

**A brief introduction by Eugene Garfield to M.R. Hyslop's paper,  
"Documentalists consider machine techniques"  
*Special Libraries* 44(5):196-198, May-June 1953**

---

A significant event in the history of information science was the Symposium on "Machine Methods in Scientific Documentation" organized and conducted by Eugene Garfield at the Welch Medical Library in Baltimore on March 3, 1953. Marjorie Hyslop who died in 2001 at the age of 93, wrote this report which is the only published record of the meeting. In addition to Garfield's discussion of punched cards methods, use of IBM 101 statistical machine for searching, and use of categorization to study MESH terms, the symposium included presentations by Saul Herner on the "Information gathering habits of scientists". Other participants included H. Peter Luhn(IBM), John Mauchly(UNIVAC), Ralph Shaw (US Dept. Agriculture), Verner Clapp (Library of Congress) and Mortimer Taube (Documentation, Inc.). At the time, this symposium received wide publicity in the Press when the Vice President of Johns Hopkins University, Lowell J. Reed, opening the session stated that mankind is drowning in a flood of information. The press reports stimulated W.C. Adair of Shepard's Citations to write Eugene Garfield which eventually lead to their meeting.

# Documentalists Consider Machine Techniques

Reported by MARJORIE R. HYSLOP

American Society for Metals

Chairman, Committee on Special Classification, Special Libraries Association

WHEN THE SYMPOSIUM on "Machine Techniques in Scientific Documentation" held at the Welch Medical Library of Johns Hopkins University on March 3 was originally conceived, it was anticipated that some fifty documentalists might assemble in a sort of informal round table affair for a day of prepared papers, discussions, and demonstrations of equipment. Instead, the surprised hosts found themselves with more than 250 enthusiastic guests and participants on their hands!

This, naturally, necessitated some last-minute readjustments to provide larger meeting rooms and changes in demonstration schedules; the negligible amount of confusion and discomfort was a tribute to the careful planning and attention to detail that went into preparation of the program. Another complication was the illness of Sanford V. Larkey, director of the MEDICAL INDEXING PROJECT and librarian at Welch Medical Library; however, his capable assistant, Eugene Garfield, rose gallantly to the occasion and presented papers, conducted demonstrations, answered questions, ushered guests, and indefatigably played the cordial host from the time the doors opened at 8:30 in the morning until the chairman of the evening panel discussion reluctantly called a halt at 10 P. M. Verner Clapp, Chief Assistant Librarian, Library of Congress, presided very effectively at both morning and afternoon sessions, and Eugene Miller, of the Armed Services Technical Information Agency and president of American Documentation Institute, conducted the evening discussion.

While the machine methods described and illustrated in the prepared program referred specifically to the MEDICAL INDEXING PROJECT being conducted at Johns Hopkins University under contract with the Armed Forces Medical Library, the techniques could be applied to any field of information, scientific or otherwise, and the individual listener had no difficulty in making mental translations of the medical terms used as illustrations into the language of his own specialty.

The PROJECT at the Welch Medical Library has two phases. First is the preparation of a subject heading authority list for the *Current List of Medical Literature* published by the Armed Forces Medical Library. The second phase is the preparation of subject indexes by punched-card techniques. While the PROJECT so far has only an experimental file of coded literature references, the difficult part of the job is the selection of appropriate subject

headings; furthermore, the principles being developed can apply to information analyses of all kinds.

## Subject Heading List

First step in the preparation of the subject heading authority list was the compilation of a list of 7,000 medical periodicals, which were analyzed according to language, country of origin, abbreviation of title, frequency, type of journal, primary and secondary subject field, coverage by some thirty-four abstracting services, and incidentals of library control such as distribution and acquisition.

All of this information was coded and recorded on IBM cards and sorted and analyzed by the so-called 101 *Electronic Statistical Machine*. This analysis, together with a compilation of the subject heading list, is said to be the only project of such size, scope, and staff utilizing machine methods for literature handling now in existence.

