The Science Citation Index®: A New Concept In Indexing

MORTON V. MALIN

THE VARIOUS INDEXING SYSTEMS presently in existence have one basic goal: to provide researchers with access to the information they require. The stimulus for development of more sophisticated systems has, of course, been the need to probe the growing mass of information documents. Developers of indexing systems are, consequently, concerned not only with establishing a tool which provides access to ideas but also one which can deal with a large mass of information. The user's interest in indexing systems is similar since his requirement is to be able quickly to extract serviceable ideas from a mass of documents which may contain information of importance to him.

The purpose of this paper is to discuss citation indexing and its present application as exemplied by the *Science Citation Index®*, published by the Institute for Scientific Information as a new, unique, and necessary tool for scientific work. It is necessary, therefore, to describe briefly, and in general terms, the nature of conventional subject indexing systems in order that a basis for contrast between these and citation indexing can be obtained. For those who wish to read extensively on the subject of indexing, references are provided at the end of the paper under the section "Additional References."

Indexing Systems

Indexing, as used in this paper, refers specifically to methods which have been developed for organizing large files in such a manner that the content of the file can be retrieved. The form in which the content appears is basically immaterial, as is also the form in which the re-

Morton V. Malin is Director, Washington Office, Institute for Scientific Information, Philadelphia, Pennsylvania.

trieved information appears. By way of illustration, a book can be a file and an index to a book can be considered a method for organizing this file to extract information. At the other extreme, a library is also a file and a subject catalog is used to organize such a file for access to its content. Since different types of files and the requirements of the users of these files vary, different methods of organization have been attempted. The various indexing systems which have been developed reflect attempts to provide comprehensive and relevant retrieval of information.

The methods used in indexing range from the relatively straightforward systems of providing bibliographic data, well described by conventional library rules for descriptive cataloging, to more sophisticated attempts to establish dictionaries and other hierarchical systems which by assigning words, terms, descriptors or other language elements, will lead the person to the desired information. Conventional indexing systems have generally followed the procedure of assigning words or subject headings to a document to describe its content. Common methods in these systems consist of choosing words from the titles of articles or books, leaving out common or functional words. An improvement on this basic approach is to choose words from the text and to add to this list synonyms of these basic words; or to provide definitions for words spelled the same but having different meanings. Further refinement results when, in addition to providing variants, attempts are made to establish generic relationships between the indexing terms. Thesauri generally follow these methods.

As a further step towards reducing ambiguity, some indexing systems have established functional or syntactical relationships between the terms used to describe the content of a single document. Role indicators are sometimes assigned to the index terms to show the application of each term. These methods begin to move indexing systems away from derived methods, i.e., those solely dependent on deriving terms from text to those methods which depend on assignment of index terms based on authority lists. These latter schemes may or may not use hierarchical structuring.

Since an assignment index or classification scheme usually represents a particular viewpoint, the classification schedule or list of descriptors may organize a body of information in such a manner that it reflects only the needs of one type of user. It is not necessary, however, to restrict the indexing of a particular item to a single approach or point of view. A collection of material can be analyzed from several points

of view and a group of indexing terms synthesized to provide this capability for the system. This is often called "facet analysis" or relational indexing. The method used is to combine in a prescribed manner the terms derived from separate indexing examinations. Ranganathan's colon classification scheme and the work done by the members of the Classification Research Group in England are outstanding examples of this method of indexing.[1]

Citation Indexing

Citation indexing represents an entirely new approach to the problem of file organization and does not depend for its indexing technique either upon word derivations from text or upon word assignment. A citation index is an ordered list of cited articles, each of which is accompanied by a list of citing articles. The citing article is identified as a source, the cited article as a reference.[2]

In such a system there is no need to provide index terms either from the text or by assignment since the indexing system employed makes use of association of ideas to establish the relevancy of material in a document rather than the arbitrary authority of an indexer. Because citation indexing does not rely on word assignment, nomenclature, or terminology, it spans the gap between the subject approach of the cataloger and the subject approach of the information seeker without need to rely on any application of a directed artificial language to describe the subject matter in a paper. The subject matter is described by the direct association that exists between authors, for when one author cites the work of another, he is associating his subject matter with that of the cited author.[3] In general, a citation implies a relationship between a part or the whole of a cited paper and a part or the whole of the citing paper. By following the subsequent citations, the history of an idea can be traced: where and how it has been applied and whether it is sustained, rejected, or absorbed into later work.

