been overruled, reversed, or limited in some way. *Shepard's Citations* enables the lawyer to do this with a minimum of trouble.

A legal case is always referred to by a code which consists of the volume and

Reprinted from:

Encyclopedia of Library and Information Science,
Volume 5, Pages 16-40, Marcel Dekker, Inc., New York (1971)

page number of the document in which the case is reported. Once a case is permanently reported, its reference code becomes fixed for all time. Thus, 301U.S.356 is a reference to the case reported in volume 301 of the United States Supreme Court Reports on page 356. Statutes are also referred to in a similar manner. Thus, Ch16Sec24NJRS refers to chapter 16, section 24 of the New Jersey Revised Statutes.

Taking advantage of this coding system, Frank Shepard devised a listing which shows every instance in which a reported decision is cited in a subsequent case. The listing also shows what statutes and journals cite the original decision.

Figure 1 is a Shepardized list of citations to a fictitious case. In the figure, the cited case is 101Mass.210; the items listed below it are the citing cases and an

CITED CASE → **101 Mass. 210**

	112	Mass.	65
a	130	Mass.	89
	165	Mass.	210
q	192	Mass.	69
	205	Mass.	113
o	221	Mass.	310
	281	U.S.	63
	35	H.L.R.	76

SUBSEQUENT CITING CASES

FIGURE 1. *Typical entry from* Shepard's Citations *showing cited and citing cases.*

article from the *Harvard Law Review* which refer to the case. The letters preceding the case codes show that the decision in the cited case was first affirmed (a), then questioned (q), and finally over-ruled (o).

To use *Shepard's Citations* a lawyer must first locate a previous decision relating to his current case. He does this by consulting a digest, index, or encyclopedia which will provide him with the case number for any given decision. The lawyer then looks up the case number in *Shepard's Citations* and finds all the subsequent citing cases. From this information, he can determine whether the original decision was affirmed or modified in any way. Thus, in the example given in Figure 1, the original decision could not be used as a precedent because it was subsequently over-ruled.

Problems with Traditional Subject Indexes

After World War II, users of scientific and technological literature were finding it increasingly difficult to find information pertinent to their own work. Several factors caused this situation. One was that the size of the literature was growing very rapidly. Projections indicated that by 1975 there would be over two million scientists in the world producing a million papers a year. These new papers would be added to the 10 million papers that would already be published by then (*1*).

As the growing volume of scientific information began to overwhelm the limited number of indexers that could be economically supported, there were delays of six months to several years before papers were classified. This, in turn, resulted in more and more scientists spending time needlessly duplicating existing work.

Another factor in the literature problems of scientists was the increasing need for exchanging information between scientific disciplines. The majority of subject indexes covered only one field or discipline. Science had become so interrelated that such arbitrary restrictions were causing the researcher to remain unaware of much valuable information. Fields such as oceanography, organic chemistry, and environmental science could no longer be placed into neat little cells. To be reasonably sure that he had most of the important information pertaining to any such field, the scientist was required to examine the literature of several disciplines. For example, a chemist selecting suitable materials for surgical implants or artificial internal organs might find useful information in chemical journals, medical journals, or engineering journals.

Although the growing size of the literature and the need for multidisciplinary information retrieval highlighted the shortcomings of traditional subject indexes, there were also other problems involved. The classification terms used in subject indexes are often ambiguous and lend themselves to different interpretations, especially when the user is not fully conversant with the details of a particular indexing system. Subject indexes also encounter the problem of assigning labels to new concepts. Many times, especially in fast-moving (or rapidly developing) fields like biochemistry, this is very difficult. Indeed, a consensus of what is the "proper" label for a concept may not be arrived at until some time after the original paper presenting the concept has been indexed under an inappropriate term.

Because indexers possess different intellectual abilities and technical skills, two different indexers will often use different key words, or headings, or subject terms when classifying the *same* document (*2*). Thus, it is not surprising to find *related* documents classified under entirely different subject headings with no clue to the searcher that this has happened. For example, an important 1963 paper on the topic of "seasonal variations in birth" (*3*) is indexed under the subject heading of "Periodicity" in the 1964 edition of *Index Medicus* (*4*). It is highly unlikely that anyone looking for information on seasonal variations in birth would ever think to look under "Periodicity" since it is quite a different concept than "Seasonal Variation."

These types of problems made clear the need for a system that would provide a unified index to the scientific literature that was current, free of semantic difficulties, and not dependent on the subject knowledge of indexers (*5*).

The Start of Citation Indexes for Science

REFERENCE TRADITION

As the doctrine of *Stare Decisis* provided the logic for *Shepard's Citations*, so did the "reference tradition" provide the rationale for citation indexes for science.

Scientific tradition requires that when a reputable scientist or technologist publishes an article, he should refer to earlier articles which relate to his theme. These references are supposed to identify those earlier researchers whose concepts, methods, apparatus, etc., inspired or were used by the author in developing his own article. Some specific reasons for using citations are as follows:

1. Paying homage to pioneers.
2. Giving credit for related work.
3. Identifying methodology, equipment, etc.
4. Providing background reading.
5. Correcting one's own work.
6. Correcting the work of others.
7. Criticizing previous work.
8. Substantiating claims.
9. Alerting researchers to forthcoming work.
10. Providing leads to poorly disseminated, poorly indexed, or uncited work.
11. Authenticating data and classes of fact—physical constants, etc.
12. Identifying original publications in which an idea or concept was discussed.
13. Identifying the original publication describing an eponymic concept or term as, e.g., Hodgkin's disease, Pareto's Law, Friedel-Crafts Reaction.
14. Disclaiming work or ideas of others.
15. Disputing priority claims of others.

In the early 1950s, the availability of this built-in system for linking scientific articles began to receive attention as the possible foundation of an indexing system for the scientific literature.

WELCH MEDICAL LIBRARY INDEXING PROJECT

In 1952, Dr. Chauncey Leake was chairman of the Committee of Consultants for the Study of Indexes to Medical Literature. This committee was supervising the Johns Hopkins Welch Medical Library Indexing Project which was sponsored by the Armed Forces Medical Library. Dr. Leake suggested that project workers should examine review articles in connection with their investigation of the problems with subject indexes to medical literature.

GARFIELD

This statement had considerable impact on Eugene Garfield, one of the Welch Project investigators. Garfield realized that nearly every sentence in a review article is supported by a citation to a previous work. Thus, a review article could really be considered a series of indexing statements. The problem then became one of transforming these statements into a consistent format that would be useful as an index.

ADAIR

In 1953, the Welch Project conducted a symposium, news of which was reportec in a Colorado newspaper. This article was read by William C. Adair, who wa a former vice president of the firm that produced *Shepard's Citations*. Adair wrote to the Welch Project and suggested that they consider the method employed by Shepard's as a possible indexing technique.

After examining *Shepard's Citations*, Garfield realized that the "citator" principle could provide a means of indexing review papers which could be extended to the scientific literature in general.

After the Welch Project ended, Garfield began graduate work in Library Science at Columbia University. During this period he continued correspondence with Adair and began writing a detailed article on citation indexes for scientific literature. The article was completed in 1954 and was edited and refereed by Professor Bentley Glass, who was then chairman of the Johns Hopkins Department of Genetics and on the editorial board of *Science*.

While his own article was awaiting publication, Garfield, who by then was an associate editor of *American Documentation*, suggested that Adair write a shorter article which would explain, in general terms, the operation of *Shepard's Citations*. The Adair article appeared in *American Documentation* in June of 1955 (*6*); Garfield's article appeared in *Science* in July of 1955 (*7*).

LEDERBERG

It was not until 1958, however, that the scientific community exhibited any specific interest in Garfield's idea. In that year, Professor Joshua Lederberg of Stanford University wrote to Garfield to inquire if any further work had been done on citation indexing. When informed of the financial problems involved in starting such a project, Lederberg suggested that Garfield should apply for a grant from the government.

GENETICS CITATION INDEX

In 1961, the National Institute of Health initiated a cooperative program with Garfield's Institute for Scientific Information (ISI) to prepare a citation index for the field of genetics (*8*). In addition to preparing the index, the program was to investigate and make recommendations on such general questions about citation indexes as:

1. Should there be a single citation index for all of science and technology, several rather broad ones, or many narrow ones, each focused on a single discipline?
2. In what ways is it possible to arrange a citation index (author, journal, etc.) and what way is the best for any given situation?
3. What techniques could be used for gathering the citation information?
4. Should books and technical reports be covered and to what degree?

Garfield soon recognized, however, that defining the genetics literature to be covered by a citation index would be quite difficult. Fine judgements would be required as to what was or was not genetics literature. At Garfield's suggestion, it was decided to undertake a comprehensive, interdisciplinary approach to preparing a citation index and then extract a genetics citation index from that base of information.

SCIENCE CITATION INDEX

The interdisciplinary data base was eventually used to produce the first *Science Citation Index* which was published in 1963. The first *SCI* covered the literature of the calendar year of 1961. It covered 613 journals, contained 1.4 million citations, and required five volumes. Nineteen per cent of the citations in the 1961 *SCI* data base were selected by special computerized procedures as "having to do with genetics" and were published separately as the *Genetics Citation Index*. The genetics index was also published in 1963 and was complete in one volume.

OTHER CITATION INDEXES

In addition to the *Science Citation Index* and the *Genetics Citation Index*, there have been several other efforts at compiling citation indexes or at using the principle of citation indexing in information retrieval systems. Most of these have been experimental in nature, extremely narrow in their coverage, or published on a one-time basis.

Some citation indexes provide coverage of the material published in just one journal. One of the earliest examples of this is the cumulative index to volumes 35 through 50 of the *Journal of the American Statistical Association*. This index was prepared with assistance from the Ford Foundation and was issued in 1959. In this index, both the cited and citing articles had to appear in the covered journal to be indexed. Another example of a citation index with single journal coverage is the one that appears in the cumulative index to volumes 1 through 31 of the *Annals of Mathematical Statistics*. Published on a one-time basis in 1962, this index is a listing of articles appearing in the *Annals* which is arranged by author and shows references to various abstracts of the article and other articles in the *Annals* which cite the original article.

Another citation index that was published only once is contained in the *Bibliography of Non-parametric Statistics*. Published in 1962, this index shows what items in the bibliography cite other items in the bibliography.

