ROBERT BALAY and JOHN GARDNER

An Inexpensive Information Retrieval System Using Coordination of Terms With Edge-Notched Cards

The retrieval system described was designed for a collection of approximately six to ten thousand documents of a wide subject range. After an analysis of the available cheap forms of manual information retrieval systems a unique method of combining coordinate indexing, together with McBee Keysort aperture cards was developed. This method also had the capability of quickly reproducing the results of a search by the use of the Keysort cards as duplicating masters and a special Handiprinter. The system was proved to be eminently practical in operation, and its main advantages were its simplicity, the clerical and professional staff time-saving it offered, and the opportunities for deep analysis of subject matter that it offered.

RESEARCH LIBRARIES ARE often made the depository for unique collections of documents which present unusual features of organization or subject matter, and for which ready-made cataloging is not available. Methods devised to index and control such collections are sometimes elaborate and expensive.[1] In this paper, by way of contrast, we describe an information retrieval system which is both inexpensive and entirely manual in operation and which features random filing, coordinate indexing, and a quick

[1] For a particularly glamorous method, see Abraham Lebowitz, "Mechanization of Legislative Materials at AEC Headquarters Library," *AEC Technical Information Bulletin*, No. 12 (September 1965), p. 3-7.

Mr. Balay is at the Kresge Science Library, Wayne State University, and Mr. Gardner is at the Technical Library, Sandia Corporation, Albuquerque, New Mexico. The work described in this article was performed when the authors were employed by General Precision Aerospace, Little Falls, New Jersey.

means of reproducing the results of a search.

PLANNING

Part of the technical information center at General Precision Aerospace consisted of a collection of proposals, amounting to about six thousand items, which had never been cataloged. A card catalog had been started some years earlier, but the only subject access it provided was a title index, and the catalog had not been kept up to date. Consequently a great deal of new material had never been indexed, and to make matters worse, portions of the card catalog had been destroyed. The proposal collection was used steadily (about fifteen queries per week) and it grew fairly quickly, since a copy of every proposal submitted by the Aerospace Group[2] was deposited in the technical information center. Proposals were filed

[2] The group is made up of the systems division, the Kearfott division and the aerospace research center.

in alphanumeric order by the proposal code and could be found in most cases only if the requestor knew the code number. A quarterly listing of proposals submitted, issued by the publications department, provided access by title but only in chronological order of publication; hence the listing had to be searched item by item.

Since a great deal of the research conducted by the group was reported only in this body of proposal literature and was not readily available in any other place, the need for some sort of index to the collection was apparent. Accordingly, we began to investigate existing methods of information retrieval. Before doing so, however, we formulated a set of requirements for an acceptable system. These were:

1. The system had to be under the control of the technical information center, so that it could be used at any time.
2. Because of the shortage of staff time, the system had to operate with an absolute minimum of clerical effort.
3. The system had to permit indexing in depth, or coordinate indexing and searching techniques, or both.
4. The system could not demand refiling of the collection in any other than its current order, both because of the confusion that would occur during refiling and because the clerical time involved would be intolerable.
5. The system had to be inexpensive.
6. The system had to be simple so that a great part of its operation could be trusted to clerks, of whom little training would be required.
7. Extreme speed of retrieval would *not* be necessary; ten to thirty minutes of retrieval time per query would be acceptable.

With these requirements in mind, we proceeded to examine the available systems.

Requirements (1) and (5) at once eliminated the use of systems requiring computers or data processing equipment. The use of Termatrex or one of its imitators was seriously considered, since it satisfied most of the requirements; but this class of retrieval device was rejected because it would exceed the available funds (5) and would require refiling the collection (4). The conventional library card catalog, while flexible and inexpensive, does not permit coordinate indexing (3) and demands a great deal of card handling, typing, and filing (2). Finally, Uniterm cards, the most attractive of the existing systems, chiefly because of their simplicity and low cost, were rejected because of the filing and posting time involved (2) and because their use would have demanded a reordering of the collection (4).[3]

In this way the readily available systems were found, for various reasons, unsuitable for the needs of the collection, and we were forced to design a unique system. Edge-notched cards had been considered during our analysis of existing systems, but the methods in use were either very complex[4] or admitted of too few coding positions.[5] If the coding capacity of the cards could be increased while their operation was simplified, edge-notched cards could be made to satisfy our needs.

