Interview with Eugene Garfield, Chairman Emeritus of ISI, Philadelphia, U.S.A.

Citation Consciousness

The Origins of Citation Indexing in Science

Wolfgang G. Stock: In the early 50s you joined an indexing project of the medical literature at the Welch library of Johns Hopkins University. This project led to early versions of medical databases, later called MEDLARS resp. MEDLINE, and to an indexing method using controlled terms or the Medical Subject Headings (MeSH). This kind of indexing was done (and is still done) by professionals reading scientific articles and using the tool of a thesaurus. Did you support this idea of intellectual indexing or did you see problems using thesauri or other documentation languages?

Eugene Garfield: The use of Medical Subject Headings (MESH) had been established at the *Current List of Medical Literature* before I came to the Welch Project in 1951. One of our tasks was to categorize the list of terms and to help make recommendations for their improvement. As a chemist, I was also given the task of standardizing chemical nomenclature. During World War II, the *Current List* was simply a typewritten compilation of medical journal contents pages. Later on it was expanded. Subject and author indexes were added. But afterwards the name and format was changed to the *Index Medicus*. The format of the AMA's *Quarterly Cumulative Index Medicus* was adopted. The use of MESH, however, was continued and is used essentially in the same way today. MESH aids human indexers who select appropriate subject headings to describe the main contents of medical articles.[1]
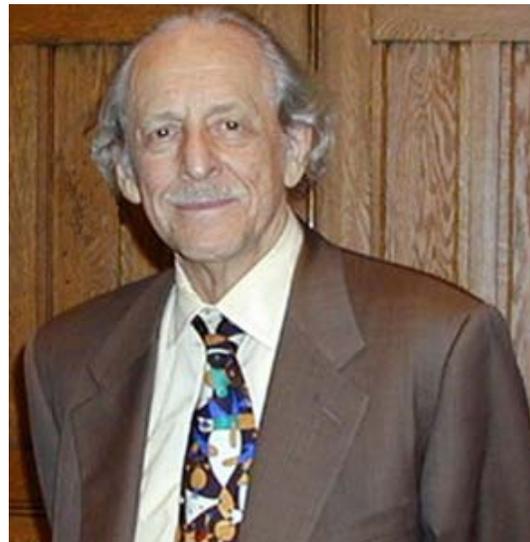
I recognized the problems in this system quite early. The original concept of publishing contents pages from current journal issues, as was done during World War II, had been abandoned in the quest to include subject and author indexes. Thus rapid timing, essential to researchers, was completely sacrificed to improve the retrieval functions important to librarians. The current awareness of the CLML and natural language was lost and an imperfect system of human indexing was substituted. Furthermore, the sheer volume and size of the monthly issues made for significant technological problems in assembling the indexing materials. So I attempted to alleviate these problems by designing a new punched card system for producing the indexes.[2] But no matter how fast the mechanical problems could be solved, the delays due to human indexing could not be. The multi-lingual nature of the project also was a factor although the excellent indexers at the NLM could handle many of the languages involved.

Controlled vocabulary has its advantages but at the same time so does natural language. Like formal classification systems, controlled vocabularies change very slowly in response to the rapid changes in science. Recognizing the need for current vocabulary when I introduced *Current Contents* weekly, we mechanized the index compilation with punched cards and computers so that no time was lost. And we relied on natural language title words. In addition to that, we introduced the use of author's addresses, which became a boon in the worldwide exchange of about ten million reprints per year. We continued to use natural language later on when we introduced the *Permuterm Indexes* for the *Science Citation Index*.[3]

Stock: The Welch project organized a Symposium on machine methods in scientific documentation. It is told that William C. Adair, a former vice-president of Shepard's Citation Index, came in contact to the project. Was that a starting point of a scientific citation index?

Garfield: It all depends upon your definition of starting point. I was on the way to discovering a new method of indexing based upon using sentences in review articles to create indexing statements automatically. When W. C. Adair wrote to the Project, after he heard about the famous March 1953 Symposium in the newspapers,[4] he told me about *Shepard's Citations* which I examined at a public library in Baltimore and realized that this was the structure I needed for my review index. But there are vast differences between legal citators as they were called and science literature indexes.

Stock: *Shepard's* uses descriptions and evaluations of citations. There we find notes about history of case (e.g. parallel citation, affirmed citation or reversed citation) and treatment of case (e.g. critized, explained or void). Your method of a scientific citation index only marks the information relation between a document A (which references to an earlier document B) and the document B (which is cited by A), but no history or treatment. Why not?