Since 1966, each monthly issue of the *Journal of Histochemistry and Cytochemistry* has contained a citation index to its own articles. This index is arranged by author and shows, for any article previously published in the journal, where and by whom the article has been cited in the preceding month in any of over 2200 other journals.*

An example of a citation index that covers more than one journal but is limited to a single field is the *Citation Index for Statistics and Probability* which is currently being produced by Dr. J. W. Tukey at Princeton University (9). This project

* This information is compiled from the Automatic Subject Citation Alert (ASCA) service of the Institute for Scientific Information.

was initiated in 1961 and is being conducted in cooperation with the National Science Foundation. The journals covered by this index are concerned with theoretical and methodological statistics. At the beginning of the project, about 50 journals were being completely covered, with another 75 journals covered on a selective basis. Currently, about 100 journals are fully covered and 150 are covered selectively. It is estimated that the first issue of this index will appear in 1971.

In 1968, the Shepard organization itself introduced *Shepard's Law Review Citations*. This new publication indexes 117 law reviews and periodicals and shows where any legal article written since 1947 has been cited in the covered journals from 1957 on.

Some citation indexes are prepared to test other aspects of citation indexing besides information retrieval. One of these was compiled by Ben-Ami Lipetz (*10*). As sources, Lipetz used four of the eight Russian physics journals published in English by the American Institute of Physics. Out of all the citations appearing in these journals, he included in his index only the citations to articles published in two heavily-used American physics journals. The resulting citation index was then distributed to a group of subscribers of the two American journals and an attempt was made to compare the frequency of use of the four Russian journals before and after the distribution of the index. The object was to see if increasing the user's awareness of relevant articles by the use of a citation index would influence their reading habits. Lipetz concluded that the index produced a measurable increase in the use of the Russian journals, although the increase was not large scale.

As a final example, special citation indexes can be prepared on demand through systems that directly connect the users with the data base, as is done with the Technical Information Project (TIP) at the Massachusetts Institute of Technology (*11,12*). TIP uses a time-sharing computer connected to remote consoles by telephone cables. The data base consists of the full bibliographies of articles from recent volumes of twenty-five physics journals. Special programs have been devised to enable a user to request a citation index to these articles that will meet his specifications. Thus, a user can obtain a citation index to all the articles, or to the articles from only one of the covered journals, or to the articles in a single volume of a covered journal.

Description and Use of a Citation Index

The multidisciplinary and interdisciplinary *Science Citation Index* is now used by the libraries of almost every major university in the United States. For all intents and purposes, it is the only citation index to the most used literature of science and technology. Because of this, the following detailed discussion on the compilation and use of a citation index will be based on the *SCI*. It is felt that the underlying principles and mechanics of citation indexing will be most clearly explained with this approach.

JOURNAL COVERAGE AND PREPARATION

The *Science Citation Index* provides an index to the contents of every issue published during a calendar year of approximately 2200 selected journals. Covered journals are considered *source* journals and the items they contain are called *source items*. All journals are indexed comprehensively to eliminate doubts as to whether or not a particular item is covered. All original articles and most other useful items in a journal are processed, including editorials, letters, and meetings. Ephemeral items such as advertisements and news notices are omitted. Although previously included, book reviews have not been used as source items since 1969.

Before keypunching, each source journal issue is edited and tagged to insure that all relevant data will be recorded. All foreign language titles are translated into English. All citations (footnoted or provided in a bibliography) are processed; where practical, citations are also extracted from the text. A separated punched card is prepared for every cited item appearing in every source item processed. For every source item, a set of punched cards containing the author(s), title, journal, etc. is also prepared. Each punched card is verified by direct comparison with the original journal.

Once the cards are punched and verified, the data are transferred from cards to magnetic tapes. During this process, the computer performs a unification routine which eliminates many errors from the original literature such as incorrect spellings in the names of cited authors and titles of cited publications (*13*).

The final data is formatted and further editing is done. Final printouts are then produced through the use of a computer-driven photo-composition machine. The final indexes are produced through photo-offset printing. Statistics are also tabulated in the course of obtaining the printouts.

REFERENCES

1. D. J. de S. Price, *Little Science, Big Science*, Columbia Univ. Press, New York, 1963, pp. 8–15.

2. C. P. Bourne, *Methods of Information Handling*, Wiley, New York, 1963, pp. 1–12.

3. K. Tietze, "On the Problems of Seasonal Fluctuations in the Dates of Birth and Conception," *Therapie der Gegenwart*, 102, 955–962 (1963).

4. *Index Medicus, 1964 (Cumulated)*, 5(IV), American Medical Association, 1965, p. S-2966.

5. E. Garfield, "A Unified Index to Science," in *Proceedings of the International Conference on Scientific Information*, Vol. 1, National Academy of Sciences–National Research Council, Washington, 1959, pp. 461–469.

6. W. C. Adair, "Citation Indexes for Scientific Literature?", *Amer. Doc.*, 6, 31–32 (1955).

7. E. Garfield, "Citation Indexes for Science," *Science*, 122(3159), 108–111 (1955).

8. I. H. Sher and E. Garfield, "The Genetics Citation Index Experiment," in *Proceedings of the American Documentation Institute, 26th Annual Meeting*, American Documentation Institute, Chicago, 1963, pp. 63–64.

9. J. W. Tukey, "Keeping Research in Contact with the Literature: Citation Indexes and Beyond," *J. Chem. Doc.*, 2, 34–37 (1962).

10. B.-A. Lipetz, "The Effect of a Citation Index on Literature Use by Physicists," in *Proceedings 1965 Congress of the International Federation for Documentation*, Spartan Books, Washington, 1966, pp. 107–117.

11. M. M. Kessler, "The MIT Technical Information Project," *Phys. Today*, 18(3), 28–36 (1965).

12. F. L. Alt and R. A. Kirsch, "Citation Searching and Bibliographic Coupling with Remote On-Line Computer Access," *J. Res. Nat. Bur. Stand., B*, 72B(1), 61–78 (1968).

13. I. H. Sher, E. Garfield, and A. W. Elias, "Control and Elimination of Errors in ISI Services," *J. Chem. Doc.*, 6(3), 132–135 (1966).

Essays of an Information Scientist, Vol:1, p.198-205, 1962-73 Current Contents, #13, June 30, 1971
Reprinted from: Encyclopedia of Library and Information Science, Vol:5, p.16-40, 1971 (Marcel Dekker, Inc., NY)

CITATION INDEXES

By
Melvin Weinstock, Senior Information Scientist
Institute for Scientific Information

PART II

FORMAT AND ARRANGEMENT

The *SCI* consists of three separate but related indexes. These are the *Citation Index*, the *Source Index*, and the *Permuterm Subject Index*. All three indexes making up the *SCI* are published quarterly for the first three quarters of the year. The indexes for the fourth quarter are incorporated in the annual cumulation for each index. Eleven volumes of about 1400 pages each were required for the annual cumulation for the 1969 *SCI* which contains 4 million citations extracted from about 341,000 source items.

The *Citation Index* is arranged alphabetically by cited author. An entry for a cited item (reference) contains the first author's name and initials, the year the cited item was published, and the name of the publication in which the cited item appeared along with its volume and page number. When there is more than one cited item for any author, these are arranged chronologically by cited year. The source items citing a particular reference work are arranged alphabetically by source author immediately under each reference line. The source item line contains the citing author's name, name of the publication in which the citing item appeared, and the publication year, volume, and page. There is also a coded symbol indicating whether the citing item was an article, abstract, editorial, etc. Cited items may be from any year in recorded history; citing items, however, are always from the current year. In the *Citation Index* only the *first* author is shown for the cited and citing items. The *Source Index* (discussed below) gives *all* authors for each citing item. Figure 2 shows part of a typical column from the *Citation Index* of the *SCI*.

A separate section of the *Citation Index* is used for anonymous items (no personal author specified for the cited work). These items are arranged alphabetically by the titles of the cited publications.

Another separate section within the *Citation Index* contains a *Patent Citation Index*. This is a listing of all patents (foreign and domestic) which have been

Reprinted from:

Encyclopedia of Library and Information Science,
Volume 5, Pages 16-40, Marcel Dekker, Inc., New York (1971)

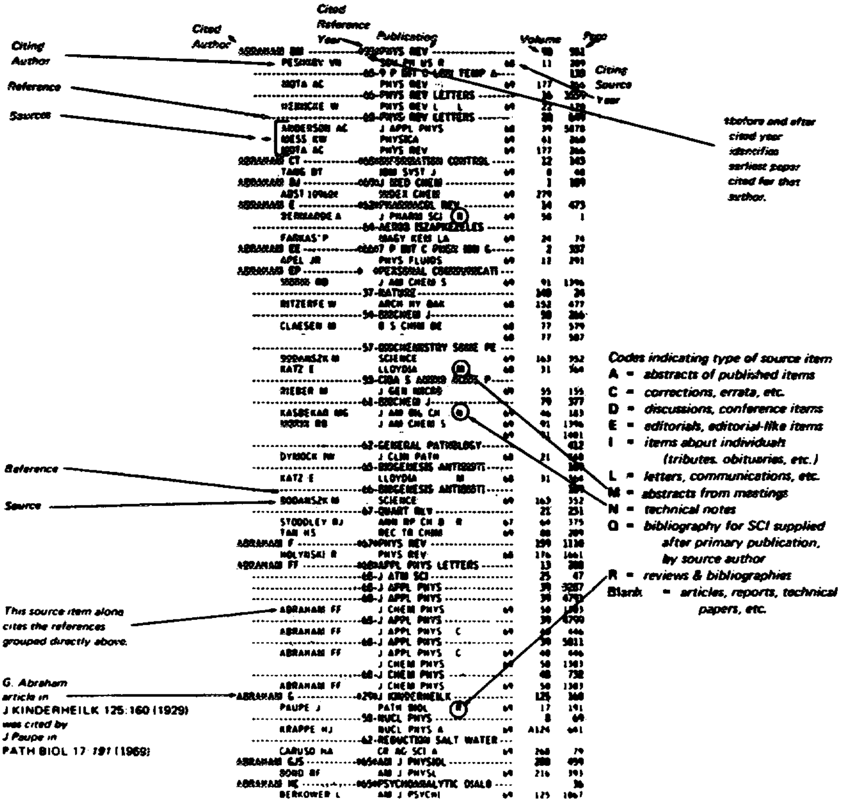

FIGURE 2. *Typical column from the* Citation Index *portion of the* Science Citation Index.

cited or referred to in any journal covered by the *SCI*. The *Patent Citation Index* is arranged in numerical order by patent number and usually provides, in addition to the patent number cited, the year of issuance, inventor, and country.

