

Control of Subject Information: Can It Be Mechanized?

Mr. Bristol is head of the Catalog Department, Peabody Institute Library, Baltimore.

TO librarians the Industrial Revolution bears the aspect of a communicational revolution, which in the past century has produced an ever-swelling flood of informational materials, many in new media. Within the limits of printed matter the mass of information is so huge that it has been difficult even to inventory it. Nevertheless the inventory approach has resulted in the building up of trade bibliographies, printed library catalogs, and union catalogs, all of which arrange in convenient form and in relatively few alphabets millions of titles, mostly by author, and which among them come close to complete coverage. In this country a complete inventory, if not within reach, is at least within sight along the lines suggested by Ellsworth in 1948.¹

The next logical step is to control the subject information in the inventory. Plainly this is vastly more difficult. For instance, serial publications can be inventoried fairly easily through union lists; but "the bibliography of serial contents is of a magnitude to dwarf the problem of monograph bibliography."² Moreover,

to subject bibliography . . . there is no definable end, no consistent national or annual

¹ Ellsworth, R. E. "Centralized Cataloging in the Library of Congress." *Library of Congress Information Bulletin*, Nov. 16-22, 1948. Appendix.

² Williams, E. E., and Ruth V. Noble. "Subject Bibliography; Preliminary Memorandum." In *Conference on International Cultural, Educational, and Scientific Exchanges*, Chicago, American Library Association, 1947, p. 17.

limitation, no reliable or even desirable uniformity of interpretation, and a great desire to reduce the mass selectively.³

The problem of recording this mass of subject information is an immense one. Yet it is at least equaled in difficulty by the problem of releasing recorded information swiftly and completely enough to satisfy modern research. In the humanities, it is true, scholars have long been accustomed to the leisurely pursuit of learning; nor is the mass of information, though rapidly spreading, bursting all bounds as is the literature of science. Scientific research cannot wait. In the field of chemistry, for instance, Jaffe⁴ points out that developments which hitherto took decades now take only a few years, and that the rate of change is increasing. Scientists are confronted with a mass of observations, facts and ideas which they are not able to put to use.

What would satisfy the scientific researcher? Without regard for the means used to obtain it, he wants a *small* record, one which will not baffle him by its immensity or its complexity; he wants a *complete* record, one which has overlooked no pertinent information; he wants a *convenient* record, one which can be used wherever he chooses; and he wants a *quick* record, one which can be had in a matter of

³ Vanderbilt, Paul. "National Bibliography and Bibliographical Control." *College and Research Libraries*, 9:163, April 1948.

⁴ Jaffe, F. "Quelques reflexions sur la division 66 et plus specialement sur la subdivision 661 de la C. D. U. et sur son emploi pour la documentation de l'industrie chimique." In *Federation internationale de documentation, 17th conference, Berne, 1947. Rapports 1. La Haye, 1947.* p.59.

hours, not weeks.⁵ His desires jibe perfectly with the aim of every reference librarian, which is surely this: the release of the maximum of accurate information with the minimum of time and effort.

Against this background classifications, subject catalogs, and indexing services—tools familiar to every librarian—though often excellent as finding devices, have grave defects both in recording and in releasing subject information. Critics of library classifications have concluded that too little information is released,^{6,7} and that the effort expended in making the schemes work is far too great.⁸ Perhaps catalogers and classifiers, by nature of the puzzle-solving type (else why do they become catalogers and classifiers?), have been so intent on fitting books into classification systems that they have lost sight of the fundamental question of the over-all effectiveness of their work. Certainly any classification scheme is gravely wanting if it is cumbersome to put information into it and to get information out of it, and if there is no assurance that more than a fraction of the information desired is obtainable by using it.

It is safe to say that subject headings, despite cogent criticism⁹ of their unplanned growth and resulting heterogeneous form, do release more information than do classifications. But do they release enough? According to Pettee, the subject catalog is an instrument to serve as a popular guide to general informative reading and not to furnish the exhaustive bibliographies scholars need. . . . The scholar uses the subject catalog only incidentally for material outside his main theme or as a first step in assembling the multitude of

books through which he expects to search patiently.¹⁰

But scientific researchers are increasingly unwilling to undergo "patiently" the drudgery of literature searching. Not only are scientists impatient; they are humanly fallible as well. According to Ball, they simply cannot remember all the terms relating to a given area of scientific effort, or even of any one major science. Moreover, a system of complete cross-referencing among all related entries in a large index creates an unwieldy bulk and a sense of frustration in trying to follow a train of association.¹¹

Researchers are demanding even more subject approaches. Eugene W. Scott, for example, states that some Research and Development Board technical papers may need from 100 to 1000 headings adequately to cover their potentialities. Standard library subject heading work is simply not equal to such a task. Present library practices are so laborious that much material is obsolete before it is usable; subject heading workers cannot know enough fields thoroughly; transactions, pamphlets and serials, which often contain the latest and best information, are usually not analyzed.