Garfield: *Shepard's* method is by no means an automatic process; attorneys are used to characterize the type of citations involved in specific court cases. The volume of legal material covered, however, is an order of magnitude smaller than the scientific literature. To introduce an intellectual effort in describing the reason for each cited reference in science would require enormous effort, delay the indexing, and be subject to human variation. You must keep in mind that ISI indexes over one million published items per year. A possible modern solution to "explaining" a citation is to use the context of the sentence in which the citation occurs. Even then it is not always possible to characterize a citation unless one includes an entire paragraph. And other citations are quite general as with book citations. However, in the *SCI* we found that the title of the citing paper was a powerful discriminator when looking at a long list of citations to the same work. Differentiating citations is really only essential when there are many citing works involved.
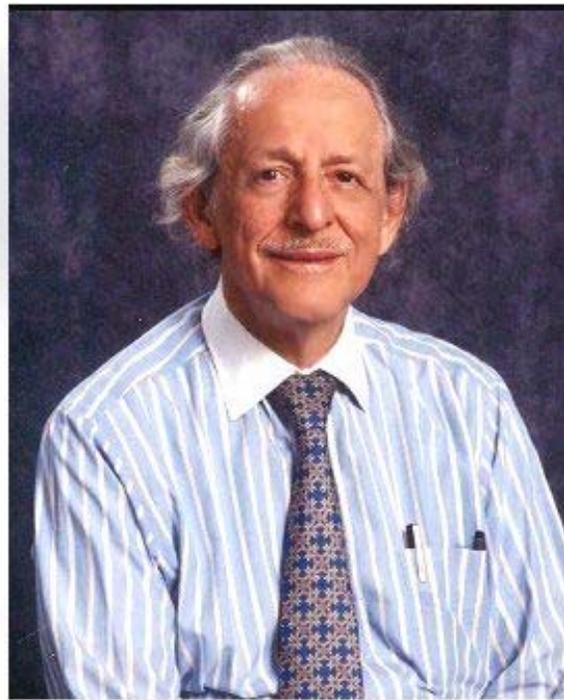
Stock: After you had finished your studies of library science you published an article in 1955 about the idea of a citation index of scientific articles in *Science*.[5] Six years later, in 1961 - a first result (the *Genetics Citation Index*) and then in 1964 the first commercial product *Science Citation Index* was published. Can you tell something about the transition of the idea from 1955 to the information product of 1964! What were the influences of Joshua Lederberg? The enterprise was done by a private company, Eugene Garfield Associates, changed later in 1960 to the Institute of Scientific Information. How did you solve financial problems?

Garfield: The history of these events is described in detail in many publications. Dr. Joshua Lederberg wrote to me in 1959 shortly after he received the Nobel Prize.[6] This eventually led to an NIH financed project to create a *Genetics Citation Index*. The pilot project tested several variants including a comprehensive multidisciplinary index of about 600 leading 1961 journals. This involved about 1,400,000 million cited references taken from about 100,000 articles. As defined by a scientific advisory group, all "genetics" related citations were extracted to produce a 1,000 page index which was distributed to about 1,000 scientists for evaluation. The field of molecular biology was about five years old and it was obvious that the need for the new genetics was multidisciplinary input.

The National Science Foundation refused to print and distribute the main file so we made the brash decision to do so. However, we had to launch an up-to-date product as well. *SCI* was started as a quarterly index in 1964. That decision almost bankrupted what was a small but profitable firm selling *Current Contents*. The response of the library community was extremely slow. So we made a deal with a group of Wall Street investors who acquired a 20% interest in the firm for a convertible debenture loan involving $500,000. As it turned out, we could have managed with ordinary bank financing but it was then too late, although the amount of financing needed was 1,000 times the original $500 I borrowed to finance *Current Contents* in the late fifties.[7]

Stock: The initial input for the *Genetics Citation Index*, was pretty small - "only" 613 journals covered and "only" 1.4 millions citations processed. The *Web of Science* today covers more than 8.500 journals. In the *Genetics Citation Index* there probably were no problems of identifying authors by their last name and the initial letters of their first names. But in a large multidisciplinary citation index a problem rises. The German name "Schmidt M", for example, stands for several researchers in physics, chemistry, clinical medicine, biology, space physics, and engineering. In informetric studies we have difficulties to separate the different scientists. Looking back, was it a mistake not to cover the full name of a person and not to identify clearly the single researcher?