The *Source Index* is arranged alphabetically by source item author. Entries provide all co-authors, the full title of the citing (source) item, journal title, volume, issue, page, year, type of item (review, letter, correction, etc.), and number of references in the bibliography of the source item. Also provided is an

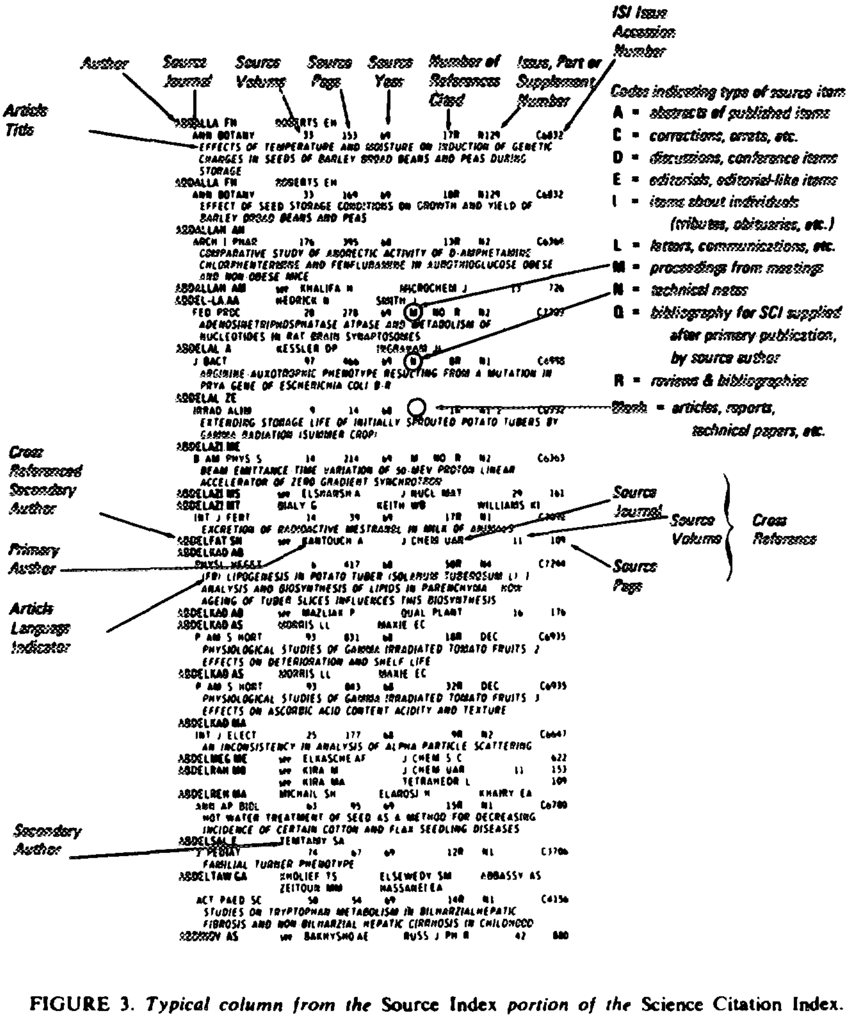

FIGURE 3. *Typical column from the* Source Index *portion of the* Science Citation Index.

" A CONFORMAL MAPPING METHOD TO PREDICT LOW-SPEED AERODYNAMIC CHARACTERISTICS OF ARBITRARY SLENDER RE-ENTRY SHAPES"

the Permuterm technique results in the following indexing entries:

PRIMARY TERM
- CO-TERM

AERODYNAMIC
- ARBITRARY
- CHARACTERISTICS
- CONFORMAL
- LOW-SPEED
- MAPPING
- METHOD
- PREDICT
- RE-ENTRY
- SHAPES
- SLENDER

ARBITRARY
- AERODYNAMIC
- CHARACTERISTICS
- CONFORMAL
- LOW-SPEED
- MAPPING
- METHOD
- PREDICT
- RE-ENTRY
- SHAPES
- SLENDER

CHARACTERISTICS
- SEE STOP LISTS

CONFORMAL
- AERODYNAMIC
- ARBITRARY
- CHARACTERISTICS
- LOW-SPEED
- MAPPING
- METHOD
- PREDICT
- RE-ENTRY
- SHAPES
- SLENDER

LOW-SPEED
- AERODYNAMIC
- ARBITRARY
- CHARACTERISTICS
- CONFORMAL
- MAPPING
- METHOD
- PREDICT
- RE-ENTRY
- SHAPES
- SLENDER

MAPPING
- AERODYNAMIC
- ARBITRARY
- CHARACTERISTICS
- CONFORMAL
- LOW-SPEED
- METHOD
- PREDICT
- RE-ENTRY
- SHAPES
- SLENDER

METHOD
- SEE STOP LISTS *

PREDICT
- AERODYNAMIC
- ARBITRARY
- CHARACTERISTICS
- CONFORMAL
- LOW-SPEED
- MAPPING
- METHOD
- RE-ENTRY
- SHAPES
- SLENDER

RE-ENTRY
- AERODYNAMIC
- ARBITRARY
- CHARACTERISTICS
- CONFORMAL
- LOW-SPEED
- MAPPING
- METHOD
- PREDICT
- SHAPES
- SLENDER

SHAPES
- AERODYNAMIC
- ARBITRARY
- CHARACTERISTICS
- CONFORMAL
- LOW-SPEED
- MAPPING
- METHOD
- PREDICT
- RE-ENTRY
- SLENDER

SLENDER
- AERODYNAMIC
- ARBITRARY
- CHARACTERISTICS
- CONFORMAL
- LOW-SPEED
- MAPPING
- METHOD
- PREDICT
- RE-ENTRY
- SHAPES

* No entries are created for the words "A", "TO", "OF". These illustrate "full stop" words and are not indexed. The words "METHOD" and "CHARACTERISTICS" illustrate "semi stop" terms. Semi stop words are suppressed as Primary Terms but do appear as Co Terms. Hyphenated words such as "RE ENTRY" or phrases such as "LOW SPEED" are treated as one term.

FIGURE 4. *Indexing entries that result when Permuterm technique is applied to an article title.*

accession number: this is the code by which the *source journal* is filed at *ISI*. Figure 3 shows part of a typical column from the *Source Index*.

Within the *Source Index* is a separate section called the *Corporate Index*. In the *Corporate Index*, all of the source items processed are listed alphabetically by author under the name of the organization where the work was performed. If more than one organization is involved in a given project, an entry is created for each organization.

The third major index contained within the *SCI* is the *Permuterm Subject Index*. Permuterm is a contraction of the phrase "permuted terms." In the *PSI*, the term "permuted" is used in its correct mathematical sense. This is to be distinguished from a Key-Word-In-Context (KWIC) index which rotates the words in an article title rather than fully permuting them.

To produce the *PSI*, a computer is used to permute all significant words within each title and subtitle of every item included in the *Source Index*. All possible pairs of terms are formed. Thus, for a title containing n significant words, there will be $n(n-1)$ pairs. With this system, every significant word takes a turn at being the primary term *as well as* being a co-term (*14*). Figure 4 shows the indexing entries that result when the Permuterm technique is used.

The *PSI* is arranged alphabetically by primary term. Terms which begin with numbers appear at the end of the index. All co-terms co-occurring with a particular primary term are indented and listed in alphabetical order under that primary term. Co-terms beginning with numbers appear at the end of the list. Dashed lines lead from each co-term to the name and initials of the author whose item contains that co-term and its associated primary term. For anonymous entries, the journal title is given in place of the author's name. Figure 5 shows part of a typical column from the *Permuterm Subject Index*.

BASIC SEARCH TECHNIQUE

Using the *Science Citation Index* involves the following steps. The searcher starts with the name of an author he has identified as having written an item related to the topic of the search. He then enters the *Citation Index* and looks up the name of that author. Once the author's name is located, the searcher can see the items that have currently cited the various items listed for this author. The searcher then notes the author, journal, volume, and page of each citing item. The searcher then turns to the *Source Index* and looks up the name of the citing author. At this entry he will find the complete bibliographic data for the citing item including the complete title and all co-authors. At this point, the searcher should examine source item titles and select those items which seem most likely to be relevant to his topic. He can then obtain the journals containing the items of interest from the library. This basic search technique is illustrated in Figure 6.

A search may be readily expanded in order to build a more extensive bibliography for a particular inquiry. For example, once he finds a number of source items, the searcher can use the bibliographies of one or several of these to provide

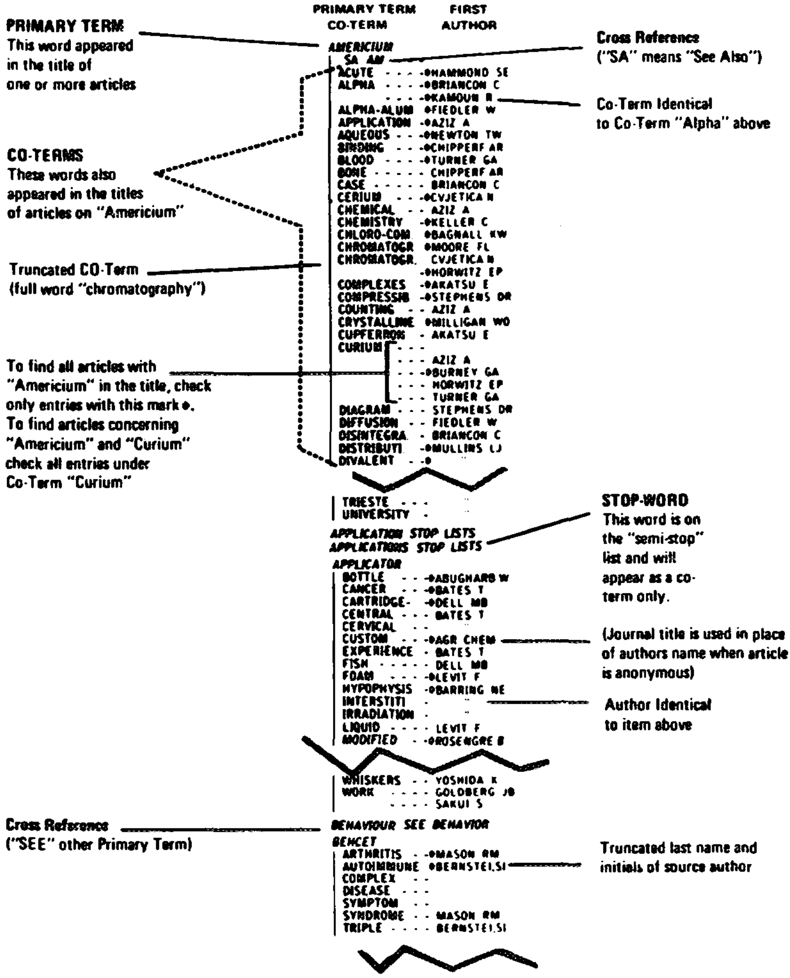

FIGURE 5. *Typical column from the* Permuterm Subject Index *portion of the* Science Citation Index.

the names of other authors to look up in the *Citation Index;* this process is called "cycling." Figure 7 is a diagram of the basic cycling procedure. [More sophisticated cycling procedures exist, but are not discussed here (*15*).]

