Essays of an Information Scientist, Vol:4, p.341-347, 1979-80 Current Contents, #51, p.5-11, December 17, 1979

Current Comments

Some Reflections on *Index Medicus*

Number 51 December 17, 1979

Recently, I attended the 23rd annual meeting of the Council of Biology Editors in Charlottesville, Virginia. I was pleasantly surprised to meet Cliff Bachrach, an old friend and colleague. Cliff and I met in 1951 just after I joined the Welch Medical Indexing Project at Johns Hopkins University in Baltimore. At that time he was the hospital biostatistician. In that capacity he was responsible for the punched card equipment used for medical record keeping and other purposes. Since I was investigating ways to use the IBM 101 statistical sorter for information retrieval, he and others in the statistics department had a keen interest in what I was doing. I'm sure that at that time Cliff never dreamt that he might someday also migrate to the information retrieval field.

But in 1966 he joined the staff of the National Library of Medicine (NLM) as chief of the Bibliographic Services Division. He has now assumed the role of editor of *Index Medicus*.[1] I congratulate him on this and wish him well. *Index Medicus* is a monthly index to articles published in about 2,500 biomedical journals from around the world. This year *Index Medicus* celebrated its 100th anniversary.

It's not customary to say much about one's competition. Many of my professional colleagues manage to write articles on one or more aspects of medical or scientific documentation without once mentioning *Current Contents®* *(CC®)* or *Science Citation Index®* *(SCI®)*. ISI®'s very existence disturbs their equilibrium! But I never hesitate to discuss our competition. In the case of *Index Medicus*, it would be intellectually dishonest. Besides, I have a sentimental attachment to *Index Medicus*. I am also well aware of its advantages and disadvantages when compared to other indexing systems because I played a role in its evolution.

Recently, the *Journal of the American Medical Association (JAMA)* singled out *Index Medicus* for special acclaim on its centenary celebration.[2] The AMA wasn't always so sanguine about the status of *Index Medicus*. In fact, the AMA published *Quarterly Cumulative Index* from 1916 to 1927 while the original *Index Medicus* was published by the Army Medical Library.[2] I won't detail the entire history of *Index Medicus* here. Jeffrey Kunz has done a good job in *JAMA* of tracing its history. But I would like to fill in some details that I find especially relevant. In particular I want to mention some of the events of the last 30 years in *Index Medicus'* history that made it one of the world's most recognized information retrieval systems.

In 1948, the Army Medical Library established the Welch Library Indexing Project under contract with Johns Hopkins University.[3] The project examined basic and applied problems of medical information retrieval, especially those involved in publishing indexes of current biomedical literature. *Index*

Medicus as it exists today is directly related to the Welch project because we resolved many of these problems.

When the project was established, three indexes of biomedical literature were available. The *Index-Catalogue of the Library of the Surgeon General's Office* was published by the Army Medical Library. The *Index-Catalogue* indexed the library's monograph, pamphlet, and thesis holdings. The library's periodical holdings were only indexed selectively. The *Current List of Medical Literature*, which led to the present *Index Medicus*, was also published by the Army Medical Library. It had been started in wartime and listed the library's periodical holdings that were available on microfilm. Later, in 1950, *Current List of Medical Literature* was expanded but the subject indexing was inconsistent and incomplete. Its new organization by author and subject headings was inadequate for reference purposes. At that time the *Quarterly Cumulative Index Medicus (QCIM)* was published by the AMA under the direction of Magdalene Freyder. *QCIM* was the standard medical index, offering the most complete listing of current biomedical literature. But World War II had made it difficult for the AMA to maintain its coverage of foreign journals, and *QCIM* was more than two years behind schedule.[4]

Both the AMA and the Army Medical Library wanted to unify these overlapping bibliographic services and to improve currency and coverage of biomedical journal articles. Under the directorship of Sanford V. Larkey, the Welch project specifically examined "the size of universe of biomedical literature, and the extent of its coverage by the existing services, the common and disparate features of subject-headings among the services, and the application of machine methods to medical indexing."[3] It was my good fortune to become a member of the project in March 1951, shortly after I was introduced to Larkey by James W. Perry, who wrote the first self-teaching book on scientific Russian.[5]

At that time, IBM machines were commonly used to process personnel records, accounting invoices, actuarial statistics, and other large volumes of information recorded on "Hollerith" punched cards. The machines performed a variety of simple functions too tedious and time-consuming for an office staff—sorting, counting, matching, alphabetizing, and printing. Since creating medical indexes involved these tasks, it was hoped that existing IBM machines could be used to perform these repetitive functions.[6] One also hoped that by slight modifications we could use them for the "searching" function.