The scientist is not interested only in obtaining specific information; browsing is also important to researchers who must keep alert and stimulated. Easy browsing should depend upon relationship of ideas rather than classification, and such relationship is provided by references which an author makes to earlier articles. This is usually achieved by assembling a bibliography of references to a special topic. Such bibliographies, although useful, are deficient because they only establish links going backward in time. Through the use of citation indexing, these references can be reorganized to provide the additional capability

of tracing the same topics forward in time and thus satisfying both the browsing instinct of the researcher and his need for specific information.[4]

The concept of citation indexing is not new, having been used for nearly three-quarters of a century by the legal profession in the form of *Shephard's Citations*, an index to legal cases. The system provides a listing of individual American court cases, each case being followed by a complete history written in a simple code. Listed under each case are publications that have referred to the case, other court decisions that have affected the case, and any other references that may be of value. This type of listing is particularly important to the lawyer because, in law, much depends upon precedent.[3]

Citation indexes may be specific to one subject area such as the *Genetics Citation Index*™, or they may be produced as indexes to material appearing in one journal, the most notable example being volumes 1-31 of the *Annals of Mathematical Statistics*. An index may also be multidisciplinary in nature, covering all fields of science as exemplified by the *Science Citation Index*®.[5]

Citation indexes can be produced using a simple system of coding entries in order to conserve space, or they can be prepared to provide more or less complete bibliographic histories for each item. Arrangements of entries are also flexible in citation indexing. They can be by author, by journal, or, if specific to a particular journal, by initial page of the cited item.

Science Citation Index®

The *Science Citation Index*®, the only regularly published citation index in science, is prepared by computer and cites source authors as entries. In May of this year, the 1966 *SCI*® was published, making the fourth complete annual accumulation. To provide proper background, it is worthwhile to summarize the content of the 1966 *SCI*® before describing its make-up and discussing its uses.

The *SCI*® uses, as its basic input, selected journals covering all the major and sub-disciplines of mathematics, the life, physical, and chemical sciences and, to a large degree, engineering. In the 1966 *SCI*® approximately 100 subject areas in these disciplines were covered by the journals providing the input to the system. During the year, 1,573 journals from forty-two countries were processed. The number of journal issues processed totaled 12,444 and the number of source journal items processed totaled 273,870. Total citations to journal and

non-journal articles amounted to 3,063,180. Over one million unique authored items are cited in the 1966 *SCI*R.[6]

The *SCI*R is a calendar year index, which means that, with minor exceptions, all specific issues of journals published and available during the time period covered by any edition of *SCI*R, are included. Journal selection has placed an important emphasis on the multidisciplinary journals in order to provide the broadest coverage and to enable the searcher to obtain information across disciplines. Each journal receives comprehensive treatment so that doubts as to whether any particular article was indexed are eliminated. Comprehensive treatment also means processing all material that has substantive information. All possible useful items, including editorials, book reviews, letters, meeting reports, critical reviews, and so on, are processed leaving nothing to chance; only ephemeral items such as advertisements and news notices are omitted.[6]

As was stated previously, arrangements for citation indexes can vary. As an illustration, let us return to an earlier example: *Shephard's Citations.* This particular index is organized by law cases to provide access to the volumes in which cases are reported. Thus, the citation 301 U.S. 356 is a reference to the case reported in volume 301 of the United States Supreme Court Reports on page 356. This reference becomes fixed for all future time and will appear in legal documents in the form noted. Statutes are cited by chapter and section number or by article, chapter and section number of the publication in which they appear. Ch. 16, Sec. 24, N.J.R.S. would thus refer to Chapter 16, Section 24, New Jersey Revised Statutes.[7]

A lawyer wishing to locate a case or cases for his authority first obtains a suitable case to fit his need. He then refers to *Shephard's,* using this case as his starting point, where he will find as his authority the subsequent cases that have cited it. The type of arrangement he would find would follow the fictional illustration shown below:

101 Mass. 210	(starting case)[7]
112 Mass. 65	
130 Mass. 89	
165 Mass. 210	
192 Mass. 69	
205 Mass. 113	
221 Mass. 210	
281 U.S. 63	
35 H.L.R. 76	

Tukey describes another type of index, *Annals of Mathematical Statistics*, in which a coding system is used.[4] The code provides the year, code for the author's last name, volume number, journal code and initial page of the cited and citing article. A sample illustration is given below:

