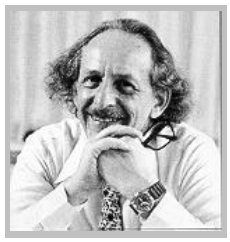
Reprinted in Essays of an Information Scientist, Vol:6, p.540-547, 1983



"World Brain" or "Memex?" Mechanical and Intellectual Requirements for Universal Bibliographic Control

Eugene Garfield

President, Institute for Scientific Information

Introduction

When Mr. Bergen asked me to speak on the "Design of Bibliographic Systems," he suggested that I discuss the experiences of the Institute for Scientific Information in using and compiling citation indexes. I shall not take advantage of Mr. Bergen's mandate simply to repeat to this audience the mechanics and philosophy of the *Science Citation Index* system. Naturally it is impossible to avoid some repetition of ideas I have expressed elsewhere. However, in this paper, I have attempted a synthesis of conventional word indexing systems as well as citation indexing systems. I shall summarize our plans and accomplishments and indicate the factors to be considered in designing a system of universal bibliographical control.

The Perennial Dichotomy—Storage Medium or Information Stored?

In several papers,[1] I have described the *Science Citation Indexes* and the Unified Index to Science[2] as preliminary steps toward achieving the dream of universal bibliographical control which H. G. Wells symbolized in the *World Brain*.[3] To some, the Wellsian term "World Brain" might seem less appropriate than "Memex," the term chosen by Vannevar Bush[4] to symbolize the ideal information retrieval device. However, there is a world of difference between Memex and "World Brain" —essentially the difference between hardware and software—between a communication carrier and the intellectual-message-carried problem. The "World Brain" symbolizes the information stored—"Memex," the storage device. In designing any bibliographic system, it is imperative to make these distinctions. It distresses me when this dichotomy is glossed over in vague generalizations about the so-called Information Explosion.

Information Systems Adapt to Many Devices

The *Science Citation Index* (SCI) system is an intellectual achievement. It can be used or take expression in various devices. The information it contains is already stored in three forms or devices—magnetic tapes, microfilm, and printed books. The latter form, the set of printed books, is probably still the "cheapest" random-access device available for servicing a large diversified audience, especially since user time is rarely accounted for in measuring costs by library administrators.

Magnetic Tapes

The magnetic tapes on which the SCI is stored are used for dissemination of information to individual scientists, either through keyword profiles[5] or through citation profiles.[6] The tapes can also be used for retrospective searching. However, the basic shortcoming of sequential search methods, namely that all items in the file must be examined one by one (as opposed to random-access search), applies to this particular file, as it does to other sequential tape search systems as, for example, MEDLARS.[7] The cumulated SCI magnetic tapes, like MEDLARS, are later used to prepare hard-copy printout for the photo-offset printed books.

Random-Access Memory Cells

Alternative random-access devices besides printed books are now available and getting cheaper. The SCI system can be stored in large banks of so-called memory cells which, like a book, would permit almost instantaneous access to any address in the file. Thus, instead of looking up a particular citation or author in the printed SCI, the reader would operate a keyboard linked to a computer with a large auxiliary memory which would store the several billion characters now stored in the printed books or magnetic tapes.[8]

Time-Shared Computers but not Time-Shared Communication Lines

Surprisingly, the main deterrent to implementing such a system is not the relatively high cost of the memory cells. The cost of the cells is shared by all the users in the network. The most significant cost is the long-distance communication line. The average basic cost of a long-distance telephone call is still too high in many instances. Furthermore, the telephone system is not yet organized so that you only pay for the line during the time you actually are using the computer. We have time-shared computers, but we do not have time-shared telephone lines. On a time-shared computer, the computer is working for someone else during your silent or thinking period. On a telephone call, you pay whether you are speaking or not. This is a basic simplification but essentially that is the problem. I fully realize that in fact many messages are transmitted simultaneously on modern lines, but time-sharing is a somewhat different problem which undoubtedly will be solved.

Allowable Search Time Is Inverse Function of Use

Why do we need such elaborate hardware when the printed book is available? In any IR system, the amount of time allowable in each look-up or search operation is an inverse function of usage. If you use the phone book once each day, you don't mind a 30-second delay in finding a number. However, if you use the phone book once an hour, then half a minute represents almost one per cent of your total working time. If you use the phone book once each minute, the operation involves half your time. A scientist doing bibliographic research for a review paper or a book may spend more time turning pages than reading citations! Consequently, any feature of an IR system that reduces look-up time is an improvement. Until recently, academic, public, and most other libraries have been unwilling to pay the cost of reduced look-up time. The user pays with his lost time.

Updating Daily

Consider the effect of computer storage on another innate deficiency of printed books—updating. To look-up where a given paper was cited in 1964, now requires about 15 seconds depending upon your manual dexterity and scanning speed. However, to determine if the paper was cited in 1964 *or* 1965 requires that you use the *1964 Annual Index* and the first two quarterly supplements for 1965. Your search time is tripled since three index issues must be consulted. In a very large random-access computer system, you would obtain the same information by a single inquiry. Furthermore, you also may want to count the time required to copy the pertinent citations by hand or by photocopy machine. In the computer system, the full bibliographic data would be printed on your console typewriter, your teletype unit, and/or your video display tube. Indeed, the computer could also generate a perforated tape or a set of punched cards, which, in turn, could be used later to activate another input device to avoid repetitive typing.

At the Institute for Scientific Information, we are acutely conscious of the potential manipulative dexterity of time-shared use of random-access systems. Such dexterity may or may not be desired by

[1] E. Garfield, *"Science Citation Index*—A New Dimension in Indexing," *Science*, 144 (1964), pp. 649-654.

[2] E. Garfield, "A Unified Index to Science," *Proceedings of the International Conference on Scientific Information, 1958* (Washington, D.C.: National Academy of Sciences-National Research Council, 1959), pp. 461-474.

[3] H. G. Wells, *World Brain* (Garden City, N.Y.: Doubleday, Doran & Co., 1938).

[4] V. Bush, "As We May Think," *Atlantic Monthly*, 176 (1945), pp. 101-108.

[5] C. R. Sage, R. R. Anderson and D. R. Fitzwater, "Adaptive Information Dissemination Without Indexing," *American Documentation*, 16 (1965), pp. 185-200.

[6] I. H. Sher and E. Garfield, "The *Science Citation Index* and ASCA—New Tools for Improving and Evaluating the Effectiveness of Research. Paper presented at the Office of Naval Research Second Conference on Research Program Effectiveness, Washington, D.C., July 27-29, 1965. (To be published in *Proceedings* for this conference.)

[7] L. Karel, C. J. Austin and M. M. Cummings, "Computerized Bibliographic Services for Biomedicine," *Science*, 148 (1965), pp. 766-772. See also "Erratum" in *Science*, 148 (1965), p. 1257.

[8] M. M. Kessler, "The MIT Technical Information Project," *Physics Today*, 18 (3) (1965), pp. 28-36.