Punched card machines were already being used on a limited basis. The subject heading authority lists of the Library of Congress' Technical Information Division and *Current List of Medical Literature* were maintained on punched cards.[7] However, these machines were severely limited in the range of information problems they could handle. Standard sorting machines operated on only one column at a time. Since the standard punched card had 80 columns and 12 rows, this was equivalent to reading one letter on a printed page![6] If you wanted to search for more than one out of the 960 possible punch positions on an IBM card, you had to pass large volumes of cards through the sorter several times. Or if you used a decimal coding system with fixed fields you had an equally difficult job.

Thus, if you wanted information on the effect of caffeine on heart rate, the entire file of punched cards would be searched for the caffeine code and then once again for the heart rate code. Only one question could be asked at a time.

Obviously, this wasted a lot of time and frustrated attempts to use punched card systems.

The IBM 101 could search all 80 columns simultaneously. This was an improvement but hopelessly inadequate for our purpose. I worked out a "superimposed" wiring system for the 101 that enabled us to search for as many as 48 codes in various combinations at one time. A number of separate complex questions could be asked simultaneously, and the effective speed of the machine was increased as much as tenfold.[6] But the IBM 101 was never used operationally by the Army Medical Library or its successor, the National Library of Medicine. Claire Schultz of Merck, Sharp & Dohme designed a clever application of the IBM 101 in combination with Calvin Mooer's system of Zator coding.[8]

One of the main research tasks of the Welch project was the revision of the subject heading authority list of *Current List*. For this we used standard tabulating equipment.[7] The list we eventually produced became the first *Subject Heading Authority List (SHAL). SHAL* was the prototype of *Medical Subject Headings (MESH),* the authoritative list of indexing terms presently used by the *Index Medicus* staff.[3]

Before the use of punched card equipment, subject headings were usually available on printed alphabetical lists or in 3 x 5 card files. It was difficult to analyze such lists and work out relationships between the 10,000 or more terms involved. Sanford Larkey, Helen Field, and Williamina Himwich had already planned to transform this data to punched cards when I joined the project. "We started...to put on punched cards all the headings used by *Current List* in the first five months of 1951. The actual heading was put down in full by alphabetical punching, and code numbers [were numerically] punched for category sorting.... By category sorting and then printing lists from the punched cards, we were able to study the terms used in restricted fields and to compare them against other authorities."[9] After the Welch project ended in 1953, many librarians continued to use these punched card files in their own studies of subject heading lists.

In the spring of 1951, during a strategy session on the philosophy of categorizing subject headings held at Johns Hopkins, I met Robert Hayne,[10] the assistant editor of *Current List.* Bob was temperamentally well-suited to his job. He previously was a classics scholar. And he could type and index faster than any individual I ever met—with equal facility in a half-dozen or more languages! After I met Bob I went to Washington to meet Seymour Taine, then the editor of *Current List* and now chief librarian of the World Health Organization. I also met Frank "Brad" Rogers, an Army physician fresh out of Columbia Library School. He had been selected as the new director of the Army Medical Library. I also met Estelle Brodman, their head of reference, and Samuel Lazerow, head of acquisitions. I can't imagine the Army Medical Library, or *any* government agency, having a more dedicated group of managers. If I am critical and skeptical about government operations, it is not because I have not met well-motivated people in government. But I know all too well that the government is inefficient despite the efforts of dedicated persons. It is very difficult for such managers to overcome the built-in inertia of an elaborate bureaucracy.