Subject work within libraries is even now, in fact, less important than subject work outside them; "the major problem of subject bibliography is the problem of indexing and abstracting serials."¹² Indexing services, under the auspices of private enterprise and of scientific societies, have made extensive inroads on the "bibliography of serial contents," but their coverage is slow, and incomplete in spite of their minuteness; according to Coblans,¹³ only a third of the real

⁵ Taube, Mortimer. "Memorandum for a Conference on Bibliographical Control of Government Scientific and Technical Reports." *Special Libraries*, 39:160, May-June 1948.

⁶ Kelley, Grace O. *The Classification of Books*. New York, H. W. Wilson Co., 1937, p.125.

⁷ Tauber, M. F. "Reclassification and Recataloging in College and University Libraries: Reasons and Evaluation." *Library Quarterly*, 12:835, October 1942.

⁸ Bush, Vannevar. "As We May Think." *Atlantic Monthly*, 176:106, July 1945.

⁹ Prevost, Marie L. "An Approach to Theory and Method in General Subject Headings." *Library Quarterly*, 16:140-151, April 1946.

¹⁰ Pettee, Julia. *Subject Headings; the History and Theory of the Alphabetical Subject Approach to Books*. New York, H. W. Wilson Co., 1947, p.63, 83.

¹¹ Ball, N. T. "Classification," p.6. Preliminary draft of unpublished chapter made available to the author, to be included in Perry and Casey, *Punched Card Systems*.

¹² Williams and Noble, *op. cit.*, p.17.

¹³ Coblans, Herbert. "Organization of the Literature of Science." *South African Libraries*, 14:47, October 1946.

total of worth-while scientific papers gets indexed. Yet the bulk, because of the minuteness of what has been done, has prevented as much cumulation as is needed. Bibliographies, more useful to scholars than subject catalogs because they take the reader's point of view, go beyond the confines of a single library; but they too do not give complete coverage of the field, nor are they rapidly compiled.

Perhaps it would be possible to achieve control of subject information by coordinating all these tools, but successful coordination is certainly far off. Such coordination, even if theoretically possible, will never succeed until libraries cease to rely on hand-craft tools and until they develop mechanical devices for the release of information which are in step with the communicational revolution which has caused their bibliographic woes.

For some 15 years sporadic experiments have been carried out on various mechanical devices—punched cards, magnetic tape, coded metal plates, and coded microfilm. Of these the most widely known are the two types of punched cards, edge-notched and over-all punched. In this country the principal supplier of edge-notched cards is the McBee-Keysort Company; the field of over-all punched cards is dominated by the International Business Machines Corporation.

Keysort cards, best known to librarians, operate on a simple principle. They vary in size, but all have rows of holes punched close to the card margins. When a card is coded, certain holes are converted into notches extending to the edge of the card. Sorting the file of cards is done by running a long needle through the proper hole; the notched cards fall out when the file is lifted on the needle.

Library uses of Keysort cards have been mostly confined to circulation control. Installations for control of subject information

have been made chiefly for individuals or for small special libraries, particularly in various enterprises where technical chemistry is involved. Information can easily be recorded on the cards; even abstracts can be typed on them or inserted in them in windows. The cards can also be searched easily, but not fast enough for large installations. They are not easily reproduced and do not supply a printed record. In short, edge-notched cards seem best adapted to fields in which there are a large number of simple, definite elements which relate to a comparatively small number of core ideas. It is unlikely that they will ever be useful in larger bibliographic installations than special libraries, and in those only if they do not use more than 10,000 cards in a single file.