Examination of the *Source Index* itself may yield additional relevant current items by a given source author, even though they may not cite any of the known starting references.

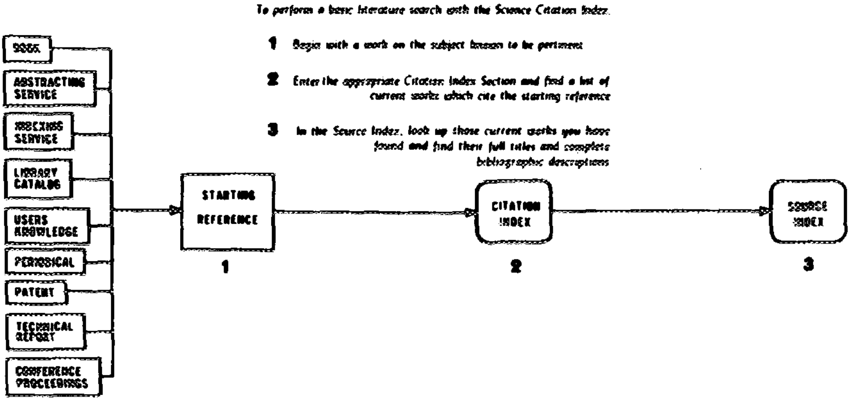

FIGURE 6. *Basic search technique for the* Science Citation Index.

The *Permuterm Subject Index.* is used when the searcher does not know a specific author of interest by which he can enter the *Citation Index* or the *Source Index*. To use the *PSI*, the searcher first compiles a list of terms that are likely to describe his topic of interest. The searcher then enters the *PSI* and locates a primary term that is the same as one of the terms he has listed. He then sees all the authors in the current year that have used that term in the title of an item. Usually, the searcher will want to be more selective. He can accomplish this by locating a co-term under the primary term that further defines his topic. He can then select only those authors that have used the co-term as well as the primary term. Once the appropriate author(s) is identified, the searcher can enter the *Source Index* to obtain the full title of an article along with other bibliographic data. The searcher can then obtain desired items and/or look up appropriate items in the *Citation Index* for subsequent citing sources. Figure 8 diagrams the use of the *PSI* to obtain a starting author to conduct an *SCI* search.

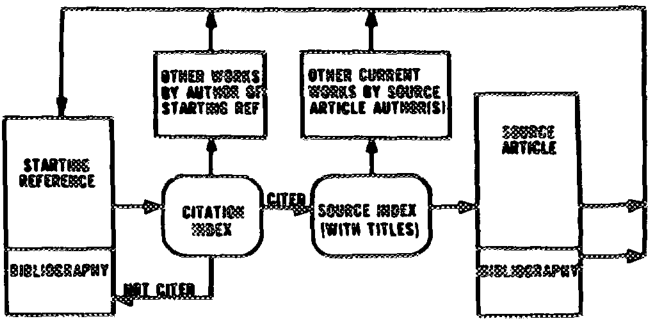

FIGURE 7. *Use of "cycling" in basic search technique for the* Science Citation Index.

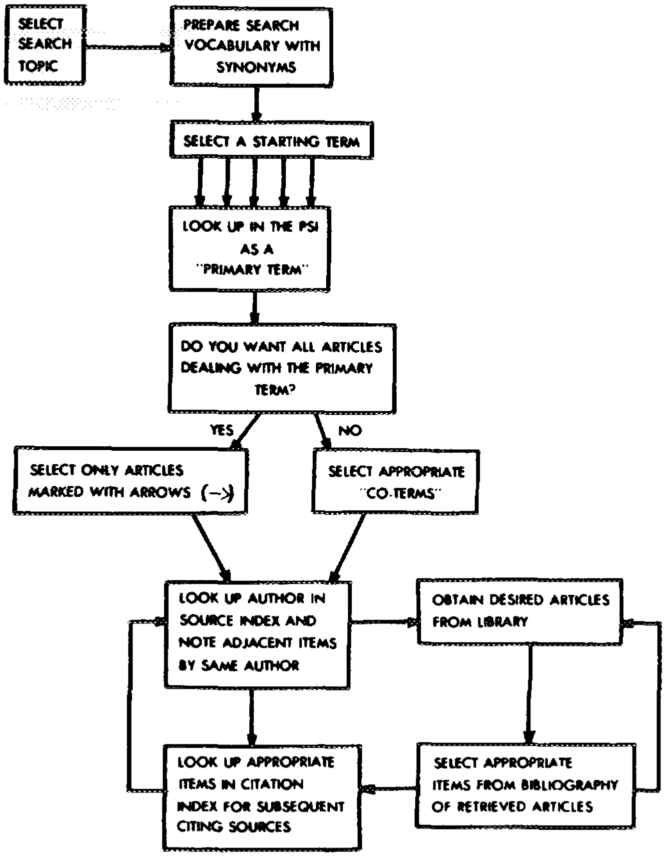

FIGURE 8. *Use of the* Permuterm Subject Index *to obtain a starting author to conduct a* Science Citation Index search.

REFERENCES

14. E. Garfield, "Citation Indexing, Historio-Bibliography, and the Sociology of Science," in *Proceedings of the III International Congress of Medical Librarianship*, Excerpta-Medica Foundation, Amsterdam (1969).

15. A. E. Cawkell, "Search Strategies Using the *Science Citation Index*," in *Computer Based Information Retrieval Systems* (B. Houghton, ed.), Bingley, London, 1968, pp. 27–44. See also: "Current Comments," *Current Contents Life Sciences*, 12(44), 89–103 (1969); "Current Comments," *Current Contents Physical Sciences*, 9(44), 89–103 (1969); "Current Comments," *Current Contents Chemical Sciences*, 3(45), 49–63 (1969); "Current Comments," *Current Contents Education*, 1(41), 33–47 (1969); and "Current Comments," *Current Contents Behavioral, Social & Management Sciences*, 1(34), 41–56 (1969).

(Part III of this article will appear in the next issue of *Current Contents®*.)

Essays of an Information Scientist, Vol:1, p.207-216, 1962-73 Current Contents,

CITATION INDEXES

By
Melvin Weinstock Senior Information Scientist
Institute for Scientific Information

PART III

Evaluation of Citation Indexing

Perhaps the best way to evaluate citation indexing is to examine how the *Science Citation* Index resolves the previously discussed deficiencies of subject indexes. For purposes of review, the deficiencies can generally be summarized as follows:

1. Inability to deal comprehensively with the growing volume of scientific literature on a timely basis.
2. Limited ability to cut across disciplines to pull together related information.
3. Semantic difficulties in preparation and use of the indexes.

COMPREHENSIVENESS AND TIMELINESS

The *SCI's* method of obtaining comprehensive coverage of the literature has its foundation in Bradford's Law (16.17). In general, this law states that a small percentage of journals account for a large percentage of the significant articles in any given field of science. In support of this, a study of Physics *Abstracts* by Keenan and Atherton (18) shows that 50% of the items abstracted are taken from only nineteen journals. Also, studies conducted by *Index Chemicus* show that 100 journals account for 98% of all new articles in synthetic chemistry (19).

Further analyses have revealed that this concentration of information in relatively few journals is characteristic, not only of the individual disciplines, but of the scientific literature as a whole. Yale Professor Derek J. de Solla Price claims that about 1000 journals contain 80% of *all* scientific articles (20). This estimate is confirmed by continuing *ISI* studies. These same studies show that fewer than 1000 journals account for 90% of the *significant* literature, that is, they are the most heavily cited journals (21).

These findings lead to the reasoning that if the 2200 journals covered by the *SCI* are properly chosen, most of the world's important scientific literature will

Reprinted from:

Encyclopedia of Library end Information Science,
Volume 5. Pages 16-40, Marcel Dekker, Inc.. New York (1971)

be indexed even though there are an estimated 30 to 50 thousand journals in existence.

The publisher of the *SCI* uses several methods to make sure that covered journals are, in fact, the significant ones. First, he has enlisted an editorial board composed of experts in the various disciplines to recommend journals for coverage. Second, subscribers are invited to suggest journals for coverage; such suggestions are then evaluated by the editorial board. Third, large-scale citation analyses are made to see which journals are cited most frequently. This information is especially helpful in determining those journals that are the most used in emerging fields of science. The coverage rationale that has evolved from this system places heavy emphasis on the multidisciplinary journals, supplementing these with the most important journals from the individual disciplines.

There has been some criticism that the *SCI* is biased in favor of covering the Western-language journals. This is, in part, a reflection of the superabundance of research conducted in the United States and abroad which is published in English, German, and French. This bias is also due to the fact that the *SCI* has, in the past, given preference to cover-to-cover English translations of Russian journals. However, it should be noted that the SCI is the only index that covers every article in the Soviet journal *Doklady Akad. Nauk SSSR*, which ranks as the fifth largest journal in the world in terms of articles published each year.

Another Russian journal covered by *SCI* is *Teploenergetica* (Thermal Engineering). It must be admitted, however, that this journal was not covered until 1968 when *ISI* studies showed that it was among the 500 most cited journals in the world. This is typical of the continuous improvement in journal coverage of *SCI*.

The *SCI* provides timely coverage partly because the type of intellectual activity required to compile traditional subject indexes is not required. The author himself "indexes" his article (by way of his citations), enabling the *SCI* to be prepared by a combination of man-machine procedures which facilitate current coverage of the literature. Thus, indexes covering the literature appearing in any calendar quarter are published within sixty days after the end of the quarter; hard-bound annual cumulations are published within four months after the end of the year.