Garfield: The decision to omit first names of authors was dictated by necessity. First of all, there was the restricted space of the 80 column punched card. Second, many, if not most, science journals did not use first names in the byline and more importantly they almost never used them in cited references. Some British journals did not even cite initials – just the last name. The practice of using first names is much more widely accepted in social science and humanities journals. Further, the inclusion of author addresses in the *SCI* made it possible to separate the works of homograph authors in most cases. Today, there are many ways that these problems could be resolved but it would take a massive effort to modify the already existing records. Of course, it would have been preferable but the expense would be enormous and as I have said the information is not universally provided. Add to this the complexity of different alphabets and naming systems. But the key point often forgotten is that the article and not the author is the basis of the system – so there never is any doubt that whatever article by M. Schmidt you are discussing, there is only one article in journal X, Volume Y, page Z.

Citation Analysis in Sociology, History, and Philosophy of Science

Stock: The *Science Citation Index* (*SCI*) is as well a bibliographic tool as an empirical basis for science studies. The sociologist Robert K. Merton says, "It is our collective good fortune that Gene Garfield initially designed the *SCI* in the first instance as an effective and wide-ranging tool for bibliographical search ..., and thus conferred an immensely valuable research tool upon his sociological and historical colleagues". Did

you realize this meta-scientific power of the *SCI* in the 60s? What role did Professor Merton play in the development of citation analysis as a tool for sociology and history of science?

Garfield: Yes, there were brief allusions to these ideas in my first papers on the *SCI* and on the topic of "Unified Indexes to Science."[8] The sociological and historical implications were much more pronounced by the 1960's. I had already known about J.D. Bernal from high school and later about the science of science from the UK.[9] This is also related to my friendship with Maurice Goldsmith.[10,11] I was the one who initiated contact with Derek Price and Merton in 1962 when I sent them samples from our *GCI* project listing citations to their work. Later when I met Professor Merton and his colleague and wife Harriet Zuckerman, my understanding of sociology grew. As it was once expressed, I was a sociologist of science but didn't know it. I had also known about the historical methodology of Von Ronke and it is the historical importance citation indexing that still pervades my interests. This was originally expressed in the work on "Use of Citation Analysis in writing the History of Science."[12]

Dr. Merton's role has been invaluable. After all, it was his students Jonathan and Steven Cole who published key works on the validity of citation analysis. I have drawn on Merton's writings in dozens of articles. I regard him as one of the chief mentors in my life. My youngest son is named Alexander Merton Garfield. Of course the same must be said of Professor Lederberg, for whom my second son Joshua was named.

Stock: Derek J. de Solla Price in his "Little Science, Big Science" assumed that we should watch more the books of scientists than only their articles when we analyse invisible colleges. And the philosopher of science Thomas S. Kuhn declares that the principles of "normal science" are codified in scientific books, precisely in textbooks for students. Now it is true that more and more scientists publish articles and partly even only articles. This kind of scientific literature is represented by ISI's Citation Indexes. But ISI does not cover books. Is it a problem for the purpose of science evaluation that we cannot watch scientific books? In my opinion, there are two ways of scientific impact. One way leads to collegues, to other scientists. This kind of impact is marked by citations in articles published in journals resp. proceedings and well documented by Citation Indexes. The other way leads to students and other beginners. This the world of textbooks with their task to reproduce normal science. Isn't it a strong scientific influence that one is cited in a textbook, is it?

Garfield: Your question reflects a common misperception about books in the ISI citation indexes. Citations to books in all the ISI indexes are quite prevalent, especially the *Social Sciences Citation Index*. I have published studies of highly cited books. So the statement that ISI does not "cover" books or monographs is inaccurate. They are not comprehensively covered as sources. ISI does cover many monographic series. For example, the festschrift *Web of Knowledge*[13] is a book in the ASIST monographic series that is covered as a source. However, I would agree that most single authored and other

monographs commonly known as books are not and should eventually be covered in a citation index. Doing so is not a trivial task and would be equivalent to processing a huge number of review articles, each of which often contain thousands of cited references.

It is estimated that 15% of cited references in *SCI* are non-journal items – mainly books, patents, dissertations, and technical reports. This is in addition to the many book reviews covered as sources. In the *Social Sciences Citation* Index, 50% of cited references are books. So one find thousands of citations to books cited in research and review articles. Adding books as sources would not change the rankings of books when judged by citation frequency since they are cited so frequently in the journal literature. In the humanities this might not be true since only about 20% of cited references are to journal articles – 80% are to books, manuscripts, and other non-journal materials.

Nevertheless, I would favor adding books as sources to the citation indexes because it would be valuable to retrieve relevant book chapters when doing a literature search. My own book *Citation Indexing* has been cited hundreds of times. Will its 'status' change if I learn that it is cited in a few dozen books on information retrieval? Probably not. I do not agree that being cited in a textbook is any greater indicator of influence than being cited in review or research articles.