Some might say that *Current List of Medical Literature* was the predecessor of *Current Contents.* This is not without foundation because the original *Current List* produced by Atherton Seidell was a pocket-sized publication similar to *CC.* It contained tables of contents for bio-

medical journals in fairly small type. Later on, *Current List* issued separate author and subject indexes. Its major transformation took place in 1950 under the direction of Seymour Taine and Brad Rogers.[4] Journals were listed alphabetically in the article registry, and each was assigned a six-digit serial code to connect with author and subject indexes. Ten regular issues of *Current List* appeared each year, and semiannual subject and author indexes appeared as the June and December issues.[11]

However, *Current List* was never current enough to satisfy the needs of researchers in academia or the drug industry. A 1954 report indicated that about half the material it listed was more than a half-year old. Moreover, the late appearance of the semiannual subject and author indexes had an adverse effect on overall currency.[11] You can be sure I remembered this when I implemented bimonthly issues of the *SCI*.

Current List also ran into problems with the number of journals it covered. By the late 1950s, *Current List* indexed about 110,000 articles from 1,600 journals. But, according to a 1957 survey conducted by Brodman and Taine for the International Conference on Scientific Information, the current world production of biomedical papers was estimated at 220,000.[3] The NLM, successor to the Armed Forces Medical Library, felt compelled to increase coverage by 70,000 additional articles.[11] This increase in coverage made it even more imperative to implement a new system of production.

A development project was established in 1958 to overcome *Current List's* coverage and currency weaknesses by applying new information processing technology to the production of *Current List*.[3] The objectives of the new Index Mechanization Project were those of "developing and demonstrating in the field of medicine improved methods for the rapid and efficient publication of comprehensive indexes to the literature of broad scientific fields, with simultaneous provisions for meeting the requirements of specialties within these fields, making use of hitherto unutilized mechanical applications."[3] Although I was not a member of the project's advisory committee (despite Scott Adams' kind assertion that I was[3]), I worked closely with the committee to recommend possible machine methods for the production of *Current List*.

The project's advisory committee decided to change the way articles were entered in *Current List*. Previously, bibliographic entries were split between the article registry and the author and subject indexes. If you wanted complete bibliographic information, you had to search each section for the six-digit numerical "addresses" connecting a specific article title with its author and subject headings. This constant referral and re-referral became more awkward and tedious as the volume of biomedical literature grew. Also, the demand for limited indexes for restricted specialties increased with the growing volume of biomedical literature. Entries had to be completely recast and recomposed before specialty indexes could be prepared.

The committee recommended the adoption of a "unit" entry. The unit citation gives complete bibliographic information in a single entry listed under subject. Once you located an article you were interested in, you needn't refer to any other source for more information. More importantly, these entries could easily be scanned, sorted, and duplicated to produce limited indexes for restricted specialties from *Current List*.[11] While the adoption of this form of entry made it easier to use *Current List* and permitted multiple use of its data file, the numerical article registry listing journal contents pages was dropped completely. Some people regretted the loss of this useful feature. However, by

that time *Current Contents* was widely used in the pharmaceutical industry and was gaining acceptance in academia as well.

Entering bibliographic information in the new form also made it easier to mechanize the production of *Current List*. I designed the punched card system that eventually was used in a slightly modified form. Taine summarized the new production system for *Current List*.[11] "Using the Justowriter, the typist would compose the copy and, at the same time, the machine would produce a coded paper tape. This paper tape would, in effect, serve as a recording of the unit citation entry. Individual IBM cards for each author and subject entry would be prepared on keypunch machines. The citation tape, previously prepared, would then be used to type out automatically the visible entry at the top of the appropriate punched cards. The next link in the chain would be the mechanical interfiling of the entries with selected heading and cross reference cards using IBM sorters and collators."[11]

Mechanical typing, interfiling, sorting, and collating greatly reduced the time it took to prepare *Current List* for printing. But the greatest timesaving machine was the Listomatic camera. "The Listomatic camera is a high speed, precision machine capable of photographing text imprinted or typed at the top of an IBM punched card. One, two, or three lines of copy, precisely positioned on the card, may be fed into the Listomatic camera and photographed at the rate of 230 cards (up to a maximum of 690 lines) per minute.... A special punch in the card 'informs' the camera of the number of lines in the entry in order that an automatic adjustment in aperture size may be made.... The product is a negative film...which, after processing, is ready for photo-offset plate making."[11] Using the Listomatic camera, an entire monthly issue of *Current List* was photographed in a single day, and the annual cumulation required about two weeks.[11]

The Index Mechanization Project succeeded in meeting its goals. The unit citation form of entry made it possible to reuse previously prepared material to produce specialized indexes without costly recasting and recomposition of entries. The new mechanized system using punched cards and the Listomatic camera further improved the preparation of specialized and standard indexes.