Over-all punched cards, typified by I.B.M. cards, differ markedly from Keysort cards. In I.B.M. cards rectangular holes are punched by machine in certain fixed locations. "All the holes punched are identical in size and shape, but are given meaning by their location in any one of the possible 960 positions."¹⁴ Sorting must also be done by machine. The newest I.B.M. machines sort cards at 650 a minute; 10,000 cards can be searched in 15 minutes.

For any type of installation, bibliographic or not, I.B.M. equipment, being much more complex than Keysort, requires careful adaptation to the job in hand; its costliness necessitates thorough advance planning; use must be fairly constant to justify the cost; unforeseen difficulties are likely to occur; but savings in money and time *may* be considerable. Two facts about punched cards are worth noting: (1) almost all library installations for any purpose whatever have been made within the last 20 years; and (2) extension of uses from one application to another has been cautious and has re-

¹⁴ Peakes, G. L., "Report Indexing by Punched Cards." *Journal of Chemical Education*, 26:139, March 1949.

quired each library staff to adapt the machines to each new use.

Unlike Keysort cards, the theoretical coding capacity of I.B.M. cards is astronomical. Though the actual capacity is far less (because the cards are generally broken down into fields), it is still more than adequate to code any present or likely future list of headings. Moreover, the I.B.M. people are currently engaged on a development which will permit random punching anywhere on the surface of the card and thus immensely increase the coding capacity.

I.B.M. selecting equipment is several times faster than Keysort. Sorting 40,000 cards an hour sounds like a tremendous speed until one stops to think of library situations in which it might be necessary to search a million cards, which would take upwards of four seven-hour days. Large libraries, the ones most likely to use I.B.M. cards for the control of subject information, are thus faced with a dilemma. The larger they are, the more they need rapid searching. But the larger their collections, the longer would be the search required, and the more searches would be called for.

Can the information on I.B.M. cards be reproduced in usable form? What might be called "invisible" bibliographies, consisting of cards with holes in them, can be produced by reproduction of the cards themselves (at a distance if desired) by means of the card-operated tape punch, followed by the tape-controlled card punch. For reproduction of the actual information on the cards in printed, readable form the best means is the I.B.M. tabulator, which has enabled various agencies to carry out printing and publication successfully. The publication which comes nearest to the appearance of ordinary printing is the index to the *Bibliography of Agriculture*, which has been issued by means of the tabulator. Use of I.B.M. cards makes it possible for monthly indexes to be cumulated into an annual

issue, and for a subject index to be prepared.

The three requirements which any system of control of subject information must fulfil are these: adequate recording, swift and thorough searching, and reproducing in usable form. In selection speed, coding capacity, and release of information in printed form I.B.M. cards excel Keysort cards. But bins of I.B.M. cards take a great deal of space; the machines are complicated and costly to install and operate; and they are still not fast enough to perform the maximum job to be done.

In the literature on punched cards the dimensions of that job have not been clearly marked. Many writers assume that what is required is a search of the material available within a single specialized library. If this is all, either catalog cards or punched cards will do the job, the latter somewhat more speedily. As the information coded into the cards of a single library is limited, so is the information extracted.

Other writers, particularly in the field of science, tacitly assume that what is desired is a search of the total subject bibliography of a special field. Some writers seem to expect also the rapid production of printed, cumulative bibliographies. Such demands make Keysort cards and even I.B.M. cards appear slow and clumsy. Some swifter and more flexible tool is needed.

Such a tool is the Rapid Selector, now being developed at the Department of Agriculture Library. It is based on coded microfilm run at enormous speed past a camera which photographs all items having a given code designation and records them on film or sensitized paper. The machine was developed during the late 1930's at the Massachusetts Institute of Technology, and although it never operated successfully there, the principles on which it was constructed are basically those of the machine at the Department of Agriculture.

Each separate frame on the film is di-

vided into two sections by a line running lengthwise down the center of the film. The information section of each frame contains an abstract, a citation, a photograph, drawings or any other kind of information which can be filmed. Alongside the information section is a code section, which in an extremely small space permits a choice among an extremely large number (at present 10,000,000) of coding positions. As the film runs at high speed through the selector, it passes over a scanning plate set to match the code designation of the particular subject on which the operator wishes information. When the code patterns match exactly, a flash lamp is set off and the selected items are photographed on recording film. Either positives or negatives can be made. Negatives 5x8 inches in size are not entirely without blur, but are readable with little difficulty.