Stock: Otto Neurath, an Austrian philosopher of the Vienna Circle, spoke about the programme of a unified science, arranged as an encyclopaedia written in a physicalistic language. Henry Small adopts the Neurath idea and says that the multidisciplinary Citation Index is a practical instrument for creating a view upon the unified science. Is the Citation Index with its pathways through science really the empirical solution of Otto Neuraths philosophical thesis? Is citation indexing not only an instrument of sociology and history of science, but of philosophy of science, too?

Garfield: I would agree that the *SCI* is the empirical solution to the proposal of Otto Neurath. In 1958, I referred to Neurath's work in the opening remarks for my paper at the International Conference on Scientific Information.[7] However, I had been familiar with the Vienna school much earlier.

And I do believe that citation indexing is the expression of a philosophical approach. I call it citation consciousness. It is a way of thinking about scholarly work and it should be quite evident to the German information community that this goes back ultimately to the documentation philosophy of Von Ronke. I believed he introduced, at a conscious level, the idea of documenting one's sources for writing history. Historiography and information retrieval in my view are inextricably interlinked. My most recent research continues the work in algorithmic historiography that I started with Dr. Irving Sher in the sixties, but this is not the place to expound on it.[14,15]

From ASIS to ASIST

Stock: In the years between 1998 and 2000 you have been president of the American Society for Information Science (ASIS). In this time the name of this society changed. Now it is called American Society for Information Science and Technology (ASIST). ASIS had had serious problems with the decreasing number of members. Could this trend be stopped by the new name - and a new society's policy?

Garfield: The change in name occurred just prior to the formation of a new SIG in ASIST about information architecture. This increased membership and reversed the downward trend. Obviously I believe the name change is positive and hopefully will lead to financial and membership stability. One of the problems of ASIST is that our profession has been too successful and has spawned many competing groups. A different kind of leadership in the past might have captured all these new groups as they were spawning, but ASIST lacked that kind of leadership. And there are other factors which affect the ability of the Society to grow but it is now up to the younger generation of informationists to decide whether ASIST will remain small or whether it will grow into a broader and internationally more influential organization.

Stock: When we look to the not very successful development of the ASIS in the last years (in Germany we have the same - negative - development), is there a future for information scientists? What are the core competencies of information scientists so that they survive in the information society and information economy - in a society in which information skills are actually necessary?

Garfield: Of course I believe there is a future for information scientists of many kinds. A discussion of core competencies for information professionals is interpreted differently in the many different institutions that teach the subject, but I cannot take the time to get into that arena. The basic competencies in mathematics, language, etc. are the same for all scientists as are information skills. This is not my area of expertise, although I did teach information retrieval at the university at one time.

Stock: Let's have finally a look on the information industry. In the last years we watch a concentration process of companies to very few, very big global players. Imagine, for example, the market power of Dow Jones, Reuters, Dun & Bradstreet, Reed Elsevier and - of course - the Thomson Corporation. In your opinion, do chances or risks prevail for an all in all market development of the information economy?

Garfield: The concentration of power in the information industry is frightening to some. Eventually some of these companies will invite regulation as monopolies if it is perceived that there is no competition or if these conglomerates abuse their power. But in spite of their size there will always be entrepreneurs who will develop competitive niches that will succeed in spite of the large firms. And most of these companies will eventually respond to user demands. That is why we see so many library purchasing consortia forming.

Stock: Dr Garfield, thank you very much for this interview. Have you any suggestions for our German information professionals?

Garfield: As an internationalist, I would not advise my German colleagues differently than those in my own country. First of all they are members of the multi-lingual European community. So your question must be based upon language and culture. German scholars should cherish their historical reputation and tradition of providing complete and accurate documentation. I deplore the failure of so many German journals to include English summaries, especially since they mostly speak English. The advent of the WWW makes it possible for them to include English translations along side their German text. That would foster the international use of their work.

I am often surprised at the level of misunderstanding of German editors about citation analysis and the impact factor.[16] There is a great need for national journals written in German and other European languages, but it is absurd to expect them to reach the same level of impact as international journals, unless they adhere to the same standards as leading journals and attract significant original research. German scientists understand that fact of life and for this reason they mainly publish their best work in English in international journals.