The eventual recommendations of the Index Mechanization Project changed the format, production, and print quality of *Current List* to such an extent that the new index bore little resemblance to the original. In 1960, *Current List* took over the name *Index Medicus*. For several years the NLM published the monthly issues while the AMA published the annual cumulations at its own expense, having dissolved the *QCIM*. This had to be one of the most bizarre cooperative relationships in history. Although the AMA was an outstanding critic of any government involvement in medicine, it gladly entered into partnership with the government to publish a key tool for the medical profession.

The Index Mechanization Project also demonstrated that using millions of punched cards was too unwieldy. It was almost impossible to process literature searches requested by individual scientists! The project concluded that "a retrieval system could not be successfully grafted onto a publication system, which deals in large measure...with the problems of composition.... To achieve both ends, in the contexts of a very large system, it would be more suitable to start with the design of a retrieval system, and then to proceed with the publication system...."[12]

In 1960, just a year after the Index Mechanization Project ended, NLM established the Medical Literature Analysis and Retrieval System (MEDLARS) project. The MEDLARS

project intended to exploit digital computers to achieve two goals: first, to continue to publish *Index Medicus* and related specialty indexes with the same high quality print and currency that the biomedical community had come to expect, and second, to process literature search requests with a thoroughness and speed currently unavailable.[13]

With these goals in mind, the MEDLARS system was designed around three subsystems: "an input subsystem in which the skills of professional indexers are used in connection with the capabilities of a large-scale digital computer; a retrieval subsystem in which the capabilities of professional literature searchers are used in connection with computer manipulations; and a publication subsystem that converts retrieved citations...into photopositive film."[14]

In the publication subsystem, NLM introduced a brand new computer-driven phototypesetter called GRACE—*GR*aphic *A*rts *C*omposing *E*quipment. Developed by the Photon 900 Company of Wakefield, Massachusetts, GRACE was a significant improvement over the standard "hot-type" printing systems used in those days. "Operating at a speed of about 300 characters per second, GRACE can use 226 different characters in preparing 23-centimeter wide photographic film or paper. Character sets include a 6-point font of regular and bold-faced upper- and lower-case characters, a 10-point font of upper-case characters only, and a 14-point font of upper-case characters only. The exposed film is developed by an automatic film processor, inspected, and cut into page-sized sheets."[14]

The August 1964 *Index Medicus* was the first book composed by GRACE. It is considered as much a landmark in the history of phototypesetting as Gutenberg's Bible is in the history of printing![15] Indeed, one might say that much of the computerized photocomposition that takes place today owes its origin to the development of GRACE by NLM's MEDLARS project. I doubt that most private companies would have spent the amount of money that was poured into developing this system. I suppose it's to the government's credit that it hadn't wasted the money on some less useful project. Regardless of its cost, GRACE was a significant innovation in publishing.

Using a digital computer also made it practical to increase the number of terms used to index material stored in the MEDLARS data bank. Old issues of *Index Medicus* used an average of two headings per article, inadequate for most search and retrieval systems. The decision was made to begin indexing articles with as many as eight or ten terms per article. If all these terms were used in the printed *Index Medicus*, the subject section would be three times as large.[13]

Consequently, the average number of indexing terms was increased but many of the terms are used only in the computer system for search purposes. In 1964 Martin M. Cummings became director of NLM. He has enjoyed great support from Congress in expanding NLM's role, especially in the development of MEDLINE. Introduced in 1971, MEDLINE gave librarians on-line access to bibliographic information on MEDLARS' magnetic tapes. NLM has been the largest supporter of on-line access to bibliographic files. It has subsidized the growth of on-line searching via MEDLINE to the tune of millions of dollars. Whether such unqualified subsidy is warranted can be debated. But the impact of MEDLINE has been widespread. Guided by the philosophy that the lowest possible charge to the searcher fosters greater use, MEDLINE continues to be one of the cheapest bibliographic services available. However, hourly rates can be deceptive. One system may cost $15 per hour but the average search may take 20 minutes at a cost of $5. Another system may cost $75 per hour but the average search requires only 4 minutes at a cost of $5. One of the main criticisms of MEDLINE by sci-

entists is that only specialists trained by NLM are able to use the system effectively. However, this is true of many other on-line systems.