MULTIDISCIPLINARY SEARCHING CAPABILITY

The reason citation indexes provide multidisciplinary searching capabilities is, once again, related to the fact that most indexers are not as qualified as the author himself to decide which previously published material is related to his current work. A citation index takes advantage of the built-in linkages between documents provided by authors' citations by listing together *all* items with common citations.

It is this unique ability to group together items that are often seemingly unrelated that is so important to the modern researcher. Everyone knows that there are important, though small parts of the literature which can be called "pure" physics or "pure" chemistry, etc. There is, however, a larger, less specialized part of the literature that is of interest to physicists, chemists, or other scientists as it relates to their specialties. For example, the chemistry of water is pertinent to oceanography, but it is also pertinent to a vast array of other problems in biology, physics, chemistry, and other applied fields. With the *SCI*, as long as a current *item cites a given previous item, it will be indexed under the cited item.* It makes no difference if the citing item appeared in a physics journal, a chemistry journal, an engineering journal, or any other type of journal. Therefore, a searcher using the *SCI* can identify a group of items whose contents are in some way related to his topic, but which were published in a variety of journals not normally considered as being related to his discipline.

An interesting example of the utility of the *SCI* in crossing scientific disciplines to retrieve isolated bits of information is found in the relationship between C. H. Whitnah's paper in the *Journal of Dairy Science* in 1959 and a paper by Albert Einstein in *Annalen der Physik* in 1906. This apparently incongruous combination proves to be a legitimate reference by Whitnah to an equation used in calculating molecular dimensions which was applied in a study of the physical properties of milk.

This same article by Einstein was cited in a 1960 article by V. V. Varadaiah in the *Journal of Polymer Science*. In this article, Einstein's equation was used as a basis for calculations relating to the Flory universal constant in the equation for intrinsic viscosity. In two other papers by P. H. Elworthy, one in 1959 in the *Journal of the Chemical Society*, and one in 1961 in the *Journal of Pharmacy and Pharmacology*, the Einstein equation was cited in a discussion on the size and shape of lecithin micelles.

In a 1961 paper by K. Yagi in *Comparative Biochemistry and Physiology*, the Einstein equation is employed in the study of mechanical and colloidal properties of amoeba protoplasm. Still again in 1961, S. G. Schultz, in the *Journal of General Physiology*, reported biophysical studies in which he used the Einstein viscosity equation to confirm atomic dimensions compiled by L. Pauling. In the *SCI*, each of these widely scattered papers would be retrieved by the use of basic search and cycling techniques.

SEMANTIC PROBLEMS

Citation indexes resolve semantic problems associated with traditional subject indexes by using citation symbology rather than words to describe the content of a document. This concept is rather difficult for most people to comprehend at first. Therefore, the following rather extended example is presented as an aid to understanding.

In 1963, Professor J. Lederberg published a paper in *Nature* entitled "Molecular Biology, Eugenics and Euphenics." In this paper, he established the word "euphenics" as a synonym for the concept of "engineering human development." As long as this paper was the only one in the literature on euphenics, there was effectively a one-to-one equivalence between the word "euphenics" and the citation which identified the document in which it first appeared. The word "euphenics" and the citation "Lederberg J., 63, *Nature* 198, 428" were essentially equivalent symbols for the subject discussed in Lederberg's paper.

Now suppose that other authors use the term "euphenics" in subsequent papers. Customarily, the subsequent authors will give credit to Lederberg as originator of the term by citing his original paper. As a result, in a citation indexing system, the new papers would be automatically grouped together under the citation "Lederberg J.. 63, *Nature* 198, 428." In a word indexing system, the subsequent papers would be grouped together under the term "euphenics." Both methods would achieve the same objective—to make information on "euphenics" retrievable. In one system, the word is the indexing term; in the other, the citation is the indexing term.

Once it is understood how a citation can serve as an indexing term, it is not difficult to show why citations are frequently better than words in this role.

Carrying the Lederberg example further, suppose that another author discusses "engineering human development" but does not mention the word "euphenics." As long as the author cites the original paper by Lederberg (which is highly probable), the new paper will be indexed under Lederberg's paper in a citation index. The odds are very slim, however, that any subject indexer would equate "engineering human development" with "euphenics." Thus, the paper that does not specifically mention euphenics has a high probability of being indexed under some other term.

Consider the same situation from the point of view of the user of an index. If the searcher is familiar with the term "euphenics," word indexes will enable him to find the Lederberg article and the subsequent articles that specifically mention "euphenics." The searcher will, however, most likely miss the papers on "engineering human development" unless he is aware that this phrase is an alternate for the word "euphenics." With the *SCI*, the searcher only needs to know that Lederberg had published on this general topic. By simply looking up the name Lederberg, he will find the original paper plus all subsequent citing papers, whether or not they specifically mention "euphenics." This is especially useful to a searcher who is not familiar with the jargon of a different discipline than his own.

Grouping items by a common citation also makes the *SCI* a self-organizing indexing system that is constantly being upgraded by the feedback of more current information.

UNIQUE CAPABILITIES OF CITATION INDEXES

Citation indexes not only resolve many of the difficulties inherent in conventional indexes, certain things can be accomplished with citation indexes that are not at all feasible with other indexes.

Probably the most important of these capabilities is the ability to bring the searcher forward in time from an earlier known reference. The *SCI* is set up so that no matter when an item originally appeared, it will be indexed in the *Citation Index* as long as it is cited at least once in the current year in any of the covered journals. As soon as the searcher locates his starting "cited item" he is brought forward to items that are currently citing the original. This could bridge a gap of fifty years or more (as in the Einstein article discussed above), or it can take the searcher forward in increments as small as a year (say from 1968 for the cited article to 1969 for the citing article).

By utilizing this ability of citation indexes, necessary research questions such as these can be answered:

1. Has this basic concept been applied elsewhere?
2. Has this theory been confirmed?
3. Has this method been improved?
4. Is there a new synthesis for this old compound?
5. Have there been errata or correction notes published for this paper?

Also, any scientist may legitimately wish to determine whether his own work has been applied or criticized by others. Citation indexes facilitate this type of feedback in the communication cycle. A further use of citation indexes is to quickly identify scientists currently working in special branches of science for the purpose of correspondence or personnel selection.

Finally, a mention should be made about the unusual ability of citation indexes to serve as a tool in evaluating literary practices and the structure of scientific literature (*22–24*). Using citation data, networks of interconnected articles may be constructed to trace the history of a subject (*25,26*). Citation counts can also be used to determine the length of time that there is any interest in a given article or topic (*27*). The impact of individual articles as well as the emergence of "superclassic" papers can be studied with citation data (*28*).

FORMAL STUDIES OF CITATION INDEXING

A number of formal studies have been conducted on citation indexing, almost all of which have been based on the *SCI*. Barlup (*29*) describes a study to test allegations that *SCI* searches are "noisy," that is, that they retrieve a high percentage of irrelevant material. In this study, searches were conducted for a range of medical subjects. A team of physicians was used to assess relevance. It was found that 72% of the citing articles located were "closely or directly related in subject content to the cited article." About 22% were "slightly or indirectly related" and about 5% could be considered noise. Of the article found to be directly related to the cited article, about 10% were judged to have titles that did not indicate any relationship.

Spencer (*30*), Rieger (*31*), and Ghosh (*32*) have conducted studies that were mainly concerned with the comprehensiveness and/or the search speed of the *SCI* as compared with discipline-oriented indexes such as *Chemical Abstracts* and *Index Medicus.* On a topic that was clearly within the specialized field covered by *Index Medicus,* Spencer found that the SCI produced better results. Reiger found that the *SCI* was less efficient than *Index Medicus* on a subject that was primarily covered in Italian journals. On the other hand, Ghosh used the *SCI* to conduct a search on "hemorrhagic fever," a narrow field in which articles are almost entirely confined to Indian journals. In this case, the *SCI* produced a high retrieval efficiency.

Future Improvements and Applications

Several aspects of citation indexing require attention if it is to deliver its full potential in the future. Included in these are improvements in the mechanics of the system itself as well as the conception, investigation, and development of new applications.

CITATION PRACTICES OF AUTHORS

One of the most obvious areas for improvement in citation indexing systems is the citation practices of authors themselves. Some scientific articles have hundreds of references; others have none at all. Part of the reason for such disparity in the number of references is the great difference in quality, not only in articles, but also in the journals that publish them. Many authors, editors, and referees are quite meticulous in ensuring that an article includes a comprehensive set of references. For some articles, especially in the less scholarly journals, the references may be inadequate or nonexistent. Information scientists and others have discussed this problem at length, although most of their suggestions are aimed at improving the author's awareness of the value of good citation habits (*33*). Garfield went a step further and discussed the possibilities of an automatic system in which a computer could read an article and determine not only if the references provided are appropriate, but also what references are missing (*34*). Needless to say, such a system is not on the immediate horizon.

STORAGE REQUIREMENTS

A problem that confronts all types of indexes is the growing amount of space required to store the cumulated volumes. One obvious way to approach this is the use of microforms. The improving technology in this area has resulted in increased data storage capabilities through higher reduction ratios and larger magnifications in the optical system of the viewer. An optimized indexing system of the future may make use of remote access to time-shared computers for the current year's indexed material, printed books for three to four year cumulative indexes, and microforms for very large scale cumulative indexes (*35*).

Of course, putting large-scale cumulative indexes on microforms presupposes that such citation indexes are available. To date, this is not the case. It is, however, *ISI*'s announced goal to produce a citation index that will provide total retrospective coverage of the literature of the twentieth century. *ISI*'s plan is to produce this index in stages over the next decade.

ON-LINE ACCESS TO INDEXED DATA

As indicated, one of the logical developments for citation indexing would be to provide remote access to the indexed data. This would be similar to what is provided on a limited basis by the previously discussed project TIP. With such a system, a searcher would be able to sit before a computer console and do his bibliographic research by operating a keyboard linked to a computer with several billion stored characters. The required data would then be automatically printed out on a typewriter or teletype unit or displayed on a cathode ray tube. A perforated tape or set of punched cards could even be produced for use as input to another data processing system which could further refine or analyze the data obtained.