Reprinted from *"The Foundations of Access To Knowledge,"* 1968, Syracuse University Press, Box 8, University Station, Syracuse, New York 13210, pp. 169-196 $5.75

absolutely consistent, if the same paper is published by an author in two different publications, both papers should always be retrieved together. We can say they are 100% similar, and, of course, their information content would be equal. For the purpose of this discussion, I shall call the degree of similarity "descriptor coupling." The degree of similarity between two or more documents is a function of their descriptor coupling. This applies to all natural language systems using words whether they are called "descriptors," "subject headings," or "uniterms," etc.

Bibliographic Coupling

The similarity of two documents can also be measured by bibliographic coupling. Kessler[45] studied bibliographic coupling extensively, though it was Fano who first expressed the notion in 1956.[46] Of course, the idea of grouping similar documents by citation relationships is the essence of citation indexing. Nevertheless, one can employ bibliographic coupling to measure similarity regardless of whether one wishes to prepare a citation index. Thus, for any bibliographic system, if one needs a method for determining the degree of similarity between two documents, one can examine the number of reference or footnote citations they share in common. As with the indexing terms in descriptor coupling, the document, in bibliographic coupling, is described by its bibliography—the set of reference citations the author has used in documenting his work. Each bibliographic citation is a descriptor.

To test the equivalence of citations to words, one can index a given source document by using words taken from the titles of the cited papers[47] or subject headings used by an indexer to index the cited papers.[48] This procedure provides additional insight to the problem of automatically or algorithmically identifying what is really new in a given source document. If the set of words extracted from the cited titles is compared to the text of the source document, then new words, such as names of new chemical compounds, stand out. They cannot appear in the list of old words compiled from the bibliography. Thus, in our previous example involving "Engineering Human Development," this phrase would not appear in a list of the words used to index the papers Crow cited, but the word "euphenics" would occur if the book he reviewed were thoroughly indexed.

Salton carried the comparison one step further. He measured similarities based on frequency of occurrence of a term in the document itself. He concluded, in comparing similarity coefficients derived from term analysis and from citation analysis, that "citations provide a large number of relevant index terms not originally available with a given document collection, and thereby create a much more flexible retrieval process."[49]

Borko has confirmed the notion that grouping of documents "according to similarity of word content" facilitates browsing and retrieval. He does not attempt to confirm experimentally a similar claim for citation analysis.[50]

An extremely exciting application of bibliographic coupling is observed in the ASCA system previously mentioned. In the ASCA system, the user creates a bibliography of about 50 papers or books. This bibliography is his field-of-interest profile. Suppose we found a paper which cited all or most of these 50 references. You can be certain that the new paper would have a direct relationship to the research of the ASCA client. Indeed, we have used this method to uncover cases of duplicate research. In the ASCA service, we frequently find that several papers in a profile may be cited in a current work. The degree of coupling is a measure of similarity between the retrieved paper and the interest profile.

Relevance is Subjective

Note that I have carefully avoided the term "relevance" even though, in general, a bibliographic coupling strength or threshold of three or more will invariably turn up a relevant paper. However, even a coupling strength of one or two may turn up an even more relevant document. Relevance is a highly subjective factor which only the user can evaluate.[51] In fact, two patrons may have similar profiles but disagree on the relevance of any retrieved document.[52] Relevance is not discernible on an *a priori* basis. Similarity can be measured objectively, whereas relevance is purely subjective.

I have described how citation coupling is used in the ASCA system. What of word coupling? Naturally this is possible and is employed in the SDI systems mentioned earlier. However, the natural ambiguity of language makes word coupling more difficult to achieve. The success of word coupling is in part determined by the type of terminology peculiar to a given field. "Euphenics" is so sufficiently specific and rare that I can rely on its high information content and discrimination value to turn up relevant information when it occurs. A word like "films," however, is so highly ambiguous, it will create more noise than music.

In the ASCA system, we also can select documents based on key words appearing in titles. To date, we have restricted the use of words to a pure word-oriented SDI system.[53] We wished to avoid giving the user the untrue impression that indexing words in titles is equivalent to the depth and specificity of indexing achieved by citation indexing. The depth or degree of specificity of citation indexing, in some instances, can be matched by conventional deep indexing as done by *Index Chemicus, Chemical Abstracts,* or for MEDLARS Demand Bibliographies. However, even these indexing systems cannot cope with the complexities of indexing mathematical formulas, complex methodology, etc. That is why these systems cannot readily answer such questions as "Where has Smith's equation XYZ, as modified by Ford, been used?"

The Indivisibility of Knowledge

Here is the final question I should like to discuss in terminating this discussion of the World Brain—its parameters, dimensions, etc. I have always believed that it is ultimately impossible to segregate knowledge into discipline-oriented compartments. Every university library administrator knows the consequences of trying to achieve this feat. What I should like to tell you now is why it would not satisfy your users even if you could accomplish the perfect classification system. One hears a great deal about the chemical literature, biomedical literature, mathematical literature, etc.

Literature of Science vs. the Literature of Interest to Scientists

There is a tendency to confuse the literature of science with the literature of interest to scientists. For example, I recently attended a meeting concerned with the documentation of oceanographic literature. There is an important, though small, percentage of the scientific literature which can be called "pure" oceanography. An experienced group of catalogers could identify this literature, most of which occurs in a small number of journals. However, we compared this literature to the literature "of interest" to oceanographic scientists who use our ASCA service. These scientists, like most other scientists, are interested in subject matter or concepts which may be found anywhere in a hard core of scientific journals. In one sense, they couldn't care less about the literature of pure oceanography, to which they may contribute, because they are in regular touch with this literature through personal contacts. It should be obvious that 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 hundreds of applied fields. Knowledge is interdisciplinary, and our many existing fragmented approaches to bibliographical control are only compromises dictated by bibliographic poverty in the midst of research affluence. To achieve universal bibliographical control requires that we think and dream, as did Wells and Bush, on a larger, though not necessarily more extravagant, scale.

[45] See footnote 43.

[46] R. M. Fano, "Information Theory and the Retrieval of Recorded Information," *in* J. H. Shera, A. Kent and J. W. Perry, eds., *Documentation in Action* (New York: Reinhold Publishing Corp., 1956), pp. 238-244.

[47] M. E. Stevens and G. H. Urban, "Training a Computer to Assign Descriptors to Documents: Experiments in Automatic Indexing," *AFIPS Conference Proceedings,* Volume 25, *1964 Spring Joint Computer Conference* (Baltimore: Spartan Books, Inc., 1964), pp. 563-575.

[48] See footnote 1.

[49] See footnote 42d.

[50] See footnote 42e.

[51] L. B. Doyle, "Is Relevance an Adequate Criterion in Retrieval System Evaluation?" SP-1262 (Santa Monica, California: System Development Corporation, 1963).

[52] R. S. Taylor, "The Process of Asking Questions," *American Documentation,* 13 (1962), pp. 391-396. As cited in Doyle in footnote 51.

[53] Institute for Scientific Information, Brochure #5-16-40-19A, "ISI Source Index Magnetic Tapes Description" (1965).

cal objectives—to make information on euphenics retrievable. In the one system, the word is the access or starting point; in the other, the citation is the starting point.