Edge-notched cards may be used in one of two ways: either the card represents a subject and documents are coded around its edge,[6] or the card represents

---

[3] In the Uniterm system, documents must either be filed by their Uniterm number or a file correlating Uniterm numbers with the document shelf list must be created. Either method uses up clerical time.

[4] See, for example, Gerald J. Cox and others, "Punch Cards for a Chemical Bibliography," *Chemical and Engineering News*, XXIII (September 25, 1945), 1623-26.

[5] John G. Wagner's system, for instance, provides only 116 coding positions (see "Manually Sorted Punched Card System for Pharmaceutical Literature," *Journal of Pharmaceutical Sciences*, LI [May 1962], 481-84).

[6] Wagner's terminology (p. 481) is perhaps preferable here; for "documents" he uses "individuals," for "subjects," "characteristics."

a document and subjects are coded around the edges. In the first application, the cards make up an inverted file and cards must be kept in strict alphabetical order.[7] This application resembles the Uniterm method and would incorporate the disadvantages of Uniterm which were described above. The second method has two advantages: it does not require refiling the collection of documents, and it permits coordinate retrieval by providing a means of comparing subject terms. Terms are compared in this application by passing the sorting needle through the deck of cards a few times; thus, if one were searching for digital integrating accelerometers, the needle would be passed through the entire deck of cards at the position coded for "accelerometer," a second time through the smaller group of cards thus selected at the "digital" position, then a third time through the remaining cards at the "integrating" position. The cards which drop out after this final sort will be the ones which deal with the subject under search.

By employing this second method, edge-notched cards could be made to satisfy most of our requirements: they were inexpensive, they allowed coordinate indexing and retrieval, they did not require refiling the collection, and they would be under the control of the technical information center. It was found that their use could be greatly simplified by reserving all the punching positions for descriptor codes and by using only direct coding. A drawback still seemed to be the small number of notching positions available; even the larger 5 x 8 inch cards contained only about two hundred and fifty holes, and if each hole represented a subject, the system would be restricted to two hundred and fifty subject terms. A solution—to expand the number of notching positions by

[7] See, for example, J. G. Roney, "Inverted Indexing on Edge-Notched Cards," *Science*, CXLII (October 1963), 227-28.

using combinations of holes—was quickly hit upon, and the proposed system now seemed to satisfy our requirements. It was therefore decided to adopt the system.

CARD DESIGN

The final card design, which evolved over a period of two months, is shown in Figure 1. The card layout was de-
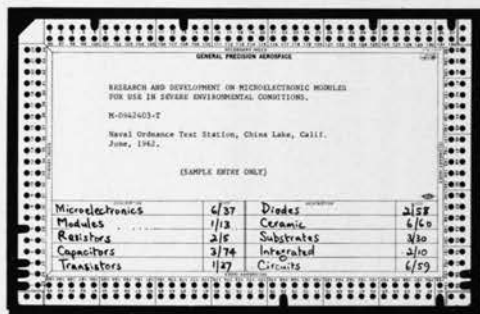

FIG. 1.—Sample of an edge-notched Keysort card for a fictitious proposal.

veloped with the assistance of a representative of McBee Systems, and after the approval of a dummy, a quantity was ordered, printed, and delivered. On the cards the two rows of notching positions are divided into two parts, a primary index along the lefthand margin and a secondary index around the other three margins. Each descriptor is assigned a two-part number, the first part being punched in the primary index, the second in the secondary index. Thus the term "reconnaissance," coded 6/58, will require a punch at the 6 position in the primary index and at the 58 position in the secondary index. By using combinations of holes in this manner, the 234 separate coding positions on the card can be made to accommodate 8,360 subject entries. Space is provided on the card for recording the descriptors used and their code numbers.