The feature of this periodical analysis that received most comment in discussion had to do with the coverage of medical periodicals by abstracting journals. Titles of 4,674 periodicals were checked as to their coverage by the thirty-four abstracting services listed, and while 546 were found to be covered by as many as four different services, and 833 by the two most comprehensive services, 2,031 titles or forty-three per cent were neither indexed nor abstracted by any of the services listed. Granted that the nature of these 2,031 titles was not indicated and that they might be covered by other abstracting services than the thirty-four studied, the consensus of the meeting was that something needed to be done for the benefit of the scientists using the medical literature. Extension of such studies as this to other fields of knowledge might reveal equally provocative statistics.

Returning to the compilation of subject heading lists, Mr. Garfield explained how the principle of categorization was applied to bring together related terms, eliminate redundancies and relegate synonyms to cross references. Each specific subject heading term was coded to indicate its place in one of several broad categories. (Typical examples of categories in the medical field are chemistry, pathology, organisms, psychology.) When cards are sorted according to category, a study of the resulting list of terms will lead to selection of the most appropriate term from a list of synonyms and also indicate appropriate subheadings. In this connection, Mr. Garfield

took issue with the commonly heard statement that no matter what type of indexing is used, you can get no more out of the system than you put into it. The very fact that machine methods can bring out combinations of sub-headings not envisaged in the original index, and hidden correlations not revealed by printed indexes, can be interpreted as new information not intentionally put into the file.

Discussion of this paper brought out the admonition that the preparation of punched cards should be regarded as a means to an end and not an end in itself. Efficiency of the system must always be examined in the light of the final product, i.e., printed information in a form readily usable by the scientist. Preparation of elaborate punched-card files is not warranted if the information is to be sorted, correlated and printed once only; to earn its salt the file must be useful for repetitive operations, the more repetition the cheaper the original investment. Many simple searching problems do not warrant machine techniques. Searching for authors by machine, for example, was compared to "using a Cadillac to cross the street". Selecting an author card from an alphabetically typed list can be done in a tiny fraction of the time it would take to run a list of cards through a machine to select all of the Joneses.

### Equipment

When it came to describing the operation of the equipment itself, Mr. Garfield and the library staff, by means of exposition, black-board, demonstration, samples for the audience, and handsome wall charts prepared by the Medical Art Department of the University, did their utmost to simplify and clarify the translation of codes into electrical impulses and vice versa—with, this reporter would judge, varying degrees of success. Without going into the involved operations of the machine's anatomy, some generalizations might be stated here. Mr. Garfield is a staunch defender and enthusiastic proponent of the capacities and flexibility for literature searching offered by the IBM statistical machine. To quote:

"The 101 punched-card machine is a most versatile device for information analysis. It can handle statistical problems, as well as cumulative counting problems such as those encountered in accounting problems, but at the same time it is capable of a wide variety of applications in punched-card systems for searching literature. These include fixed field systems, superimposed coding systems, binary coding systems, as well as the type described today, namely a fluid coding system where there are no fixed requirements for particular categories of information . . . it is capable of searching for any particular punched hole in any location of the card, or any combination of

holes . . . Further research is of course required—but there can be little doubt that this machine already incorporates many features of the ideal machine for literature searching."

The fluid coding system mentioned is apparently a major contribution to punched-card techniques contributed by the MEDICAL INDEXING PROJECT, and expands immeasurably the searching possibilities of the equipment. The method does result in a certain number of false sorts, but these are readily eliminated by the machine on a second very quick sort of the selected cards. The necessity for this second quick run is more than compensated for by the system's ability to search simultaneously for a large number and variety of items of information (specifically, a theoretical capacity for forty-eight combinations of five-digit codes).

While the IBM 101 machine naturally played the stellar role in the performance, minor but popular attractions were also provided by operating installations of various types of equipment for reproduction and processing of data, and a "standing room only" movie of the justly famous Univac. Present in person to answer questions and provide additional explanatory material were such luminaries as John W. Mauchly, Remington Rand's Univac pioneer, H. P. Luhn and J. Nolan of IBM, and other representatives of the equipment manufacturers.