Dialogue Between Librarian and Scientist

The expert geneticist does not have to be told that Lederberg has written on euphenics;[15] nor will he have to be told other key events in the development of his particular specialty. On the other hand, the reference librarian is, of necessity, a generalist and cannot be expected to remember that which has become second nature to the specialist. This means that the reference librarian must usually engage the scientist in a dialogue, the purpose of which is to simplify the use of the systems available. If I ask my librarian to "find me papers on euphenics," there are a number of reasonable questions that can or must be asked by the librarian. If the librarian is embarrassed to do so, or feels that any sacrifice of additional time on the part of the patron is unjustified, then he pays a heavy price. He employs a series of bibliographic twists and turns taught to him in library school or which have become second nature through experience.

In the Process of Question Translation, What is Reasonable?

Detailed algorithmic descriptions of the reference question-answer procedure will have to be developed before computers can take over even a small part of the reference function in libraries. This type of programmed or algorithmic probing of the reference collection is implied in Swenson's recent work on using flow charts in reference work.[16] It all boils down to this: How do you translate the reference question into the form a given system can cope with? For example, what can the system do with a question like "Find papers on engineering human development." The answer to the question happens to be the same as the previous question "Find me papers on euphenics." Is it unreasonable of the reference librarian to ask, "Do you know of anyone working in this field, or do you know the alternate terminology used to symbolize the subject?"

Before elaborating this example, it is important to remind you that there are many questions for which answers cannot be provided. In designing a new system, you would do well to ask how long it takes to provide a reliable negative answer to a question.[17] In fact, a large number of your patrons are hoping to be told that the "original" idea they are investigating is original. Perhaps some librarians have become unpopular because of their uncanny ability to find something on almost any topic. This might be fine for an examiner of patents at the Patent Office, but what about the inventor who thinks he has an original invention?

Is the Completely Up-to-date Thesaurus Necessary?

Since I was privileged to know Dr. Lederberg at the time he gave his Ciba Foundation talk on euphenics, it would not have been a problem for me to translate the phrase "engineering human development" into the equivalent word "euphenics." But what is the librarian to do in answering such a reference question presented by a patron who is interested in a similar concept but does not know the term "euphenics" exists? Right now, it would be impossible for the librarian to use any existing system successfully. The general reference and lexicographic apparatus in libraries is quite far behind the advance of scientific terminology. Even today, three years after Lederberg's original paper, the word "euphenics" cannot be found in a dictionary or encyclopedia. And even if it were, how would the librarian know that euphenics is the key word that is synonymous to the subject "engineering human development?" Presumably, one might find the term through Roget's *Thesaurus* if it, too, were sufficiently up-to-date. And presumably, if the various indexing services, like *Index Medicus* and *Biological Abstracts,* issued subject heading lists which took into account every new biomedical word, phrase, or eponym, and were completely up-to-date, then we would be able to identify such key words more easily. And while we're dreaming, our up-to-date thesaurus would have to contain other cognate or synonymous terms in English, German, and other languages for the same primordial concept. Yet we have seen that this concept, no matter how it is expressed in natural language, is consistently symbolized by the citation "Lederberg J, 63, *Nature* 198, 428."

One might parenthetically ask, "If this idealized thesaurus were completely up-to-date, wouldn't it be just as simple to identify the key citation in which the term just appeared as to list descriptive words? Would not an up-to-date encyclopedia, or at least the author of a recent textbook on genetics, mention the inventor of a term?" There is no escape from this circle—if one can identify the primordial word, he can identify the primordial citation!

Ordinarily the scientist would not wade through numerous reference compendia as described above. He might engage a friend in a dialogue, asking if he knew who was working on the problem. Together, in a scientist-to-scientist dialogue, they would identify Lederberg's association with euphenics and thus gain access to current information through an author or a citation index. Hopefully, the wise librarian would try the equivalent method by contacting a special librarian.

All Existing Systems Make Demands on User

The previous analysis is necessary not only to indicate why a citation is a subject, but also to illustrate that in any non-ideal system, some accommodation must be made to the system itself in order to derive benefit from it. Of course, any system which makes *excessive* demands of the user is doomed to failure, but no system yet available makes *no* demands on the user. Indeed, if such a system existed, it would be the "World Brain." More than likely, it would be a network of human brains which were linked by telepathy to one another—a sort of community thinking machine.

The Ideal Gas Theory vs. Network Theory of Bibliographic Organization

Are documents dependent or independent entities? The traditional approach to information system design is to treat individual documents as independent entities. There is a basic fallacy in this approach which results not only in the loss of information links but in basic indexing inefficiency.

In the average conventional system, if an author publishes two papers on the same general theme in two different journals at two different times, essentially the same indexing procedure will be followed for both papers as though the papers are unrelated publishing events. In this procedure, the same selection of key words or headings should be made to describe the main theme of each paper. In practice, we know this does not occur consistently. In this *isolationist* kind of indexing, generally no effort is made to establish a relationship between the document being indexed and the documents already indexed in the collection. There are exceptions to this rule, but generally the building-block development of human knowledge is not perceptibly reflected in word indexing systems. It is somewhat detected in cumulative author indexes. But once a new specific concept has entered the literature, one wants to be able to trace an unambiguous and uncluttered path from its first occurrence to its subsequent occurrence in the literature. In conventional word indexing systems, since the indexer cannot afford to take the necessary time to establish them, these linkages are lost. As a consequence, the literature is treated as a series of isolated events, like molecules of gas. This would account for the use of much probability thinking in the literature as, e.g., the work of Mooers.[18]

Network Model

However, the literature is a rather heavily cross-linked network of interrelated events. These linkages, through citations, are ordinarily provided by authors. Although this is not a perfect system, due to human fallibility, it has a natural redundancy that overcomes the imperfections. The network model of the literature has enabled us to eliminate a rather arbitrary boundary that has hitherto existed between bibliog-

[15] J. F. Crow, "Modifying Man: Muller's Eugenics and Lederberg's Euphenics," *Science*, 148 (1965), pp. 1579-1580. ("It is a measure of their impact . . . that the word has already come into general usage.")

[16] S. Swenson, "Flow Chart on Library Searching Techniques," *Special Libraries*, 56 (1965), pp. 239-242. See also G. Carlson, *Search Strategy by Reference Librarians* (Part 3, Final Report on the Organization of Large Files under NSF Contract C-280), (Sherman Oaks, California: Hughes Dynamics, Inc., 1964).

[17] E. Garfield, "The Illogical Calculus of Information Retrieval." Paper presented at the First Annual Symposium on Biomathematics and Computer Science in the Life Sciences, Houston, Texas, March 28-30, 1963. (Unpublished.)

[18] C. N. Mooers, "Zatacoding Applied to Mechanical Organization of Knowledge," *American Documentation*, 2 (1951), pp. 20-32.

raphy and historiography. The ideal information system will enable the user to move in, around, and through this network with great facility. This is precisely the facility needed by the historian. As we have seen in using the citation network for information retrieval, a key factor is always the identification of suitable starting points. Using the new historical-bibliographical methodology we have created,[19] the creative function of the historian is intensified. His job will be the selection and identification of the starting points; and from there, he will guide the computer to develop that portion of the historical-bibliographical network pertaining to his interests. It is my contention, therefore, that the historian, like the IR system user, is involved constantly in a process of detecting key starting points. These starting points may be called "subjects," "words," "citations," "eponyms," "names," or "events." In the network model, they are all essentially equivalent. It is our symbolic descriptions of the nodes or vertices in the network that vary. For the sake of simplicity, several citations may be clustered together to represent a milestone event (See Figure 1).