Cited			*Citing*		
31PRN	23BMTA	11	34TRR	5AMSX	324
31PRN	23BMTA	23	58SKE	29AMSX	60
31PRN	23BMTA	114	39MCY	10AMSX	337
32PRN	24BMTA	404	50GRS	21AMSX	27
32 "	"	"	52MSN	23AMSX	126
"	"	"	59HRR	30AMSX	9800
32FLR	24BMTA	428	36RTZ	7AMSX	144

The first two digits are the year, the next three letters represent the author's last name, the next set of digits are the volume number. The letters following the volume number indicate the journal and the last set of digits are the initial page of the article.[4]

The *SCI*[R]*'s* arrangement provides more complete bibliographic information than the two systems described. The *SCI*[R] also provides anonymous citations and patent and corporate indexes. Moreover, it is in effect three major indexes containing in its eight volumes a *Citation Index*, a *Source Index*, and a *Corporate Index*.

1. The *Citation Index* of the *Science Citation Index*[R]. The *Citation Index* provides indexing entries to the current literature by means of the ordered listing of all items cited during a current year. It is arranged alphabetically by cited author and within this arrangement chronologically by cited year. A citation to a reference contains the author's name and initials, the cited reference year, the publication name, volume, and page number. Under the name of each cited author appears the source article citing this work. This line is arranged by citing author's name, publication, code identifying type of source item, citing year, volume and page. Only the first author for both the cited and citing articles are given in the *Citation Index*. While reference years may be any year in recorded history, the source year is always the current year. The form of the source article, i.e., letter, editorial, book review, etc. is indicated by a code appearing just before the source year. Asterisks flanking the reference year identify the first reference line and usually the first year cited for a given author. Dashed

lines indented under the reference author specify other cited works published by the same author.[6]

The *Citation Index* contains a section for anonymous items (no personal author specified for the cited work). These items are arranged alphabetically by the titles of the cited publications. Within each title the arrangement is chronological by year of publication, and within each given year by the reference volume and page number.[6]

The *Patent Citation Index* of the *Science Citation Index*R is a listing of foreign and domestic patents that have been cited or referred to in any of the foreign or domestic journals covered by the *SCI*R. The index is arranged in numerical order by patent number and usually provides also the year of issuance, the inventor's name, and country.

The following illustration shows the arrangement of the *Citation Index.*

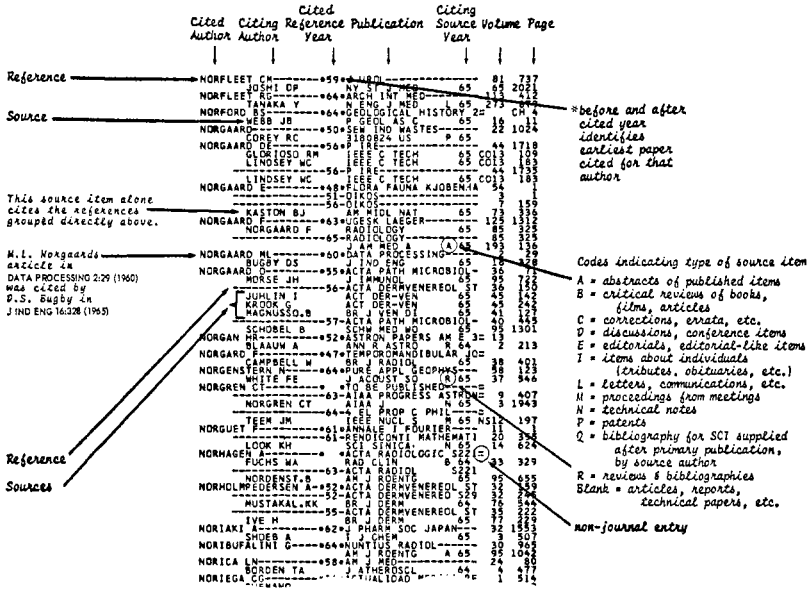
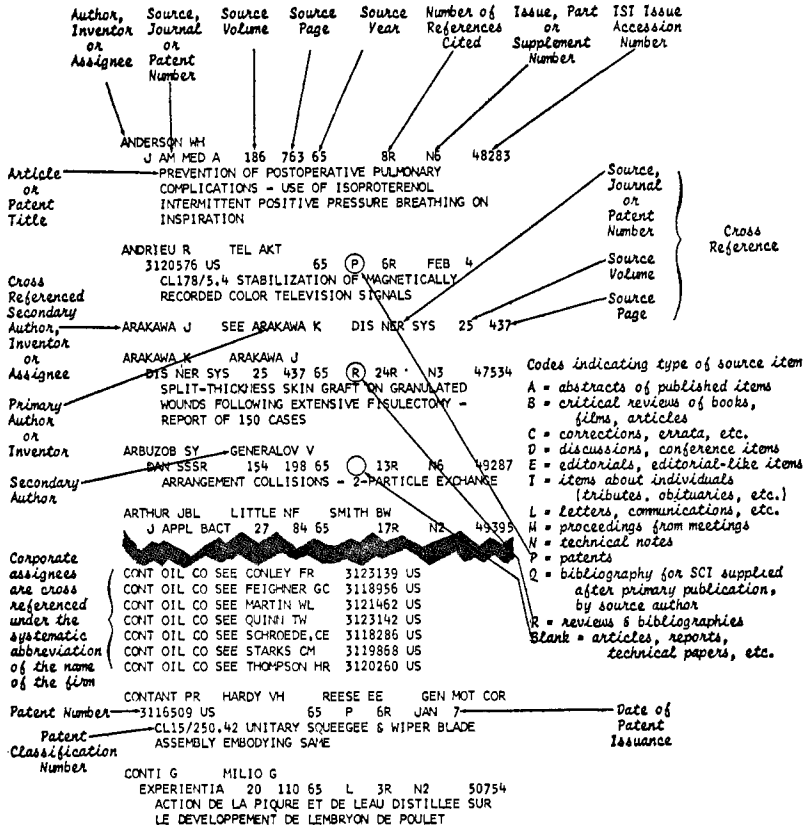