The $150 which libraries pay for a subscription to the monthly *Index Medicus* covers only the cost of printing and postage. NLM's annual $40 million budget has to cover the cost of acquiring journals, indexing, and coding in addition to funding grants and cooperative efforts with other institutions. A significant part of the indexing work is now done on contract by Herner & Co., the Franklin Institute, Raytheon Service Co., and Kessler Communications Inc.

NLM has found some interesting ways to stretch its budget. Foreign governments and institutes cover the literature of their respective countries in exchange for access to the MEDLARS data base. Thus the Japanese index the Japanese journals, the Germans do the German, and so on. One day the Chinese will get into the act unless the Russians get there first.

Index Medicus certainly has come a long way since it first appeared in print in 1879. Its principal founder, John Shaw Billings, only intended "to prepare a comprehensive catalog and index which should spare medical teachers and writers the drudgery of consulting ten thousand or more different indexes."[2] In view of the present scope of *Index Medicus,* including MEDLARS and MEDLINE, Billings' intention seems almost naive. Medical educators, students, and writers are indeed spared some of the drudgery in literature searching. But no one has reduced the job of reading all this literature. However, physicians and clinicians now have almost immediate access to the literature.

Undoubtedly, the quality of medical care and education has been positively affected over the years by *Index Medicus* and its predecessors. As it turns out, in most medical libraries today *Science Citation Index, Current Contents,* and other ISI services complement *Index Medicus* and other indexes. Obviously when budgets are tight one of several competitors may lose out. But experienced librarians and users learn to use the advantages of all systems in a complementary way. We wish *Index Medicus* well on its 100th anniversary.

REFERENCES

1. Editor of *Index Medicus* appointed. *Nat. Lib. Med. News* 34(6):2, June 1979.
2. **Kunz J.** *Index Medicus.* A century of medical citation. *J. Amer. Med. Ass.* 241(4):387-90, 26 January 1979.
3. **Adams S.** The way of the innovator: notes toward a prehistory of MEDLARS. *Bull. Med. Libr. Ass.* 60:523-33, 1972.
4. **Brodman E.** Contribution of the Army Medical Library to the bibliographic control of medical literature. *Spec. Lib.* 43:48-54; 73-6, 1952.
5. **Garfield E.** Reflections on learning foreign languages. Part 2: Electronic translators. *Current Contents* (38):5-9, 17 September 1979.
6. ----------. Preliminary report on the mechanical analysis of information by use of the 101 statistical punched card machine. *Amer. Doc.* 5:7-12, 1954.
7. ----------. The preparation of subject-heading lists by automatic punched-card techniques. *J. Doc.* 10:1-10, 1954.
8. **Schultz C K.** An application of random codes for literature searching. (Casey R S, Perry J W, Berry M M & Kent R A, eds.) *Punched cards: their application to science and industry.* New York: Reinhold, 1958. p. 232-47.
9. **Larkey S V.** The Welch Memorial Library Indexing Project. *Bull. Med. Libr. Ass.* 41:32-40, 1953.
10. **Garfield E.** To remember my brother Robert L. Hayne. *Current Contents* (34):5-6, 22 August 1977.
11. **Taine S I.** New program for indexing at the National Library of Medicine. *Bull. Med. Libr. Ass.* 47:117-23, 1959.
12. The National Library of Medicine Index Mechanization Project. *Bull. Med. Libr. Ass.* 49:1-96, 1961.
13. **Taine S I.** The Medical Literature Analysis and Retrieval System. *Bull. Med. Libr. Ass.* 51:157-67, 1963.
14. **Karel L, Austin C J & Cummings M M.** Computerized bibliographic services for biomedicine. *Science* 148:766-72, 1965.
15. **Lechéne R.** Printing. *Encyclopaedia Britannica.* Chicago: H.H. Benton, 1974. Vol. 14, p. 1058.