Graph Theoretic Model

Conventional bibliography essentially describes the structure of man's accumulated knowledge simply as a neatly piled brick wall. It is primarily descriptive of what man has created—a simple inventory of publications without regard to the interrelationships between the items in the inventory. In contrast, in citation indexing the conception of man's knowledge is a huge graph or network. A graphical description of the literature considers each document as a node or vertex in a huge multi-dimensional network. By analogy, this model of the literature (which I am here equating with man's knowledge) is like a huge road map in which the cities and towns share varying degrees of connectivity. In a recently completed study, Garner[20] has provided a graph-theoretic description of citation networks. His work includes an extremely useful bibliographic notation system which describes the nth degree relationships which exist between the individual documents or sets of documents in the total universe of documents. The notation not only provides a useful shorthand for such questions as "the set of documents, which are cited by the documents, which cite reference X," but the notation will facilitate the transition, in a computer search facility, from question to answer. The notation provides an unambiguous means of expressing complex search parameters when exploring the network or any segment of it.

Historio-Bibliography—A New Methodology

In previous work, I have referred to this very same type of graph as an historical map.[21] Since each document has associated with it a date, it should not be difficult to see why a complete bibliography expressed as a network is in fact an approximation of the history of the subject covered. In conventional bibliography, the mere chronological listing of publications only gives a hint of the historical development of a particular subject. In the citation index, a second vital aspect of historical description is provided. Both conventional and citation indexing provide an inventory of the events. But the citation index and the network that can be obtained from it show the interrelationships among events. In a recently concluded experiment on "The Use of Citation Data in Writing the History of Science,"[22] we have been convinced that this historio-bibliographic model is a legitimate starting point for the historian. The citation-network technique can eliminate a great deal of the drudgery associated with scholarly historical writing. The historian can now devote more time to the evaluation of documents and less to the search and initial preparation of the framework of his historical narrative.

Computer-drawn Topological Historical Maps

When there is sufficient chronological depth to the *Science Citation Index,* it may be possible for the historian of science to instruct the computer to automatically draw complete topological-historical maps. These maps will show the major and minor events that have intervened between specific events selected as the starting and end points. More importantly, the map will show the linkages or the interrelationships between the events selected.

Bibliographies Annotated by Network Coordinates

Similar capabilities may also become available to any library patron in the future. When conducting a literature search, he will receive not only a conventional bibliography, but also suitable notations for each item indicating the interrelationship with other items in the bibliography. In addition, he will receive a graph showing these relationships more clearly. The graph will be drawn by a plotting device attached to the computer. For a relatively short bibliography, this can be done by relatively inexpensive equipment. In fact, a useful map could undoubtedly be prepared by using a conventional typewriter as output once the drawing instructions had been completed on the computer.

Extroverted vs. Introverted Systems

The term "bibliographic systems" has hitherto had a rather static connotation. Old-fashioned libraries have an introverted quality which is quite different from the dynamism, vivacity, and extroversion connoted by terms like "current awareness," "Selective Dissemination of Information (SDI)," or "Automatic Subject Citation Alert (ASCA)." Any bibliographic system has the inherent capability for retrospective search as well as current dissemination. All such terms are relativistic.[23] However, a comparison between systems like Selective Dissemination of Information (SDI) and the citation aspect of Automatic Subject Citation Alert (ASCA), from the user's viewpoint, reiterates the important differences between word and citation systems. If the comparisons I have made regarding retrospective search are not as simple as one would like, this may be attributed to the wide variation in search strategy that is possible from search to search. However, in SDI or ASCA, a fixed procedure is required which demands a self-discipline from the user, to which he often is not accustomed, when constructing his field of interest profile. Incidentally, as the computer is used more often for retrospective search, a similar self-discipline will also be required.[24]

Problems in Constructing Word Profiles

Since he is accustomed to using natural language expressions, the average user thinks it will be easy to construct an interest profile based on words. Suppose he is interested in "thin films." At first he thinks this term is the key to all he needs. Not only does he find later on that the amount of literature on this broad subject may be rather voluminous, but more importantly, many of the significant papers do not include the term "thin films" in the title or the abstract. Subsequently, one may find that he is really primarily interested in the problem of "conduction through thin films," but he *is* also interested in all methods of depositing thin films. This procedure goes on at a varying pace, and some of the more cleverly designed SDI systems[25] allow the user to adjust his profile from week to week. The main point is that few scientists are interested in one precise and neat topic which can be spelled out by a single term or combination of terms. Rice has described some of these difficulties.[26] The average scientist today is not merely interdisciplinary but multi-disciplinary. Once a scientist realizes that he is missing a great deal of literature due to the vagaries of language or that he is getting too much irrelevant literature for the same reason, he may appreciate better the relatively unambiguous character of subject citations which can be used to symbolize his interests in the ASCA system.

[19] E. Garfield, I. H. Sher, and R. J. Torpie, *The Use of Citation Data in Writing the History of Science* (Philadelphia: Institute for Scientific Information, 1964), 76 pp. See also D. J. de S. Price, "Networks of Scientific Papers," *Science*, 149 (1965), pp. 510-515.

[20] R. Garner, *A Graph Theoretic Analysis of Citation Index Structures*, thesis, Drexel Institute of Technology (1965).

[21] E. Garfield, "Citation Indexing: A Natural Science Literature Retrieval System for the Social Sciences," *The American Behavioral Scientist*, 7(10) (1964), pp. 58-61.

[22] See footnote 19.

[23] E. Garfield, "Needed—A Relativistic Theory of Information Science," *Automation and Scientific Communication* (Proceedings of the 26th Annual Meeting of the American Documentation Institute), (Washington, D.C.: American Documentation Institute, 1963), Part III, pp. 419-420.

[24] See footnote 8.

[25] See footnote 5.

[26] C. N. Rice, "A Computer-Based Alerting System for *Chemical Titles*, *Journal of Chemical Documentation*, 5(3) (1965), pp. 163-165.

Citation Profiles

In the ASCA system, citations and authors can be used to compose an interest profile. Citations can symbolize simple one-word concepts, like "euphenics," as well as more complex terms like "conduction through thin films." Citation indexing overcomes the syntactic problems associated with Boolean systems which cannot distinguish between "dog bites man" and vice versa.