Figure 1. Arrangement of the *Citation Index.*

2. The *Source Index* of the *Science Citation Index*R. The *Source Index* provides a complete author index to the current literature and a full description of each citing item. This *Index* contains all source items processed for the *SCI*R, arranged by author. The entries include

all co-authors of items (maximum of ten), journal name, title, volume, page number, year, type of item (review, letter, correction, etc.) number of references in the bibliography of the source, the issue number, part or supplement number, and the ISI accession number for the journal issue, followed by the title of the article. Cross-references for every co-author are provided, even though the primary entries only list a maximum of ten. The "see reference" refers the user to the first author, journal, volume, and page. Anonymous source articles appear at the beginning of the *Source Index* and are arranged by journal, volume and page.[6]

Typical *Source Index* entries are illustrated below.



The data shown here simulate the type of material which appears in the Science Citation Index
The data in these entries are fictitious.

Figure 2. Typical *Source Index* entries.

3. *The Corporate Index* of the *Science Citation Index*[R]. The SCI also contains a *Corporate Index* in which all items published in the source journals processed are listed under the organization where work was performed. The *Index* will indicate under each organization the names of staff who have authored articles indicating journal, volume and page.[6]

Undertaking a Search in the SCI[R]

Using the *Science Citation Index*[R] is a relatively simple affair. In a citation index the subject of a search is symbolized by the starting reference rather than a word or subject heading. Searching, consequently, is independent of special nomenclatures or artificial languages. The searcher starts with a reference or an author he has identified through a footnote, book, encyclopedia or conventional word or subject index. He then enters the *Citation Index* section of the *Science Citation Index*[R] and searches for that particular author's name. When he locates the author's name, he then checks to see which of several possible references fits the particular one that he is interested in. Under the year, journal, volume and page number of this particular reference, he then looks to see who has currently cited this particular work. Having noted the bibliographic citations of the authors who are citing the work with which he started, the searcher then turns to the *Source Index* section and obtains the complete bibliographic data for the works which he has found.

Citation indexing is highly specific, but a search may also be readily expanded in order to build a more extensive bibliography for a particular inquiry. For example, having found a number of source articles, the searcher can use the bibliographies of one or several of these as other entries into the *Citation Index*; this process is called "cycling." Since authors frequently write more than one closely related paper, additional articles by the author of the first starting reference can also be used as entry points and citations to these articles can be examined to obtain additional information. The *Source Index* itself may yield relevant current articles by a given author, even though they may not cite any of the known starting references.