Surprisingly, the main deterrent to implementing such a system on a large scale is not the high cost of the computer. The cost of the computer would be shared by all the users in the network. Each subscriber, however, would have to bear the entire cost of the long distance telephone call required to access the computer. The telephone system is not yet organized so that you pay for the call only during the time you are actually using the computer; you also pay for waiting time. A time-shared computer works for someone else during the silent or thinking period of any one subscriber. You pay for the entire telephone call whether you are using the computer for the whole period or not.

EVALUATION OF SCIENTIFIC PERSONNEL

Although the *SCI* was originally designed to be a retrieval tool for use in library and information science work, there are indications that it will have important applications as a tool for evaluating scientific personnel. By using the *SCI* data base, it is possible to count the number of citations to a given author. Although there are exceptions, frequently cited authors are usually those who have done the most important work in a given field (*36–39*).

For example, by using the *SCI* data base, it was possible to list the fifty most cited authors for 1967. Two of the 1969 Nobelists—Derek H. R. Barton and Murray Gell-Mann appeared on the list (*40*). Since there are over a million scientists in the world's population, producing a list of fifty that contains two Nobel prize winners is no small achievement. This is especially impressive since the list was compiled by a purely mechanical method which did not require reading the works of these men.

The ability of the citation index to measure the impact of a scientist's work has practical economic consequences. Research administrators could use such a tool as an aid in evaluating present scientific personnel or in hiring new people. Officers of various foundations could use it in awarding prizes, grants, fellowships, and other forms of research assistance.

MARKETING RESEARCH

Another possible use for the data used to compile the *SCI* is in marketing research. In the *Corporate Index* section of the *SCI*, new papers are listed under the companies where the work was performed. Proper analysis of the published information could give a good indication of the type of equipment or supplies needed by a company to conduct its research work. This could be especially valuable information, for example, to a scientific instrument manufacturer trying to anticipate the needs of potential or existing customers.

Conclusion

When citation indexes for scientific literature were first introduced, they were considered supplements to traditional subject indexing methods (*41*). Time has made it clear, however, that citation indexes that are comprehensive and timely are entitled to be considered as independent, fully integrated, library and information science tools. Further, citation indexes can now perform important evaluative, analytical, and predictive roles that were never imagined for subject indexes.

It seems likely, then, that given the right amount of attention and constructive criticism, citation indexing will continue to grow in usefulness and acceptance in the scientific, academic, and industrial communities.

REFERENCES

1. D. J. de S. Price, *Little Science, Big Science*, Columbia Univ. Press, New York, 1963, pp. 8–15.

2. C. P. Bourne, *Methods of Information Handling*, Wiley, New York, 1963, pp. 1–12.

3. K. Tietze, "On the Problems of Seasonal Fluctuations in the Dates of Birth and Conception," *Therapie der Gegenwart*, 102, 955–962 (1963).

4. *Index Medicus, 1964 (Cumulated)*, 5(IV), American Medical Association, 1965, p. S-2966.

5. E. Garfield, "A Unified Index to Science," in *Proceedings of the International Conference on Scientific Information*, Vol. 1, National Academy of Sciences–National Research Council, Washington, 1959, pp. 461–469.

6. W. C. Adair, "Citation Indexes for Scientific Literature?", *Amer. Doc.*, 6, 31–32 (1955).

7. E. Garfield, "Citation Indexes for Science," *Science*, 122(3159), 108–111 (1955).

8. I. H. Sher and E. Garfield, "The Genetics Citation Index Experiment," in *Proceedings of the American Documentation Institute, 26th Annual Meeting*, American Documentation Institute, Chicago, 1963, pp. 63–64.

9. J. W. Tukey, "Keeping Research in Contact with the Literature: Citation Indexes and Beyond," *J. Chem. Doc.*, 2, 34–37 (1962).

10. B.-A. Lipetz, "The Effect of a Citation Index on Literature Use by Physicists," in *Proceedings 1965 Congress of the International Federation for Documentation*, Spartan Books, Washington, 1966, pp. 107–117.

11. M. M. Kessler, "The MIT Technical Information Project," *Phys. Today*, 18(3), 28–36 (1965).

12. F. L. Alt and R. A. Kirsch, "Citation Searching and Bibliographic Coupling with Remote On-Line Computer Access," *J. Res. Nat. Bur. Stand., B*, 72B(1), 61–78 (1968).

13. I. H. Sher, E. Garfield, and A. W. Elias, "Control and Elimination of Errors in ISI Services," *J. Chem. Doc.*, 6(3), 132–135 (1966).

14. E. Garfield, "Citation Indexing, Historio-Bibliography, and the Sociology of Science," in *Proceedings of the III International Congress of Medical Librarianship*, Excerpta-Medica Foundation, Amsterdam (1969).

15. A. E. Cawkell, "Search Strategies Using the *Science Citation Index*," in *Computer Based Information Retrieval Systems* (B. Houghton, ed.), Bingley, London, 1968, pp. 27–44. See also: "Current Comments," *Current Contents Life Sciences*, 12(44), 89–103 (1969); "Current Comments," *Current Contents Physical Sciences*, 9(44), 89–103 (1969); "Current Comments," *Current Contents Chemical Sciences*, 3(45), 49–63 (1969); "Current Comments," *Current Contents Education*, 1(41), 33–47 (1969); and "Current Comments," *Current Contents Behavioral, Social & Management Sciences*, 1(34), 41–56 (1969).

16. S. C. Bradford, "Complete Documentation," in *Report of the Royal Society Empire Scientific Conference*, London, 1946, pp. 729–748.
17. B. C. Brooks, "Bradford's Law and the Bibliography of Science," *Nature*, 224(5223), 953–956 (1969).
18. S. Keenan and P. Atherton, *The Journal Literature of Physics*, Report No. AIP/DRP PA1, American Institute of Physics, New York, 1964, 156 pp.
19. E Garfield and G. Foeman, "Statistical Analyses of International Chemical Research by Individual Chemists, Languages, and Countries," paper presented at the American Chemical Society, Division of Chemical Literature Meeting, Chicago, 1964.
20. D. J. de S. Price, "Is Technology Historically Independent of Science? A Study in Statistical Historiography," *Technol. Culture*, VI(4), 553–568 (1965).
21. ISI intends to release these and other data during 1970 in the form of a *Journal Citation Index* (*JCI*). The *JCI* will show how often each of 2200 journals cite each other as well as how often they cite any of the estimated 25,000 other journals in existence.
22. E. Garfield and I. H. Sher, "New Factors in the Evaluation of Scientific Literature through Citation Indexing," *Amer. Doc.*, 14(3), 195–201 (1963).
23. E. B. Parker, W. J. Paisley, and R. Garrett, "Bibliographic Citations as Unobtrusive Measures of Scientific Communication," Institute for Communication Research, Stanford University, Stanford (1967).
24. R. D. Whitley, "Communication Nets in Science: Status and Citation Patterns in Animal Physiology," *Sociological Rev.*, 17(2), 219–233 (1969).
25. E. Garfield, I. H. Sher, and R. J. Torpie, "The Use of Citation Data in Writing the History of Science," Institute for Scientific Information, Philadelphia, 1964, 76 pp.
26. E. Garfield and I. H. Sher, "Diagonal Display—A New Technique for Graphic Representation of Complex Topological Networks," Final report to U.S. Air Force, Office of Scientific Research under Contract AF 49(638)-1547, Institute for Scientific Information, Philadelphia, 1967, 94 pp. See also: E. Garfield and M. V. Malin, "Diagonal Display—A New Technique for Graphic Representation of Network Diagrams," in *1969 Transactions of the American Association of Cost Engineers, 13th National Meeting*, Pittsburgh, 1969, American Association of Cost Engineers, Alabama University, pp. 222–232.
27. J. Margolis, "Citation Indexing and Evaluation of Scientific Papers," *Science*, 155(3767), 1213–1219 (1967).
28. D. J. de S. Price, "Networks of Scientific Papers," *Science*, 149(3683), 510–515 (1965).
29. J. Barlup, "Mechanization of Library Procedures in the Medium-Sized Medical Library (Relevancy of Cited Articles in Citation Indexing)," *Bull. Med. Lib. Assoc.*, 57(3), 260–263 (1969).
30. C. C. Spencer, "Subject Searching with *Science Citation Index:* Preparation of a Drug Bibliography Using *Chemical Abstracts, Index Medicus*, and *Science Citation Index*," *Amer. Doc.*, 18(2), 87–96 (1967).
31. L. A. Rieger, "Subject Searching with the *Science Citation Index, Chemical Abstracts*, and *Index Medicus* in the Preparation of a Drug Bibliography," paper presented at a symposium of the Pharmaceutical Division of the Special Libraries Association, Los Angeles, 1968. Submitted for publication in *Amer. Doc.*
32. J. S. Ghosh, "Literature Searching with the *Science Citation Index;* an Experimental Study," *Ann. Lib. Sci. Doc.*, 14(3), 133–142 (1967).
33. D. J. de S. Price, "On the Side of Citations," letter, *Agr. Eng.*, 5(2), 94 (1970).
34. E. Garfield, "Can Citation Indexing Be Automated?" in *Statistical Association Methods for Mechanized Documentation: Symposium Proceedings* (M. E. Stevens, V. E. Giuliano, and L. B. Heilprin, eds.), National Bureau of Standards, Washington, D.C., Miscellaneous Publication No. 269, 1965, pp. 189–192.
35. E. Garfield, "World Brain or Memex? Mechanical and Intellectual Requirements for Universal Bibliographic Control," in *The Foundations of Access to Knowledge* (E. B. Montgomery, ed.), Syracuse Univ. Press, Syracuse, 1968, pp. 169–196.
36. A. E. Bayer and J. Folger, "Some Correlates of a Citation Measure of Productivity in Science," *Sociology of Educ.*, 39(4), 382–390 (1966).

37. S. Cole and J. R. Cole, "Scientific Output and Recognition: A Study in the Operation of the Reward System in Science," *Amer. Sociological Rev.*, **32**(3), 377–390 (1967).
38. J. P. Martino, "Research Evaluation Through Citation Indexing," *AFOSR Research*, AD 659 366 (D. Taylor, ed.), USAF Office of Aerospace Research, Arlington, 1967, pp. 226–227.
39. E. Garfield, "Citation Indexing for Studying Science," *Nature*, 227, 669–671 (1970).
40. E. Garfield and M. V. Malin, "Can Nobel Prize Winners be Predicted?" Paper presented at 135th Meeting, AAAS, Dallas (1968).
41. E. Garfield, "*Science Citation Index*—A New Dimension in Indexing," *Science*, **144**(3619), 649–654 (1964).