The Two Meanings of "Generic"

Citations can be "generic" or "specific." A "classical" paper is bound to act as a "generic" catch-all for a given field, while the rarely cited paper often covers a "specific" topic which is of interest only to a small number of investigators. These are relativistic terms which must be used very carefully. Frequently "generic" is used in a quantitative sense; while at other times, it is used in a hierarchical sense. Some of you may find the use of "generic" in a quantitative sense somewhat strange, but this is what is meant when some people say a generic as well as specific search capability is desired. The taxonomic implication of "generic" in a tree-structured system, such as a Dewey Decimal System, is quite different from the quantitative sense. Classical taxonomy classifies the animal kingdom in hierarchical trees making the term "mammal" generic to "man" and "dog." However, the new classification of numerical taxonomy[27] considers additional biochemical information elements shared by two or more organisms. Since all types of animal life share the same DNA, this will have a profound effect on classical conceptions of relationships.

Man vs. Machine, Intellect vs. Routine

I cannot, in the brief time allotted me, go into an extended discussion of the fundamental questions, "What is thinking?" and "Can the machine replace the man?" There is a great deal of muddled thinking on this problem, and those who use computers are sometimes guilty of confusing the issues further because they tend to over-dramatize. "Computer" is the okay word today. And its indiscriminate use is rarely challenged by librarians. For example, who in this learned audience would challenge the statement that, "Citation indexing is a computer procedure?" Yet nothing could be further from the truth. Indeed, there are no bibliographic applications of the computer, with which I am familiar, in which man's intellect is replaced by the computer. Many procedures have been mechanized which are purely algorithmic. They are operations which, given enough time, a clerk could perform.[28] This does not minimize the value of the computer or the accomplishments of those who have designed systems in which the computer *facilitates* indexing work.[29]

Can Citation Indexing Be Automated?

A few years ago, I was asked: "Can citation indexing be automated?"[30] I replied with a resounding "no." My audience seemed startled. They had just heard me state that computers are used in the compilation of the *Science Citation Index*. But citation indexing epitomizes a kind of intellectual activity that goes far beyond conventional *a priori* subject indexing. A computer may, in part, match the so-called intellectual activity of an indexer who is instructed to select key words in titles.[31] This is really algorithmized indexing. But no computer can perform *a posteriori* indexing. In its best and purest form, citation indexing is *a posteriori*. The author provides an insight—establishes a relationship—that is *a posteriori*, whether he is writing a critical review or an original thesis.

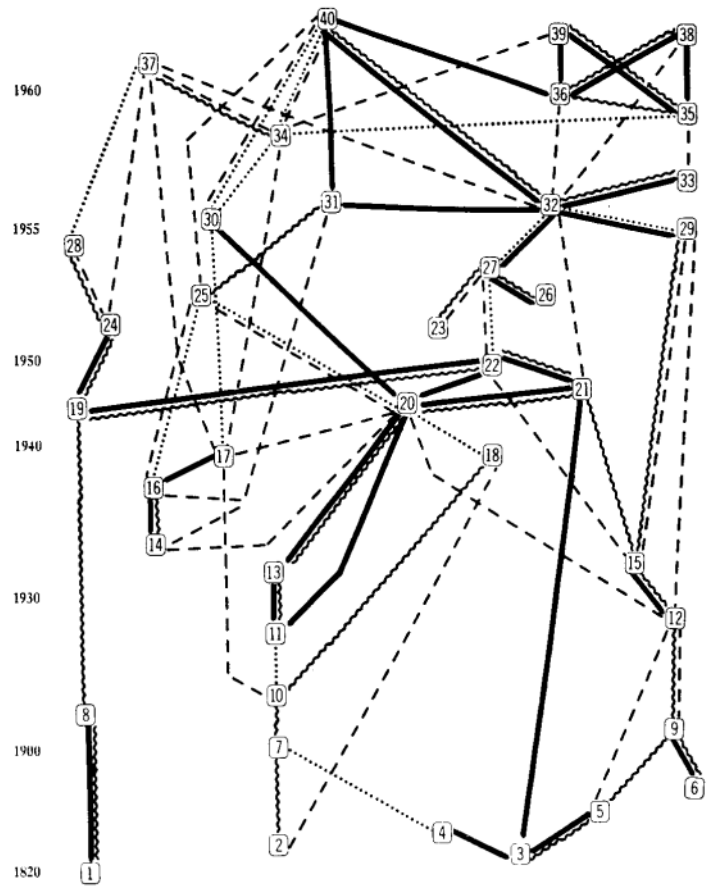
[27] R. R. Sokal and P. H. A. Sneath, *Principles of Numerical Taxonomy* (San Francisco: W. H. Freeman and Company, 1963), 359 pp.

[28] C. Montgomery and D. R. Swanson, "Machinelike Indexing by People," *American Documentation*, 13 (1962), pp. 359-366.

[29] E. Garfield, "Bang Those Robot Heads Together," *Chemical and Engineering News*, 30 (1952), p. 5232. (Letter to the Editor.)

[30] E. Garfield, "Can Citation Indexing Be Automated?", M. E. Stevens, V. E. Giuliano and L. B. Heilprin, eds., *Statistical Association Methods for Mechanized Documentation; Symposium Proceedings, Washington, 1964.* (Washington, D.C.: National Bureau of Standards Miscellaneous Publications No. 269, 1965), pp. 189-192.

[31] See footnote 28.



Network diagram for history of DNA based on Asimov's book, The Genetic Code. Composite of six network diagrams as reported in E. Garfield, I.H. Sher, and R.J. Torpie, The Use of Citation Data in Writing the History of Science (Philadelphia: Institute for Scientific Information, 1964), 76 pages.

KEY

Direct citation connections

Indirect citation connections

Asimov's specified historical connections

Asimov's implied historical connections

NODE	DATE	NAME	NODE	DATE	NAME
1	1820	Braconnot	21	1947	Chargaff
2	1865	Mendel	22	1950	Chargaff
3	1871	Meischer	23	1950-51	Pauling and Corey
4	1879	Fleming	24	1951-53	Sanger
5	1886	Kossel	25	1952	Hershey and Chase
6	1891	Fischer and Piloty	26	1953	Wilkins
7	1900	DeVries	27	1953	Watson and Crick
8	1907	Fischer	28	1953	DuVigneaud
9	1909	Levene and Jacobs	29	1955	Todd
10	1926	Muller	30	1954-56	Palade
11	1928	Griffith	31	1955-57	Fraenkel-Conrat
12	1929	Levene, Mori and London	32	1955-56	Ochoa
13	1932	Alloway	33	1956-57	Komberg
14	1935	Stanley	34	1957-58	Hoagland
15	1935	Levene and Tipson	35	1960-61	Jacob and Monod
16	1936-37	Bawden and Pirie	36	1960	Hurwitz
17	1938-39	Caspersson and Schultz	37	1961	Dintzis
18	1941	Beadle and Tatum	38	1961-62	Novelli
19	1943-44	Martin and Synge	39	1962	Allfrey and Mirsky
20	1944	Avery, MacLeod and McCarty	40	1961-62	Nirenberg and Matthaei

FIGURE 1

From Cancer to Meteorites

As an example, consider a recent paper by Warburg[32] in which he cites Urey's work on the origin of planets.[33] In this paper, Warburg discusses a postulated chemical relationship between the development of cancer and the origin of life on earth! What super-omniscient indexer, at the earlier time he is cataloging the work by Urey, can predict such a correlation? Or similarly, can we expect even a subject specialist to perceive the more obvious relationship between Urey's work on lifelike forms in meteorites[34] and Mueller's work on the carbonaceous content of meteorites?[35] These are after-the-fact observations which few catalogers and no computer can match.