### Sources of Information

The title of one paper on the printed program, *The Information Gathering Habits of Johns Hopkins Scientists*, might appear at first glance to be somewhat off the general subject, but its inclusion was only another indication of the perspicacity of those who arranged the symposium. As Ralph Shaw put it, in introducing the speaker, "We've got the cart; now let's look at the horse." And a lively horse it was, with plenty of meat on its bones, since Mr. Hermer's paper presented the first-stage results of a large-scale statistical study being conducted at Johns Hopkins. Mr. Hermer's statistics were gathered in personal interviews with three hundred scientists individually characterized as to age, academic degree, primary interest in physics, chemistry, mathematics or engineering and emphasis on pure or applied research. The purpose of the interviews was to determine the types of information sources most widely used by the scientists in these various categories. Of the fifteen primary sources of information considered, the following were found to be most popular (ranged approximately in order of importance, although there were variations depending on the groupings of the scientists): Advanced publications, research periodicals, reports and handbooks. Of the six secondary or indirect sources, conversations, regular per-

usal of periodicals, and references cited in books and papers ran a close race in that order. Abstracts and indexes constituted a popular fourth, and the library card catalog was found to be used primarily as a guide to physical location of material. The prominence given to lists of references provided in written papers led Mr. Herner to the conclusion that critical and selective reviews constitute an area that demands more attention from the publisher\*.

The animated and informal panel discussion in the evening, drew together some loose ends and also brought forth some new ideas not developed during the daytime sessions. Among the latter was a debate centering on the comparative merits of attacking the literature problem from a standpoint of order and logic (citing Mortimer Taube's system of *Coordinate indexing* as an example) and the use of random numbers in coding for machine retrieval of information.

During this session, Mr. Clapp improvised a good summation of the problems facing the literature specialist by listing the following regions requiring further exploration:

1. More needs to be known about the information-gathering habits of the investigators.
2. The various literature-handling projects in limited fields of interest need to be coordinated into a whole; methods must be ap-

*\* This statistic was particularly surprising to this reporter, since the editor is inclined to eye long lists of references somewhat askance. Their prevailing inaccuracy and incompleteness frequently lead him to regard them as a bit of fancy dress added to the paper merely to testify to the author's erudition without any proof that he has ever read them.*

plicable to large as well as to small fields of knowledge so that a universal classification scheme based on these new concepts may possibly be developed.

3. The limitations of subject heading lists should be borne in mind. The finer the subdivision of knowledge being considered, the more difficult it becomes to define a concept in terms of a single word or phrase.

4. Handling and indexing of periodicals is one of the major problems facing the information specialist. The great bulk of the periodical literature requires an indexing device which must not only select the desired information, but must also bypass the material that is not wanted.

5. Cooperative indexing might offer one solution.

6. A copying device that would permit the scholar to go into the library, scan the material line by line, and leave with a convenient roll of tape containing an exact transcription of the desired information, would be a tremendous labor-saver. Such a device, "electronic pencil", has been developed, but is not yet commercially available.

7. Complete coverage of periodical as well as book literature is essential to the user of indexing and abstracting services.

Unfortunately, it is impossible to enumerate all of the contributors to discussions which invariably had to be terminated long before the questions and comments were exhausted. The device of appointing qualified individuals to lead the discussion after each paper paid off handsomely, and the sponsors of the project could chalk up an unqualified success in their avowed purpose of providing "a symposium with active participation of visitors, rather than the passive attendance common to scientific gatherings today".

## SLA Exhibit Guide

A practical guide to aid SLA members in planning exhibits has been prepared by the Public Relations Committee and is now available on request, Genevieve Ford, chairman, announces.

The "SLA Exhibit Guide" describes in detail how to plan and prepare a professionally-executed display, the personnel required, exhibit materials available from the Association, procedures, suitable presentation, financial arrangements, servicing and related information.

Comprehensive coverage provides an indispensable source of authoritative information designed to simplify the problems and to achieve the purpose of the successful exhibit.

Members of the Sub-Committee on Exhibits primarily responsible for the "Exhibit Guide" include Mary Agee, Mrs. Marie S. Goff, Mrs. Marjorie R. Hyslop, and Meredith Wright, chairman.

Further acknowledgment is made to the following librarians and their organizations who have made it possible to extend available exhibit materials: Dorothy Wescott of Rohm & Haas; Alberta Brown of the Upjohn Company; and Caroline Lutz of General Motors.