BIBLIOGRAPHY

Abelson, P. H., "Coping with the Information Explosion," Editorial, *Science*, **154**(3745), 75 (1966).

Bottle, R. T., "Abstracts, Reviews and Bibliographies as Keys to the Literature," in *The Use of Biological Literature* (R. T. Bottle and H. V. Wyatt, eds.). Archon Books, Butterworth, London, 1967, pp. 33–57.

Cawkell, A. E., "Using References to Retrieve Current Articles," *Radio Electronic Eng.*, **35**(6), 352–353 (1968).

Dorr, H. A., and I. H. Sher, "The Science Citation Index System and Pharmaceutical Education," *Amer. J. Pharm. Educ.*, **32**(2), 177–188 (1968).

Garfield, E., "Breaking the Subject Index Barrier—a Citation Index for Chemical Patents," *J. Patent Office Soc.*, **39**(8), 583–595 (1957).

Garfield, E., "Citation Indexing: a Natural Science Literature Retrieval System for the Social Sciences," *Amer. Behavioral Sci.*, **7**(10), 58–61 (1964).

Garfield, E., "Science Citation Index—Answers to Frequently Asked Questions," *Rev. Int. Doc.*, **32**(3), 112–116 (1965).

Garfield, E., "Patent Citation Indexing and the Notions of Novelty, Similarity, and Relevance," *J. Chem. Doc.*, **6**(2), 63–65 (1966).

Garfield, E., "The Role of the Medical Librarian in SDI Systems," *Bull. Med. Lib. Assoc.*, **57**(4), 348–351 (1969).

Giuliano, V. E., "Analog Networks for Word Association," *IEEE Trans. Military Electronics*, MIL-7(2&3), 221–234 (1963).

Hayman, M., and R. Peskin, "Need of the Research-Oriented Psychiatrist for Information Retrieval," *Diseases Nervous System*, **28**(12), 798–803 (1967).

Malin, M. V., "The *Science Citation Index*: A new Concept in Indexing," *Lib. Trends*, **16**(3), 374–387 (1968).

Martyn, J., "An Examination of Citation Indexes," *Aslib Proc.*, **17**(6), 184–196 (1965).

Martyn, J., "Citation Indexing," *The Indexer*, **5**(1), 5–15 (1966).

Sher, I. H., and E. Garfield, "New Tools for Improving and Evaluating the Effectiveness of Research," presented at the Second Conference on Research Program Effectiveness, Washington (1965); published in *Research Program Effectiveness*, Gordon and Breach, New York, 1966, pp. 135–146.

Essays of an Information Scientist, Vol:1, p.188-195, 1962-73 Current Contents, #12, June 23, 1971
Reprinted from: Encyclopedia of Library and Information Science, Vol:5, p.16-40, 1971 - Marcel Decker Inc.

CITATION INDEXES

By
Melvin Weinstock, Senior Information Scientist
Institute for Scientific Information

PART I

Also See Citation Indexes - Part II

Citation Indexes - Part III

Introduction

Citation indexing is a relatively new method of organizing the contents of a collection of documents in a way that overcomes many of the shortcomings of the more traditional indexing methods. The primary advantage of citation indexing is that it identifies relationships between documents that are often overlooked in a subject index. An important secondary advantage is that the compilation of citation indexes is especially well suited to the use of man-machine indexing methods that do not require indexers who are subject specialists. This helps to make citation indexes more current than most subject indexes. Furthermore, citations, which are bibliographic descriptions of documents, are not vulnerable to scientific and technological obsolescence as are the terms used in subject indexes.

Citation indexing is based on the simple concept that an author's references to previously recorded information identify much of the earlier work that is pertinent to the subject of his present document. These references are commonly called citations, and a citation index is a structured list of all the citations in a given collection of documents. Such lists are usually arranged so that the cited document is followed by the citing documents.

The first practical application of this concept was *Shepard's Citations*, a legal reference tool that has been in use since 1873. *Shepard's Citations* owes its existence to the fact that American law, like English law, operates under the doctrine of *Stare Decisis*. *Stare Decisis* means that all courts must follow their own precedents as well as those established by higher courts. The precedents are the decisions handed down in previous cases.

To try a case under *Stare Decisis*, a lawyer must base his argument on previous decisions regarding a similar point of law. Before presenting the previous decision as a precedent, however, the lawyer must make sure that the decision has not been overruled, reversed, or limited in some way. *Shepard's Citations* enables the lawyer to do this with a minimum of trouble.

A legal case is always referred to by a code which consists of the volume and

Reprinted from:

Encyclopedia of Library and Information Science,
Volume 5, Pages 16-40, Marcel Dekker, Inc., New York (1971)

page number of the document in which the case is reported. Once a case is permanently reported, its reference code becomes fixed for all time. Thus, 301U.S.356 is a reference to the case reported in volume 301 of the United States Supreme Court Reports on page 356. Statutes are also referred to in a similar manner. Thus, Ch16Sec24NJRS refers to chapter 16, section 24 of the New Jersey Revised Statutes.

Taking advantage of this coding system, Frank Shepard devised a listing which shows every instance in which a reported decision is cited in a subsequent case. The listing also shows what statutes and journals cite the original decision.

Figure 1 is a Shepardized list of citations to a fictitious case. In the figure, the cited case is 101Mass.210; the items listed below it are the citing cases and an

CITED CASE → **101 Mass. 210**

	112	Mass.	65
a	130	Mass.	89
	165	Mass.	210
q	192	Mass.	69
	205	Mass.	113
o	221	Mass.	310
	281	U.S.	63
	35	H.L.R.	76

SUBSEQUENT CITING CASES

FIGURE 1. *Typical entry from* Shepard's Citations *showing cited and citing cases.*

article from the *Harvard Law Review* which refer to the case. The letters preceding the case codes show that the decision in the cited case was first affirmed (a), then questioned (q), and finally over-ruled (o).

To use *Shepard's Citations* a lawyer must first locate a previous decision relating to his current case. He does this by consulting a digest, index, or encyclopedia which will provide him with the case number for any given decision. The lawyer then looks up the case number in *Shepard's Citations* and finds all the subsequent citing cases. From this information, he can determine whether the original decision was affirmed or modified in any way. Thus, in the example given in Figure 1, the original decision could not be used as a precedent because it was subsequently over-ruled.

Problems with Traditional Subject Indexes

After World War II, users of scientific and technological literature were finding it increasingly difficult to find information pertinent to their own work. Several factors caused this situation. One was that the size of the literature was growing very rapidly. Projections indicated that by 1975 there would be over two million scientists in the world producing a million papers a year. These new papers would be added to the 10 million papers that would already be published by then (*1*).

As the growing volume of scientific information began to overwhelm the limited number of indexers that could be economically supported, there were delays of six months to several years before papers were classified. This, in turn, resulted in more and more scientists spending time needlessly duplicating existing work.

Another factor in the literature problems of scientists was the increasing need for exchanging information between scientific disciplines. The majority of subject indexes covered only one field or discipline. Science had become so interrelated that such arbitrary restrictions were causing the researcher to remain unaware of much valuable information. Fields such as oceanography, organic chemistry, and environmental science could no longer be placed into neat little cells. To be reasonably sure that he had most of the important information pertaining to any such field, the scientist was required to examine the literature of several disciplines. For example, a chemist selecting suitable materials for surgical implants or artificial internal organs might find useful information in chemical journals, medical journals, or engineering journals.

Although the growing size of the literature and the need for multidisciplinary information retrieval highlighted the shortcomings of traditional subject indexes, there were also other problems involved. The classification terms used in subject indexes are often ambiguous and lend themselves to different interpretations, especially when the user is not fully conversant with the details of a particular indexing system. Subject indexes also encounter the problem of assigning labels to new concepts. Many times, especially in fast-moving (or rapidly developing) fields like biochemistry, this is very difficult. Indeed, a consensus of what is the "proper" label for a concept may not be arrived at until some time after the original paper presenting the concept has been indexed under an inappropriate term.

Because indexers possess different intellectual abilities and technical skills, two different indexers will often use different key words, or headings, or subject terms when classifying the *same* document (*2*). Thus, it is not surprising to find *related* documents classified under entirely different subject headings with no clue to the searcher that this has happened. For example, an important 1963 paper on the topic of "seasonal variations in birth" (*3*) is indexed under the subject heading of "Periodicity" in the 1964 edition of *Index Medicus* (*4*). It is highly unlikely that anyone looking for information on seasonal variations in birth would ever think to look under "Periodicity" since it is quite a different concept than "Seasonal Variation."

These types of problems made clear the need for a system that would provide a unified index to the scientific literature that was current, free of semantic difficulties, and not dependent on the subject knowledge of indexers (*5*).

The Start of Citation Indexes for Science

REFERENCE TRADITION

As the doctrine of *Stare Decisis* provided the logic for *Shepard's Citations*, so did the "reference tradition" provide the rationale for citation indexes for science.

Scientific tradition requires that when a reputable scientist or technologist publishes an article, he should refer to earlier articles which relate to his theme. These references are supposed to identify those earlier researchers whose concepts, methods, apparatus, etc., inspired or were used by the author in developing his own article. Some specific reasons for using citations are as follows:

1. Paying homage to pioneers.
2. Giving credit for related work.
3. Identifying methodology, equipment, etc.
4. Providing background reading.
5. Correcting one's own work.
6. Correcting the work of others.
7. Criticizing previous work.
8. Substantiating claims.
9. Alerting researchers to forthcoming work.
10. Providing leads to poorly disseminated, poorly indexed, or uncited work.
11. Authenticating data and classes of fact—physical constants, etc.
12. Identifying original publications in which an idea or concept was discussed.
13. Identifying the original publication describing an eponymic concept or term as, e.g., Hodgkin's disease, Pareto's Law, Friedel-Crafts Reaction.
14. Disclaiming work or ideas of others.
15. Disputing priority claims of others.

In the early 1950s, the availability of this built-in system for linking scientific articles began to receive attention as the possible foundation of an indexing system for the scientific literature.

WELCH MEDICAL LIBRARY INDEXING PROJECT

In 1952, Dr. Chauncey Leake was chairman of the Committee of Consultants for the Study of Indexes to Medical Literature. This committee was supervising the Johns Hopkins Welch Medical Library Indexing Project which was sponsored by the Armed Forces Medical Library. Dr. Leake suggested that project workers should examine review articles in connection with their investigation of the problems with subject indexes to medical literature.