Perfecting the Citation Method

Not all citation indexing is of this kind. Some of it, as we have seen in the example of euphenics, is equivalent to conventional subject word indexing. Let me use this case to illustrate some of the ways we can perfect the citation indexing method. While preparing my talk, a review entitled, "Modifying Man: Muller's Eugenics and Lederberg's Euphenics" appeared in *Science*.[36] By simple title indexing, this work might be found under "Euphenics" in a KWIC, KWOC, and other conventional word indexes. However, there is no direct reference to Lederberg's original paper in *Nature*[37] in this review. There is a reference to the book in which the full talk (of which the *Nature* article is a condensation) is reprinted. By a more elaborate and expensive editing process, our citation editors could have established this direct citation. At present, this is not economically sensible. Unfortunately, the reviewer did not provide the direct citation, and the editor of *Science* did not insist on this improvement. But the case illustrates that citations have properties like words.

Synonymous Citations

There are such things as synonymous citations. In this case, the citation "Lederberg J, 63, *Nature* 198, 428" is synonymous with "Lederberg, 64, *Man and His Future,* p. 263" and with "Wolstenholme, 64, *Man and His Future,* p. 263."[38] (See Figure 2.) The flaw in the present system is simply that we do not process books or short collections of papers as sources. If we did, then the book by Wolstenholme would have been included in the SCI as a source. In this way, the equivalence of the two synonyms would be established whether or not they were cited. However, as you can see, some other author, like myself, will inevitably establish the connection. This is what was meant before by "natural redundancy" in the citation network.

Common Linguistic Properties of Citations and Words

I mention these details not merely to indicate the fully recognized weaknesses in the existing citation index systems. Obviously, in many cases we must rely on a combination search involving word and citation indexes. My prime intention, however, is to demonstrate that citations share properties ordinarily associated with conventional subject headings. These properties need to be extensively studied. In the process, we will learn much about the general problem of subject and content analysis. The work of O'Connor[39] has implied that word indexing, like citation indexing, can involve *a posteriori* indexing. Any indexing that involves the establishment of new correlations on the part of the indexer is *a posteriori*. But this is precisely where the computer is, as yet, unable to match the intellect of man. As soon as it does, then the indexing achieved will be, by definition, *a priori*.

[32] O. Warburg, *et al.*, "Experimente sur Anaerobiose der Krebszellen," *Klinische Wochenschrift*, 43 (6) (1965), pp. 289-293.
[33] H. C. Urey, *The Planets* (New Haven, Connecticut: Yale University Press, 1952).
[34] H. C. Urey, "Lifelike Forms in Meteorites," *Science*, 137 (1962), pp. 623-628.
[35] G. Mueller, "Interpretation of Micro-structures in Carbonaceous Meteorites," *Nature* 205 (1965), pp. 1200-1201.
[36] See footnote 15.
[37] See footnote 14.
[38] J. Lederberg, "Biological Future of Man," *In* G. Wolstenholme, ed., *Man and His Future* (Boston: Little, Brown and Co., 1964), pp. 263-273.
[39] J. O'Connor, "Mechanized Indexing Methods and Their Testing," *Journal of the Association for Computing Machinery*, 11 (1964), pp. 437-449.

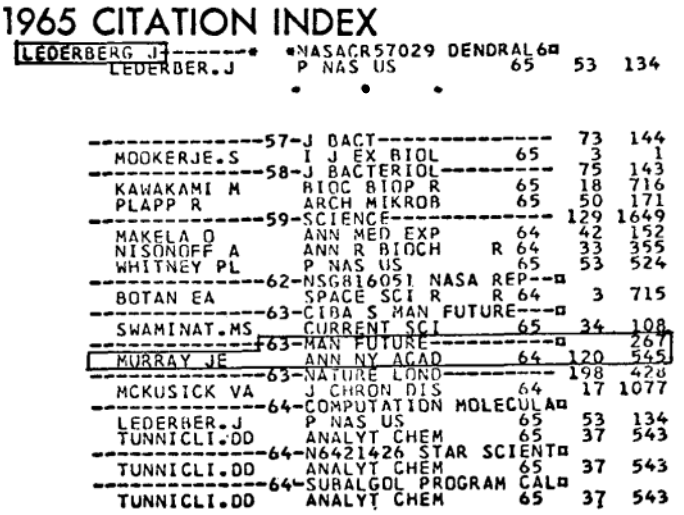


FIGURE 2. Sample of 1965 *Science Citation Index* illustrating synonomous citations.

Near Synonyms

The comparison between conventional and citation indexing systems can be extended from pure synonyms to near synonyms. In a citation index, we ultimately could establish a cross-reference structure that would instantly alert the reader to key papers on closely related topics. However, the equivalent effect is usually achieved by authors themselves. The user sometimes uncovers these "cross references" in a cyclic search in which cited paper X leads to a citing paper which, in turn, leads to cited paper Y, etc. Papers X and Y may be near synonyms. (See Figure 3.)

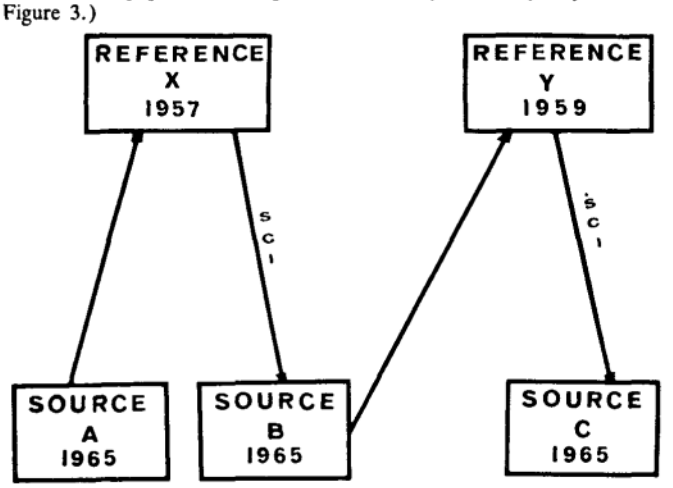


FIGURE 3. Illustration of cycling. Source A cites Reference X leading via SCI to Source B which in turn cites Reference Y. Reference Y then leads via SCI to Source C.

Is a Perfect Bibliographic System Possible?

If we lived in a perfect world, could every paper published be relied upon to provide a "perfect" bibliography? To provide a perfect bibliography, one needs a perfect retrieval system. This is a vicious circle! To achieve a perfect retrieval system through word indexing is probably impossible and certainly expensive. Similar ideas simultaneously discovered will always be expressed in various languages and in various nomenclatures. "Perfection" through citation indexing may not be practical for several years, but our present efforts appear quite satisfactory for the costs involved and the results achieved. To offset the deficiencies associated with real retrieval systems, we compromise by using a combination of approaches. However, since I assume natural language is to

be used in advanced systems, we must inevitably bridge the gap between words and citations. This may not be a serious problem for the scientist as we have seen. He knows the classical references—sometimes even remembering page numbers; but the librarian may not. The librarian or the student needs a link between the word and its citation synonym. That link, at present, is provided by the basic reference armamentarium of dictionaries, encyclopedias, card catalogs, word indexes, and the scientist himself.

