

# A Tribute To Calvin N. Mooers, A Pioneer Of Information Retrieval

Author: Eugene Garfield

*The Scientist*, Vol:11, #6, p.9, March 17, 1997

Last October at its annual meeting in Baltimore, the *American Society for Information Science* presented a session entitled "History of Information Science: Reminiscences and Assessments." Part of the session memorialized Calvin N. Mooers, a pioneer of information science who passed away in December 1994.

Mooers was responsible for many innovations in computer and information science. He is perhaps best known for coining the term "information retrieval" while writing his master's thesis at the Massachusetts Institute of Technology. The *Oxford English Dictionary*, Second Edition, Online cites the original source for this term as the *Zator Technical Bulletin* No. 48 (1950), a publication of the Cambridge, Mass.-based Zator Co.-which Mooers founded in 1947-with the following definition: "The requirements of information retrieval, of finding information whose location or very existence is a priori unknown. . . ."

I first met Calvin in 1951 when he visited the Welch Medical Library Indexing Project at Johns Hopkins University. The goal of the project, of which I was a member until its termination in 1953, was to develop computer-derived indexes to the scientific and medical literature. This mechanization of bibliographic information involved the use of IBM tabulating equipment designed for statistical analysis. The Welch project used standard punched-card machines to

prepare subject-heading lists for the Armed Forces Medical Library, the precursor to the National Library of Medicine (E. Garfield, "The preparation of subject-heading lists by automatic punched-card techniques," *Journal of Documentation*, 10:1-10, 1954).

Mooers had developed a clever method to store a large number of document descriptors on a single specially notched card, which he called Zato coding. He was able to do this by superimposing random, eight-digit descriptor codes. The result was a small but tolerable number of "false drops" in a bibliographic search-that is, retrieved documents that were not relevant to the search parameters. Calvin founded the Zator Co. to market his system.

Computer and information scientists may recognize Zato coding as a variation on what was later called hash coding. Many years later, hash coding was used in the design of the Institute for Scientific Information's (ISI) SciMate software for information retrieval, which is still used by thousands of fans. Its search capabilities have been significantly enhanced by the enormously increased speed of personal computer chips. A sequential search of thousands of records can be performed in a matter of seconds.

When he visited us at the Welch project to demonstrate his Zato coding system, Calvin struck me as quite single-minded,

even opinionated. I remember resenting the fact that he was "selling" us on a commercial, for-profit product, which I inherently mistrusted. I hadn't yet overcome the idealistic notion that all good things were nonprofit, which was probably a reflection of youthful naïveté. However, my later experiences with large government agencies and nonprofit institutions changed that view. When I had to survive as a private consultant and for-profit entrepreneur, as Calvin did, I appreciated how difficult and frustrating it was to compete with the arrogance and market advantages of these nonprofit establishments.

I took inspiration from Calvin's superimposed coding system when I designed a retrieval system based on Hollerith (IBM) punched cards and the IBM 101 Statistical Machine (E. Garfield, "Preliminary report on the mechanical analysis of information by use of the 101 statistical punched card machine," *American Documentation*, 5:7-12, 1954). Medical subject headings and subheadings were represented by seven-digit numbers. Instead of being superimposed, the numbers were spread out over 10 separate fields on these 80-column, 12-row cards. What was superimposed was the special wiring for programming on the control panel of the 101 machine.

After spending much of his effort on improving mechanized information-retrieval systems, Mooers recognized that "an information retrieval system will tend not to be used whenever it is more painful and troublesome for a customer to have information than for him not to have it." In a recent telephone conversation, Joshua Lederberg at

Rockefeller University succinctly rephrased what has come to be known as Mooers' Law as follows: "People will resist information unless the price of not knowing it greatly exceeds the price of learning it." I heard Calvin present his law in 1959 at the annual meeting of the American Documentation Institute at Lehigh University. A reprint of his remarks, which provide insights on the behavior of those who rely on information-retrieval systems, is presented on the following page. Implicit in Mooers' Law is the recognition that the quantity of information retrieved itself is not necessarily valued or even desired. The mechanical trawling of massive databases to catch thousands of matches to simple subject-heading search terms is primitive. Rather, quality of information is desired—that is, the more relevant the retrieved information is, the more valuable it will be to the user. Thus, a corollary to Mooers' Law would propose that the more relevant information a retrieval system provides, the more it will be used.

I realized early on in my career that to develop a truly qualitative information-retrieval system, one must break the simple subject-index barrier. Over the years, novel indexing and retrieval systems have been created that overcome the limitations of traditional subject headings. These systems usually succeed in providing users with more directly relevant "hits." An example is the cited reference searching of the *Science Citation Index* (SCI). Natural-language searching has also been adopted in SCI, BIOSIS, and other systems. While no system can guarantee 100 percent efficiency in providing only the most relevant hits to a search query,

the technological and conceptual advances made since Mooers' Law was first proposed have come a long way in improving the precision of information retrieval systems.

Calvin Mooers' papers are stored at the Charles Babbage Institute, University of Minnesota. Scholars may access them at <http://www.cbi.umn.edu/inv/mooers.htm>

---