GARFIELD

This statement had considerable impact on Eugene Garfield, one of the Welch Project investigators. Garfield realized that nearly every sentence in a review article is supported by a citation to a previous work. Thus, a review article could really be considered a series of indexing statements. The problem then became one of transforming these statements into a consistent format that would be useful as an index.

ADAIR

In 1953, the Welch Project conducted a symposium, news of which was reportec in a Colorado newspaper. This article was read by William C. Adair, who wa a former vice president of the firm that produced *Shepard's Citations*. Adair wrote to the Welch Project and suggested that they consider the method employed by Shepard's as a possible indexing technique.

After examining *Shepard's Citations*, Garfield realized that the "citator" principle could provide a means of indexing review papers which could be extended to the scientific literature in general.

After the Welch Project ended, Garfield began graduate work in Library Science at Columbia University. During this period he continued correspondence with Adair and began writing a detailed article on citation indexes for scientific literature. The article was completed in 1954 and was edited and refereed by Professor Bentley Glass, who was then chairman of the Johns Hopkins Department of Genetics and on the editorial board of *Science*.

While his own article was awaiting publication, Garfield, who by then was an associate editor of *American Documentation*, suggested that Adair write a shorter article which would explain, in general terms, the operation of *Shepard's Citations*. The Adair article appeared in *American Documentation* in June of 1955 (*6*); Garfield's article appeared in *Science* in July of 1955 (*7*).

LEDERBERG

It was not until 1958, however, that the scientific community exhibited any specific interest in Garfield's idea. In that year, Professor Joshua Lederberg of Stanford University wrote to Garfield to inquire if any further work had been done on citation indexing. When informed of the financial problems involved in starting such a project, Lederberg suggested that Garfield should apply for a grant from the government.

GENETICS CITATION INDEX

In 1961, the National Institute of Health initiated a cooperative program with Garfield's Institute for Scientific Information (ISI) to prepare a citation index for the field of genetics (*8*). In addition to preparing the index, the program was to investigate and make recommendations on such general questions about citation indexes as:

1. Should there be a single citation index for all of science and technology, several rather broad ones, or many narrow ones, each focused on a single discipline?
2. In what ways is it possible to arrange a citation index (author, journal, etc.) and what way is the best for any given situation?
3. What techniques could be used for gathering the citation information?
4. Should books and technical reports be covered and to what degree?

Garfield soon recognized, however, that defining the genetics literature to be covered by a citation index would be quite difficult. Fine judgements would be required as to what was or was not genetics literature. At Garfield's suggestion, it was decided to undertake a comprehensive, interdisciplinary approach to preparing a citation index and then extract a genetics citation index from that base of information.

SCIENCE CITATION INDEX

The interdisciplinary data base was eventually used to produce the first *Science Citation Index* which was published in 1963. The first *SCI* covered the literature of the calendar year of 1961. It covered 613 journals, contained 1.4 million citations, and required five volumes. Nineteen per cent of the citations in the 1961 *SCI* data base were selected by special computerized procedures as "having to do with genetics" and were published separately as the *Genetics Citation Index*. The genetics index was also published in 1963 and was complete in one volume.

OTHER CITATION INDEXES

In addition to the *Science Citation Index* and the *Genetics Citation Index*, there have been several other efforts at compiling citation indexes or at using the principle of citation indexing in information retrieval systems. Most of these have been experimental in nature, extremely narrow in their coverage, or published on a one-time basis.

Some citation indexes provide coverage of the material published in just one journal. One of the earliest examples of this is the cumulative index to volumes 35 through 50 of the *Journal of the American Statistical Association*. This index was prepared with assistance from the Ford Foundation and was issued in 1959. In this index, both the cited and citing articles had to appear in the covered journal to be indexed. Another example of a citation index with single journal coverage is the one that appears in the cumulative index to volumes 1 through 31 of the *Annals of Mathematical Statistics*. Published on a one-time basis in 1962, this index is a listing of articles appearing in the *Annals* which is arranged by author and shows references to various abstracts of the article and other articles in the *Annals* which cite the original article.

Another citation index that was published only once is contained in the *Bibliography of Non-parametric Statistics*. Published in 1962, this index shows what items in the bibliography cite other items in the bibliography.

Since 1966, each monthly issue of the *Journal of Histochemistry and Cytochemistry* has contained a citation index to its own articles. This index is arranged by author and shows, for any article previously published in the journal, where and by whom the article has been cited in the preceding month in any of over 2200 other journals.*

An example of a citation index that covers more than one journal but is limited to a single field is the *Citation Index for Statistics and Probability* which is currently being produced by Dr. J. W. Tukey at Princeton University (9). This project

* This information is compiled from the Automatic Subject Citation Alert (ASCA) service of the Institute for Scientific Information.

was initiated in 1961 and is being conducted in cooperation with the National Science Foundation. The journals covered by this index are concerned with theoretical and methodological statistics. At the beginning of the project, about 50 journals were being completely covered, with another 75 journals covered on a selective basis. Currently, about 100 journals are fully covered and 150 are covered selectively. It is estimated that the first issue of this index will appear in 1971.

In 1968, the Shepard organization itself introduced *Shepard's Law Review Citations*. This new publication indexes 117 law reviews and periodicals and shows where any legal article written since 1947 has been cited in the covered journals from 1957 on.

Some citation indexes are prepared to test other aspects of citation indexing besides information retrieval. One of these was compiled by Ben-Ami Lipetz (*10*). As sources, Lipetz used four of the eight Russian physics journals published in English by the American Institute of Physics. Out of all the citations appearing in these journals, he included in his index only the citations to articles published in two heavily-used American physics journals. The resulting citation index was then distributed to a group of subscribers of the two American journals and an attempt was made to compare the frequency of use of the four Russian journals before and after the distribution of the index. The object was to see if increasing the user's awareness of relevant articles by the use of a citation index would influence their reading habits. Lipetz concluded that the index produced a measurable increase in the use of the Russian journals, although the increase was not large scale.

As a final example, special citation indexes can be prepared on demand through systems that directly connect the users with the data base, as is done with the Technical Information Project (TIP) at the Massachusetts Institute of Technology (*11,12*). TIP uses a time-sharing computer connected to remote consoles by telephone cables. The data base consists of the full bibliographies of articles from recent volumes of twenty-five physics journals. Special programs have been devised to enable a user to request a citation index to these articles that will meet his specifications. Thus, a user can obtain a citation index to all the articles, or to the articles from only one of the covered journals, or to the articles in a single volume of a covered journal.

Description and Use of a Citation Index

The multidisciplinary and interdisciplinary *Science Citation Index* is now used by the libraries of almost every major university in the United States. For all intents and purposes, it is the only citation index to the most used literature of science and technology. Because of this, the following detailed discussion on the compilation and use of a citation index will be based on the *SCI*. It is felt that the underlying principles and mechanics of citation indexing will be most clearly explained with this approach.

JOURNAL COVERAGE AND PREPARATION

The *Science Citation Index* provides an index to the contents of every issue published during a calendar year of approximately 2200 selected journals. Covered journals are considered *source* journals and the items they contain are called *source items*. All journals are indexed comprehensively to eliminate doubts as to whether or not a particular item is covered. All original articles and most other useful items in a journal are processed, including editorials, letters, and meetings. Ephemeral items such as advertisements and news notices are omitted. Although previously included, book reviews have not been used as source items since 1969.

Before keypunching, each source journal issue is edited and tagged to insure that all relevant data will be recorded. All foreign language titles are translated into English. All citations (footnoted or provided in a bibliography) are processed; where practical, citations are also extracted from the text. A separated punched card is prepared for every cited item appearing in every source item processed. For every source item, a set of punched cards containing the author(s), title, journal, etc. is also prepared. Each punched card is verified by direct comparison with the original journal.

Once the cards are punched and verified, the data are transferred from cards to magnetic tapes. During this process, the computer performs a unification routine which eliminates many errors from the original literature such as incorrect spellings in the names of cited authors and titles of cited publications (*13*).

The final data is formatted and further editing is done. Final printouts are then produced through the use of a computer-driven photo-composition machine. The final indexes are produced through photo-offset printing. Statistics are also tabulated in the course of obtaining the printouts.

REFERENCES

1. D. J. de S. Price, *Little Science, Big Science*, Columbia Univ. Press, New York, 1963, pp. 8–15.

2. C. P. Bourne, *Methods of Information Handling*, Wiley, New York, 1963, pp. 1–12.

3. K. Tietze, "On the Problems of Seasonal Fluctuations in the Dates of Birth and Conception," *Therapie der Gegenwart*, 102, 955–962 (1963).

4. *Index Medicus, 1964 (Cumulated)*, 5(IV), American Medical Association, 1965, p. S-2966.

5. E. Garfield, "A Unified Index to Science," in *Proceedings of the International Conference on Scientific Information*, Vol. 1, National Academy of Sciences–National Research Council, Washington, 1959, pp. 461–469.

6. W. C. Adair, "Citation Indexes for Scientific Literature?", *Amer. Doc.*, 6, 31–32 (1955).

7. E. Garfield, "Citation Indexes for Science," *Science*, 122(3159), 108–111 (1955).

8. I. H. Sher and E. Garfield, "The Genetics Citation Index Experiment," in *Proceedings of the American Documentation Institute, 26th Annual Meeting*, American Documentation Institute, Chicago, 1963, pp. 63–64.

9. J. W. Tukey, "Keeping Research in Contact with the Literature: Citation Indexes and Beyond," *J. Chem. Doc.*, 2, 34–37 (1962).

10. B.-A. Lipetz, "The Effect of a Citation Index on Literature Use by Physicists," in *Proceedings 1965 Congress of the International Federation for Documentation*, Spartan Books, Washington, 1966, pp. 107–117.

11. M. M. Kessler, "The MIT Technical Information Project," *Phys. Today*, 18(3), 28–36 (1965).

12. F. L. Alt and R. A. Kirsch, "Citation Searching and Bibliographic Coupling with Remote On-Line Computer Access," *J. Res. Nat. Bur. Stand., B*, 72B(1), 61–78 (1968).

13. I. H. Sher, E. Garfield, and A. W. Elias, "Control and Elimination of Errors in ISI Services," *J. Chem. Doc.*, 6(3), 132–135 (1966).