Dictionary of Key Citations

But we could provide a more direct link. We could establish a dictionary of key citations that would contain as a typical entry "Euphenics see Lederberg, 63, *Nature* 198, 428" and "Euphenics see also Wolstenholme, 1964, *Man and His Future,* 263–273." The dictionary might also contain an entry for *"Origin of Planets* see Urey, 1952." From this last example, it can be readily seen why the title entry in the card catalog may be a useful and sufficient step for beginning a search of the SCI. The proposal by Tukey[40] that a permuted title index accompany a citation index is an attempt to satisfy this facet of a multifaceted system. But one should not confuse the value of a permuted title index for current literature with the similar need for the "older" literature which is the real problem on beginning a search for a starting point to a citation index. A word index to this year's literature does not provide a starting point for last year's or this year's citation index. However, a word index to last year's literature can help in using a current citation index.

Streamlining the Indexes

There is an important role this new dictionary of key citations could play in future word indexing. Through this master dictionary, we may eliminate a great deal of repetitious word indexing. If a new document enters the system, the indexer could, using the random-access file described earlier, enter the title and citations just as we do now. The computer would compare the title words, or any words chosen by the indexer, with the dictionary. The synonymous citations would then be compared with the citations chosen by the author. The computer would then list words or phrases not yet included in the dictionary. For example, suppose a title appears "Engineering Human Development." Examination of the paper might reveal a citation to Lederberg's euphenics which *is* the engineering of human development. In this way, a cross-reference to euphenics and/or its synonymous citation "Lederberg J., 63, *Nature* 198, 428," could be established.

ISI is planning an experiment along these lines in which suitable cross-references between key words and key citations are established for the 1,000 papers most frequently cited in 1961. This list of papers, and others like it, will be of great value to librarians, publishers, and others because it will identify works that ought to be readily available in libraries and have had great impact on scientific work. A frequently cited work is inevitably a frequently requested work.

Consider the efficiency of the total indexing complex when we know that a single paper by Lineweaver and Burk[41] was cited several hundred times per year during the last decade. Instead of creating an index term 'The determination of enzyme dissociation constants," indexers would avoid such an entry knowing that a citation to Lineweaver is provided in the bibliography.

The ultimate dictionary of key citations can significantly alter conventional word indexing. The indexer would look for the gaps in the author's bibliography by comparing the terminology employed by the author with that connoted by his reference citations. If the word "euphenics" is prominent in the title, abstract, or text, and a citation synonym for euphenics is not one of the papers cited, then this indicates the need for that citation. If it is cited, then there is no need for a word entry if the word is already in the master dictionary. If it isn't in the dictionary, then there is a high probability the needed reference will be cited; if not, it should be and the new term added to the master file.

[40] J. W. Tukey, "Keeping Research in Contact with the Literature: Citation Indices and Beyond," *Journal of Chemical Documentation,* 2 (1962), pp. 34-37.

[41] H. Lineweaver and D. Burk, 'The Determination of Enzyme Dissociation Constants," *Journal of the American Chemical Society,* 56 (1934), pp. 658-666.

Indexing New Chemicals Epitomizes the Combined Word-Citation Method

This process is epitomized in the *Index Chemicus* registry system which by-passes the expensive repetitious indexing of chemicals as done by *Chemical Abstracts.* Since a new chemical compound, by definition, cannot be found in the literature, else it would not be new, a molecular formula and/or name entry must be created for it. If it is an old compound, its previous occurrence in the literature will or should be cited. In this case, there will be no problem retrieving the old and new information on the compound, either by use of the Cumulated Formula Index, the Citation Index, or the Word Index. The efficiency of this procedure, as compared to the method used by *Chemical Abstracts,* is not trivial.

The advantages of the approach suggested here will be less obvious as one departs from the so-called systematized nomenclature of chemistry (sic) and enters the domain of the life and behavioral sciences where concepts may be expressed in word phrases not easily susceptible to direct look-up in a dictionary. An example would be "Engineering Human Development," a synonym created by J. F. Crow for "euphenics." A phrase like this is ambiguous and would inevitably be confused with topics other than euphenics. However, it is precisely this intellectual differentiation that we expect from a human indexer on whom we rely to make the appropriate choice when confronted with homonymic expressions.

Indeed, it is this negative role—to say what a topic is not—that has not been stressed in evaluating human indexing. A KWIC index system may employ a stop-list to prevent articles from being indexed under such expressions as "the" and "etc." but it is precisely the rare occurrence when "etc." has significance that we want it retained.

Similarity and Coupling

I would like to conclude my comparison of word and citation indexing systems by introducing the concepts of similarity[42] and bibliographic coupling.[43] Several years ago, I published what I consider to be one of my best papers.[44] It is somewhat distressing to me as a citation indexer to find that this paper is rarely, if ever, cited. Evidently the paper has made little impact. In that paper, I tried to dramatize its significance by the statement that, "The information content of the Woodbury Public Library was possibly the same or higher than that of the Library of Congress." The information content of a library is not a function of the number of books shelved, but rather a function of the indexing or cataloging done to produce the catalog. If two libraries are using the same indexing technique and subject heading authority list, each may contain essentially the same information content even though one may contain more books than the other. What matters is the probability of occurrence of descriptors. In short, a library with many duplicates contains no more information in its index than a library with no duplicates.

Descriptor Coupling

Information content is a function of the probabilities of occurrence of each descriptor. The document is defined in the search system by the set of descriptors or headings used to catalog the document. Theoretically, the real document does not exist in the search system. If one document is described by a given combination of descriptors and another document is described by the same set of descriptors, the two documents are equivalent. And surely, if the indexing procedures are

[42] A number of workers have used the term "similarity" with varying implications, including:

a. M. E. Maron and J. L. Kuhns, "On Relevance, Probabilistic Indexing and Information Retrieval," *Journal of the Association for Computing Machinery,* 7 (1960), pp. 216-244. (As cited in Becker and Hayes, p. 144.)

b. A. F. Parker-Rhodes and R. M. Needham, *The Theory of Clumps,* CLRU Report, February, 1960. (As cited in Becker and Hayes, p. 371.)

c. J. Becker and R. M. Hayes, *Information Storage and Retrieval: Tools, Elements, Theories* (New York: John Wiley & Sons, Inc., 1963), pp. 236, 144, 371.

d. G. Salton, "Associative Document Retrieval Techniques Using Bibliographic Information," *Journal of the Association for Computing Machinery,* 10 (1963), pp. 440-457. (Quote from p. 456.)

e. H. Borko and M. D. Bernick, "Toward the Establishment of a Computer Based Classification System for Scientific Documentation," *Technical Memorandum TM-1763* (Santa Monica, California: System Development Corporation, 1964), 47 pp.

f. R. R. Sokal and P. H. A. Sneath as cited in footnote 27.