The fundamental question one can answer quickly through the *Citation Index* is where and by whom has this paper been cited in the literature? The *SCI*[R] is also used by scientists to determine whether their work has been applied or criticized by others. It can facilitate feedbacks in the communication cycle. Any author may choose to

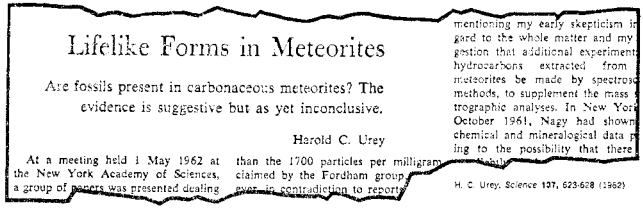
ignore the citations to his own work and still use the *Index* to retrieve publications which cite work by other scientists. The *SCI*[R] can be used to identify scientists currently working on special problems or to determine whether a paper has been cited, whether there has been a review of a subject, whether a concept has been applied, a theory confirmed, or a method improved. Because indications of corrections are published in the *SCI*[R], it is useful as an aid in following particular articles. Only the user's imagination limits the extent to which the *SCI*[R] can be a useful tool for the scientist and librarian.

The usefulness of the *SCI*[R] can best be illustrated by conducting a hypothetical search.

The interest of the searcher is in life-like forms in meteorites. He has a starting reference, an article by Harold C. Urey published in *Science*, 1962, Volume 137, pages 623-628. To determine what other work has been done in this particular area, the searcher goes to the *Citation Index* and looks under Urey identifying the '62 *Science* article. Indented under this particular citation is a citation to a work by Mueller, G., that appeared in *Nature*, 1965, Volume 205, page 1200. Moving from the *Citation Index* to the *Source Index*, one looks under Mueller, finds the particular journal, year, volume and page reference and there sees that Mueller published a letter in 1965 with the title "Interpretation of Micro-Structures in Carbonaceous Meteorites." If the searcher is interested in obtaining even more information than this one particular reference, he can now obtain Mueller's article, look at the ten references that are given and using each reference as a starting point, enter the *Citation Index* once again. Since the probabilities are very high that if Mueller cited Urey's paper, the other papers that he would cite would also deal with the question of life-like forms in meteorites, then one can expect that the additional references will yield more current source articles on this particular subject.

The procedure of cycling permits the searcher to go backwards and forwards in time because, to cite the Urey-Mueller example again, he has begun with a past citation and moved forward to a current citation. If he follows through and uses the references in Mueller's paper as other starting points, then he is again going backward in time; and when he finds the current citations to these particular references, he has moved forward once again.

Citation indexing has found applications in other fields. It can be used by librarians in selecting journals for collections, and it is useful for certain types of sociological research.[8-10] However, it is less



Lifelike Forms in Meteorites

Are fossils present in carbonaceous meteorites? The evidence is suggestive but as yet inconclusive.

Harold C. Urey

At a meeting held 1 May 1962 at the New York Academy of Sciences, a group of papers was presented dealing

than the 1700 particles per milligram claimed by the Fordham group, ever, in contradiction to reports

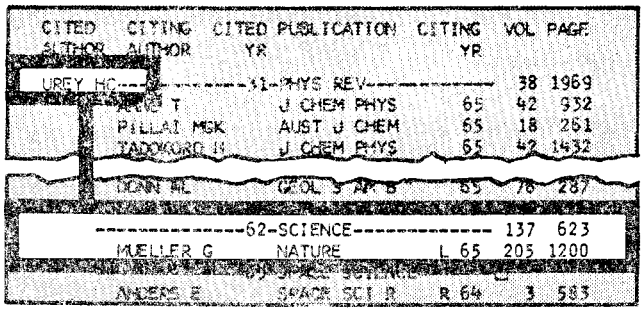
mentioning my early skepticism in gard to the whole matter and my gestion that additional experiment hydrocarbons extracted from meteorites be made by spectros methods, to supplement the mass trographic analyses. In New Yor October 1961, Nagy had shown chemical and mineralogical data p ing to the possibility that there