Currently the Rapid Selector scans over 75,000 subjects a minute. The mechanism which moves the recording film into position is not yet able to keep up with the passage of the information film across the scanning plate. Thus if similar subjects are extremely close together, the second may be missed by the recording film unless the projection machine is slowed down. When this technical difficulty is obviated, even higher speeds will be possible.

To date, development and research on the Rapid Selector at the Department of Agriculture has been entirely paid for by the Office of Technical Services in the Department of Commerce. Moreover, the first 30,000 O.S.R.D. reports were run on the selector for purposes of experimentation. Except for the P.B. report¹⁵ which was published by the Office of Technical Services in the summer of 1949, little publicity has

¹⁵ For a summary cf. "Rapid Selector is Newest Research Aid." *Library Journal*, 74:1119-1120, August 1949.

been given the research done, and comparatively few people have even seen the machine in operation. Those who have, agree as to its practicability.

The Rapid Selector is clearly superior to punched cards as a tool for the control of subject information. It can record a variety of materials on a single reel; it is highly economical of space; it handles materials thousands of times faster than punched cards; it gives the searcher his information in the form of a copy of the original. To scientific researchers it can supply bibliographies, with or without abstracts, almost instantaneously.

In short, the Rapid Selector is making so great a forward step in solving the problems of selection of information and of reproduction that the next step must be taken by those who insert material into the record. Since the selector is capable of searching an immense literature and sifting out complete information on a topic, it is now possible to consider whether total coverage shall be attempted. The development of selection by mechanical means does not solve that problem; but it does focus our attention on it more sharply.

Regardless of the degree of totality that may some day be achieved, it is plain that mechanized selection, especially by means of the Rapid Selector, will greatly affect the older bibliographic tools. Classification systems in libraries will become less and less important and will become useful chiefly as a means of assigning call numbers. As for catalogs, at a recent meeting of the Maryland-Virginia group of catalogers Shaw said of the Rapid Selector: "I am convinced it may eventually take the place of the conventional card catalog. I am also convinced it will never take the place of the cataloger." Whether or not the card catalog is doomed to ultimate extinction, it is certain that it will decline in importance; even subject catalogs will dwindle into finding lists, for which

they are today used far more than is admitted. But subject cataloging, the most intellectually demanding of present-day cataloging techniques, will come into its own.

One of the most fertile fields for the new tools is the control of materials which have already been indexed and abstracted, rather than directly of subject content. For example, "it would take the Rapid Selector only about 15 minutes to review all the entries that have appeared in the last 30 years in *Chemical Abstracts*."¹⁶ Hence indexes will continue to expand. And hence subject cataloging, in the sense of analysis of information, will require greater skill and higher standards of subject knowledge. Librarians who can qualify as subject special-

¹⁶ *Ibid.*, p. 1119.

ists will rise in esteem in the eyes of both their own and other professions.

It should be noted that it has been assumed that libraries and librarians will take the initiative in guiding the new bibliographic tools. But it is equally possible that just as indexes and bibliographies were given over to outside agencies, so the new devices will be grasped most eagerly by those who first realize their potentialities. If this comes to pass, if libraries do not seize the opportunity afforded by the new tools, it is all too probable that their research collections will be left to wither into innocuous desuetude while other agencies take over their research functions. Machines do not change human nature; but automotive transportation has certainly disposed of the horse and buggy.

Subject Bibliography Versus Subject Catalog

(Continued from page 214)

MANY; EDUCATION; EVOLUTION; PHILOSOPHY; SCIENCE—STUDY AND TEACHING; MEDICINE—STUDY AND TEACHING; and U.S. DEPARTMENT OF DEFENSE.

7. The time has passed when we need to debate the merits of good subject bibliography. The library profession wishes to assume some leadership in the production of

subject bibliography. However, we do not have an adequate number of capable people in our profession to provide the leadership let alone the army of trained scholars necessary to do the spade work. Perhaps the latter is not a function of our profession, but if we are to assert leadership we must either develop capable bibliographers in adequate numbers within our profession or import them into librarianship from other fields.

Rare Books in the University Library

Rare Books in the University Library, published as Part II of the July 1949, issue of *College and Research Libraries*, and issued as a separate by A.C.R.L., received an impressive review in *The (London) Times Literary Supplement* for April 28, 1950. Copies of *Rare Books in the University Library* are still available at A.C.R.L. Headquarters, American Library Association, Chicago, Ill.