[43] M. M. Kessler, "Bibliographic Coupling Between Scientific Papers," *American Documentation,* 14 (1963), pp. 10-25.

[44] E. Garfield, "Information Theory and Other Quantitative Factors in Code Design for Document Card Systems," *Journal of Chemical Documentation,* 1 (1961), pp. 70-75.

the user. For the same cost, he might prefer another alternative to large scale storage of information—microforms. And do not forget that, at present, at least 15 seconds would be required to ask the computer a question, during which time you might be finished using a book!

Microforms

Microforms offer several systems advantages over computers and printed books. Reduced storage space is the most obvious advantage. As yet, microforms do not always provide significant overall cost advantages of libraries[9] when all factors are considered. But cost is not the only consideration—especially when considering the systems needs of the small library or an individual. The prime audience for ISI's services is the individual scientist. He has a personal library. He would like his library to be as complete as costs and space will allow. Consequently, his acquisitions policy is dictated by his budget and the physical limitations of his office. He prefers the printed book, but if he had a really practical microform system, he might readily compromise preferences born of habit and culture. If we reduced his look-up time, this would help a lot. Using microforms we could design bibliographic systems that would significantly affect average look-up time. A comparable reduction would be obtained through printed books, but this may not be practical.

Increase Index Space to Reduce Search Time

In the *SCI Source Index*, as in the author indexes to the *Index Medicus, Chemical Abstracts*, etc., we cross-reference secondary source authors to primary authors. This almost halves the space required to provide complete citations for all authors since the average number of authors per paper exceeds two.[10] Similarly, if we included a full bibliographic citation for each entry in the *Citation Index* proper, its size would easily quadruple. Instead of eight large volumes per year, the SCI would need 32 volumes.

If we enlarged the typography in the *Citation Index*, using a three-column instead of a four-column format, we would double the size once again. Consequently, the *Science Citation Index* would require over 60 large volumes. If we doubled our coverage, which is our five-year objective, we could then exceed 100 volumes per year—the equivalent of the *Encyclopediae International, Britannica, Americana*, and *Collier's* combined.

However, when using microforms, an increase in the number of reels of film, while not trivial, is not as alarming. In addition, the technology is still being improved so as to increase storage per cubic foot by either using higher reduction ratios and/or larger magnifications in the camera and viewer optical systems. Indeed, in preference to a time-shared microform facility—that's what a library with one microfilm viewer is—we might prefer to use several viewers and a relatively large number of reels of film or microfiches to reduce the chances for a busy signal.

In the new format made practical by the use of microforms, however, the average search time would be drastically reduced, and user satisfaction would thereby increase. For this reason, even in its present state of technological development, microforms have much to offer the user of an index. However, as with the printed book, there is as yet no solution to the problem of day-to-day updating. An erasable random-access computer memory does resolve this problem. We can't afford to issue and use new films or books each day. We may be able to afford daily computer updating.

Combined Use of Computers, Microforms, and Books

A solution to the dilemma may be found in the combined use by libraries of time-shared computers for the current year's material, printed indexes for the last three or four years' material, and microforms for very large scale cumulative indexes. By contrast, the individual scientist may be served better by microform even for last year's material, printed quarterly or monthly supplements, with weekly updating for special personalized supplements prepared by the computer as in our ASCA (Automatic Subject Citation Alert) system.

Sophisticated Hardware Requires Large-Scale Information Problems

It should be evident that *Memex* is merely the symbolic representation for the best in all the hardware systems I have discussed above. Probably the closest we have come to *Memex* is the AMFIS system,[11] which subsequently took form as CRIS. However, sophisticated hardware developments, like CRIS, will not be used unless we also have comparatively sophisticated large-scale information reservoirs to manipulate with the hardware. The rare book or manuscript will not usher in the microform revolution; large compilations like *Science Citation Index, Chemical Abstracts*, and *Index Chemicus* will.

Having considered the hardware—the *Memex* view of the information problem—let me now discuss the more significant intellectual problem symbolized by the *World Brain*. While this topic has been given many names, it is the perennial problem Tauber and others called "subject analysis."[12]

What Is a Subject?

One of the most frequently expressed criticisms of the citation index is that it is not a "subject" index. What is really meant, however, is that the citation index is not a "word" index. Therefore, "What is a subject?" is fundamental not only in evaluating citation indexing, but all types of systems for subject analysis. In establishing why a citation index *is* a subject index, some important insights to conventional subject and content analysis also are obtained.

Primordial Word-Document Events

Theoretically every word in the dictionary could be traced to an important historical event—the time and place that word first occurred. Especially in science, newly coined words can usually be traced to, or identified with, a particular paper or book by an individual author. And frequently, the main theme or subject of this primordial document is the same topic symbolized by the new word.[13]

Example of Lederberg's Euphenics

For example, in 1962, Professor J. Lederberg coined the word "euphenics" which first appeared in a paper entitled, "Molecular Biology, Eugenics, and Euphenics" published in *Nature* 198, 428.[14] I think you would agree that the "subject" of that paper is indeed euphenics. 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 identifies the document in which it appeared. The word "euphenics" and the citation "*Nature* 198, 428" are both symbols. They are equivalent symbols for the topic discussed in Lederberg's paper. The subject matter of the paper is the same whether we symbolize it by the word "euphenics" or the short citation "*Nature* 198, 428" or the citation "Lederberg J 63 *Nature* 198, 428" as it would be identified in the *Science Citation Index*.

Now suppose that some other author discusses the subject of euphenics in a subsequent paper. It is the usual custom in scholarly research, when using new terms, to provide a footnote reference to the source of the new term. As a result, in a citation index system, the second citing paper is indexed under the term "Lederberg J, 63, *Nature* 198, 428."

If the "main theme" of the second paper is euphenics, one expects the document to be indexed under the term "euphenics" in a conventional word indexing system. Both methods have achieved identi-

[9] a. H. Marron, "Science Libraries—Consolidated/Departmental?" *Physics Today*, 16 (7) (1963), pp. 34-38.

b. A. B. Veaner, "Microtext Materials in Libraries," *Journal of Medical Education*, 40(1) (Part 1) (1965), pp. 43-45.

c. A seemingly opposite point of view has been taken by V. W. Clapp. *The Future of the Research Library* (Urbana: University of Illinois Press, 1964), pp. 18-23, when he states that "the real advantages should derive not from space saving, but from inexpensiveness of dissemination," (p. 19).

[10] B. L. Clarke, "Multiple Authorship Trends in Scientific Papers," *Science*, 143 (1964), pp. 822-824.

[11] E. A. Avakian and E. Garfield, "AMFIS—The Automatic Microfilm Information System," *Special Libraries*, 48 (1957), pp. 145-148.

[12] M. R. Tauber, ed., *The Subject Analysis of Library Materials* (New York: Columbia University School of Library Service, 1953), 235 pp.

[13] For an interesting discussion of citation and cognate "historical" methods see R. A. Fairthorne, *Towards Information Retrieval* (London: Butterworths, 1961), p. 192.

[14] J. Lederberg, "Molecular Biology, Eugenics and Euphenics," *Nature*, 198 (1963), pp. 428-429.