H. C. Urey, Science 137, 623-628 (1962)

```
UREY HC-----62-SCIENCE----- 137  623
```

A

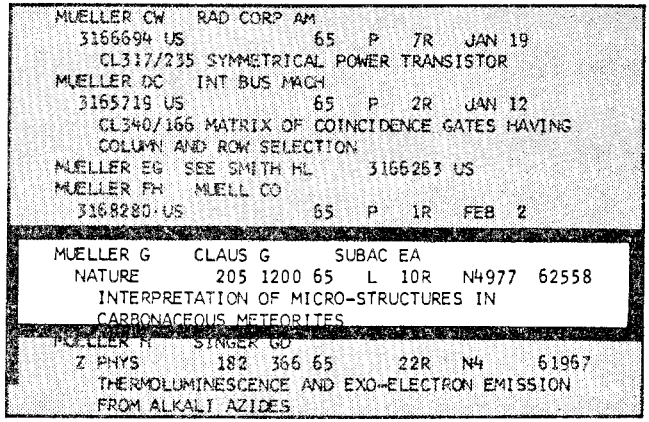
CITATION INDEX SECTION



CITED AUTHOR	CITING AUTHOR	CITED PUBLICATION YR	CITING YR	VOL	PAGE
UREY HC		-----51-PHYS REV-----		38	1969
	[illegible] T	J CHEM PHYS	65	42	332
	PILLAI MGK	AUST J CHEM	65	18	261
	TADOKORO H	J CHEM PHYS	65	42	1432
	DOAN AL	GEOL S AM B	65	76	287
		-----62-SCIENCE-----	137	623	
	MUELLER G	NATURE L	65	205	1200
	ANDERS E	SPACE SCI R R	64	3	583

B

SOURCE INDEX WITH TITLES



```
MUELLER CW   RAD CORP AM
  3166694 US              65   P   7R   JAN 19
    CL317/235 SYMMETRICAL POWER TRANSISTOR
MUELLER DC   INT BUS MACH
  3165719 US              65   P   2R   JAN 12
    CL340/166 MATRIX OF COINCIDENCE GATES HAVING
    COLUMN AND ROW SELECTION
MUELLER EG   SEE SMITH HL      3166253 US
MUELLER FH   MUELL CO
  3168280 US              65   P   1R   FEB  2

MUELLER G    CLAUS G        SUBAC EA
  NATURE       205 1200 65   L   10R   N4977  62558
    INTERPRETATION OF MICRO-STRUCTURES IN
    CARBONACEOUS METEORITES
MUELLER FI   SINGER GO
  Z PHYS       182  366 65        22R   N4     61967
    THERMOLUMINESCENCE AND EXO-ELECTRON EMISSION
    FROM ALKALI AZIDES
```

C

Figure 3. Hypothetical search of the *SCI*[R].

the intent of this paper to discuss the multiple uses of citation indexing than to present the general ideas behind the development of this indexing technique and its basic use.

General Comments

Relevancy of information obtained through use of citation indexes and the *SCI*[R] in particular, is generally very high. A citation does not suffer from failure to include descriptive terminology—rather it is a brief representation of the content of the documents it identifies. Those who have studied citation indexes know that only a small number of reference citations are needed to isolate a particular document from all others in a collection.[11]

Irrelevancy or "noise" can be eliminated in using citation indexes by careful selection of references as access points. In particular, if narrow specificity is desired in a search, the starting reference should be one which deals with only one subject.[12] Another method of reducing "noise" is to apply the concept of bibliographic coupling to the search. This involves recording sources obtained from the index which have two or more starting references in common.

Citation indexing has advantages that conventional indexing systems do not supply. A major advantage is the fact that there is no terminology problem and there is no need to guess how an indexer might have indexed a particular item. Multidisciplinary coverage is another distinct benefit afforded by citation indexes such as *SCI*[R]. Of importance too, are the speed and convenience which citation indexing provides to the searcher. Searches have been conducted using *SCI*[R] on a variety of subjects and have yielded excellent results, some of which are available from the Institute for Scientific Information upon request. Carol Spencer [12] reports on an experiment using *SCI*[R], *Chemical Abstracts*, and *Index Medicus* to develop a bibliography on the drug thalidomide. The results of the experiment showed that, for a period up to eight hours, an *SCI*[R] search yielded the highest number of references of the three indexes. Many searches take only a few minutes; consequently, several hours of searching will yield relatively complete bibliographies.