[1] Garfield, E. "The Preparation of Subject-Heading Lists by Automatic Punched-Card Techniques." *The Journal of Documentation* , 10(1):1-10 (March 1954). Reprinted in *Essays of an Information Scientist*, Volume 6, pages 444-453. Philadelphia: ISI Press (1984).
http://www.garfield.library.upenn.edu/essays/v6p444y1983.pdf

[2] Garfield, E. "The Preparation of Printed Indexes by Automatic Punched-Card Techniques." *American Documentation*, 6(2), p.68-76, 1955. Reprinted in *Essays of an Information Scientist*, Volume 6, pages 459-467. Philadelphia: ISI Press (1983).
http://garfield.library.upenn.edu/essays/v6p459y1983.pdf

[3] Garfield, E. "The *Permuterm Subject Index:* an autobiographical review ," *Journal of the American Society for Information Science*, 27 (5/6):288-291 (1976). *Essays of an Information Scientist*, Volume 7, pages 546-550 (1985).
http://www.garfield.library.upenn.edu/essays/v7p546y1984.pdf

[4] Hyslop, M. "Documentalists consider machine techniques," *Special Libraries*, 44:196-8 (1953).

[5] Garfield, E. "Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas." " *Science*, 122(3159): 108-11 (July 1955). Reprinted in *Readings in Information Retrieval*, (H.S. Sharp. Ed.), p.261-274. Metuchen, N.J.: Scarecrow Press, Inc. (1964).
http://www.garfield.library.upenn.edu/papers/science_v122(3159)p108y1955.html

[6] see Lederberg-Garfield correspondence on NLM web site:
http://profiles.nlm.nih.gov/BB/Views/AlphaChron/alpha/10010/10007/10013/
 and also Eugene Garfield website: http://www.garfield.library.upenn.edu/lederberg/list.html

[7] Garfield, E. "How it all begain – with a loan from HFC. *Current Contents* 3:5-8 (January 21, 1980). Reprinted in *Essays of an Information Scientist,* Volume 4, pages 359-362. Philadelphia: ISI Press (1981).
http://www.garfield.library.upenn.edu/essays/v4p359y1979-80.pdf

[8] Garfield, E. "A Unified Index to Science," *Proceedings of the International Conference on Scientific Information. 1958. Washington, D.C.*: National Academy of Sciences - National Research Council, 1, p.461-74, 1959. Reprinted in *Essays of an Information Scientist*, Volume 2, pages 672-688. Philadelphia: ISI Press (1977). http://garfield.library.upenn.edu/essays/v2p674y1974-76.pdf

[9] Garfield, E. "J. D. Bernal – The sage of Cambridge – 4S Award memorializes his contributions to the Social Studies of Science," *Current Contents* No. 19, pages 15-17 (May 10, 1982). Reprinted in *Essays of an Information Scientist*, Volume 5, pages 511-523 (1983).
http://www.garfield.library.upenn.edu/essays/v5p511y1981-82.pdf

[10] Goldsmith, M. *The Science Critic*. New York: Methuen, 217 pgs. (1987).

[11] Garfield, E. "From Information Scientist to Science Critic," *Current Contents* No. 36, pages 3-7 (September 4, 1989). Reprinted in *Essays of an Information Scientist*, Volume 12, pages 251-255. Philadelphia: ISI Press (1991).
http://www.garfield.library.upenn.edu/essays/v12p251y1989.pdf

[12] Garfield E. "J. D. Bernal – The sage of Cambridge – 4S Award memorializes his contributions to the Social Studies of Science," *Current Contents* No. 19, pages 15-17 (May 10, 1982). Reprinted in *Essays of an Information Scientist*, Volume 5, pages 511-523 (1983).
http://www.garfield.library.upenn.edu/essays/v5p511y1981-82.pdf

[13] Stock, W.G. "Eugene Garfield und die Folgen: Der Weg der Fussnote bis in die Wissenschaftspolitik," *Password*, pages 14-19 (March 2002).

[14] Garfield, E. "From computational linguistics to algorithmic historiography" Lazerow Lecture held in conjunction with panel on "Knowledge and Language: Building large-scale knowledge bases for intelligent applications" presented at the University of Pittsburgh on September 19, 2001.
http://garfield.library.upenn.edu/papers/pittsburgh92001.pdf

[15] Garfield, E. "From bibliographic coupling to co-citation analysis via algorithmic historio-bibliography" A citationist's tribute to Belver C. Griffith, Lazerow Lecture presented at Drexel University, November 27, 2001. http://garfield.library.upenn.edu/papers/drexelbelvergriffith92001.pdf

[16] Garfield, E. "The Impact Factor and using it correctly", *Der Unfallchirurg*, 101(6):413 (June 1998).
http://www.garfield.library.upenn.edu/papers/derunfallchirurg_v101(6)p413y1998.pdf

PASSWORD 06/2002, p. 22-25.