A citation index also has the advantage of being self-organizing. Each new reference to an old paper and each new citation modifies the previous store of information. The system is constantly being upgraded by current information, which in turn helps bring old information up to date.[13]

Conclusion

The crucial fact that must be faced by both the user of information and the librarian is the continuing growth in the number of documents reporting the results of scientific investigation. Concomitant with this growth are the continued needs of the scientist to search narrow areas, and to browse effectively. Association of ideas provides an effective means for coping with the information explosion by providing focused access and ample browsing capacity. Classification of a field changes rapidly, but the relationship of one article to another changes little. The relationship between articles is provided by references authors make to earlier works, and a citation index is the best means for establishing this continued relationship both forward and backward in time.[4]

References

1. Bourne, Charles P. *Methods of Information Handling.* New York, J. Wiley, 1963, pp. 13-33.

2. Martyn, John. "An Examination of Citation Indexes," *Aslib Proceedings,* 17:184, June 6, 1965.

3. Garfield, Eugene. "Citation Indexes for Science," *Science,* 122:108-111, July 15, 1955.

4. Tukey, John W. "Keeping Research in Contact with the Literature: Citation Indices and Beyond," *Journal of Chemical Documentation,* 2:34-35, Jan. 1962.

5. Martyn, John. "Citation Indexing," *The Indexer,* 5:5-9, Spring 1966.

6. *Science Citation Index 1966.* Vol. 1, Philadelphia, Institute for Scientific Information, 1966, pp. 8-20.

7. Adair, W. C. "Citation Indexes for Scientific Literature?" *American Documentation,* 6:31, Jan. 1955.

8. Garfield, Eugene, *et al. The Use of Citation Data in Writing the History of Science.* Philadelphia, Institute for Scientific Information, 1964.

9. Bayer, Alan E., and Folger, John. "Some Correlates of a Citation Measure of Productivity in Science," *Sociology of Education,* 39:381-390, Fall 1966.

10. Margolis, J. "Citation Indexing and Evaluation of Scientific Papers," *Science,* 155:1213-1219, March 10, 1967.

11. Garfield, Eugene. "Can Citation Indexing be Automated?" *In* M. E. Stevens, *et al.,* eds. *Statistical Association Methods for Mechanized Documentation: Symposium Proceedings, Washington 1964.* (National Bureau of Standards Miscellaneous Publication 269.) Washington, D.C., U.S.G.P.O., 1965, p. 190.

12. Spencer, Carol C. "Subject Searching with *Science Citation Index:* Preparation of a Drug Bibliography Using *Chemical Abstracts, Index Medicus,* and *Science Citation Index* 1961 and 1964," *American Documentation,* 18:88, April 1967.

13. Garfield, Eugene. "Citation Indexing: A Natural Science Literature Retrieval System for the Social Sciences," *American Behavioral Scientist,* 7:58, June 1964.

ADDITIONAL REFERENCES

Association of Special Libraries and Information Bureaux. *Proceedings of the International Study Conference on Classification for Information Retrieval.* Pergamon Press, London, 1957.

Bernier, Charles L. "New Kinds of Indexes," *Journal of Chemical Documentation*, 1:62-67, Jan. 1961.

Collison, Robert L. *Indexes and Indexing.* 2nd ed. New York, J. de Graff, 1959.

Foskett, D. J., *et al.* "Classification Research Group Bulletin No. 7," *Journal of Documentation*, 18:65-88, June 1962.

Gull, C. D., "Alphabetic Subject Indexes and Coordinate Indexes: An Experimental Comparison," *College & Research Libraries*, 14:276-281, July 1953.

Ranganathan, Shiyali. *Colon Classification.* 6th ed. New York, Asia Publishing House, 1960.

Ranganthan, Shiyali. "Classifying, Indexing, Coding." *In* Allen Kent, ed., *Information Retrieval and Machine Translation.* (Based on the International Conference for Standards on a Common Language for Machine Searching and Translation, Western Reserve University, 1959.) New York, Interscience Publishers, 1960.

Shera, Jesse H., *et al.*, eds. *Information Systems in Documentation.* (Based on the Symposium on Systems For Information Retrieval, held at Western Reserve University, 1957.) New York, Interscience Publishers, 1957.

Taube, Mortimer. "Specificity in Subject Headings and Coordinate Indexing," *Library Trends*, 1:219-223, Oct. 1952.

Taube, Mortimer, *et al. Studies in Coordinate Indexing.* Vol. 1. Washington, D.C., Documentation Inc., 1953.

Vickery, Brian C. *Classification and Indexing in Science.* 2d ed. London, Butterworths, 1959.

The reader is also referred to the excellent bibliography covering the subject of indexing that can be found in the footnotes to Chapter 2 *in* Bourne, Charles P., *Methods of Information Handling*, New York, J. Wiley, 1963.