Another feature of the card is the aperture on which the title and other bibliographic information are typed. The aperture is covered with a special duplicating paper plate; before typing

the bibliographic notation on this paper, the typist backs the aperture with a sheet of hectograph carbon, supplied by McBee Systems; when the notation is typed, a reproducible master is deposited on the back of the aperture paper. This can be reproduced on 3 x 5 cards for making auxiliary or supplementary indexes, or the cards resulting from a search may be reproduced to form a bibliography. A portable spirit-type duplicator, called a Handiprinter, which consists of a pad and roller together with a tubular spirit tank serving as a handle (see Figure 2), is sold by the card manu-

FIG. 2.—The McBee Systems Handiprinter. The knob at left will release the spirit duplicating fluid.

facturer for this purpose. The Handiprinter is filled with spirit and the damp felt pad is passed over the card on which one wishes to reproduce. The aperture card is then laid face up on the moist paper, the Handiprinter is rolled over it, and the information typed on the master is transferred to the paper.

INDEXING

One of the aims of this retrieval system was to use as little time—both pro-

fessional and clerical—as possible. Accordingly, some shortcuts were adopted. Since it is rare to see an author's name on a proposal, this item was eliminated from the bibliographic notation. Because all the items to be indexed were proposals, and all originated at General Precision, there seemed to be no reason for recording this information. This left only four items to be recorded in the bibliographic entry: the proposal's code number, its date, its title, and the agency to which it was addressed.

All typing was done by a clerk from information provided by the cataloger. Descriptors and their codes were written directly on the Keysort card (an average of 10-15 descriptors for each document) and the bibliographic particulars were marked on the title page of the document with appropriate symbols (title in " ", date circled, and proposal number and addressee underlined). The report, with the Keysort card enclosed, was then passed to a clerk who typed the bibliographic information on the card according to a predesigned format and notched the appropriate numbers on the edge of the card. The document was then returned to the shelves and the Keysort card filed with others already prepared. Since the documents on the shelves were kept in proposal code order, they constituted a shelf list of the collection; there was, therefore, no need to file the Keysort cards in any particular order or to refile them in a special order after they were used, and they could be kept in random sequence.

Subject control was maintained by two devices: a numerical code list and an alphabetic descriptor list. The numerical code list was prepared in advance and consisted of a sequential listing of code numbers, thus: 1/1, 1/2, 1/3, . . . 1/95; 2/1, 2/2, 2/3, and so on. When a new descriptor was used, a code number was assigned from this list, and the number was then crossed off so it

could not be used again. The descriptor was written on a 3 x 5 card with its code number and filed alphabetically. The cataloger assigned descriptors and code numbers from this file or made up new cards with new code numbers when new descriptors were required. The file was reviewed periodically for synonymous terms for a trial period during which 150 proposals were indexed. If it had been considered necessary, title and addressee indexes could have been prepared using the Handiprinter. These indexes would have been arranged in alphabetical order.

SEARCHING

To query the system, the indexing procedure is reversed. The requestor announces his needs and is quizzed about his topic in accordance with good reference practice. Descriptors are arrived at which characterize his needs. These terms are noted, looked up in the alphabetical descriptor file, and the corresponding code numbers noted. The pack of Keysort cards is then needled for these code numbers.[8] Two passes are required for each descriptor, one in the secondary index, one in the primary index. If more than one descriptor is being searched, the term likely to occur least often is needled first in order to reduce the number of cards to be needled on passes two, three, four, and so on. This process provides comparison of terms, the effect being similar to that obtained in the Uniterm system where document numbers on descriptor cards are compared.

It should be pointed out that it is seldom necessary to needle twice for each term being searched. For example, if one is searching for documents on the fabrication of ceramic diodes for microelectronic modules for use in severe en-

vironments, a descriptor list such as this might be compiled:

DESCRIPTOR	CODE
Ceramic	6/60
Diode	2/58
Microelectronic	6/37
Modules	1/13
Environment	9/71

To conduct the search, one might disregard the needling strategy described above and proceed arbitrarily, needling in the secondary index 60, 58, 37, 13, and 71, and in the primary index 6, 2, 1, and 9. In practice it would not often be necessary to make as many passes as this example enumerates; after the third or fourth pass, the pack of remaining cards will ordinarily be reduced so that their titles can be scanned quickly without making further passes.

CONCLUSIONS

The information retrieval system described here has proved satisfactory in operation. Its simplicity makes it easy for clerical assistants to understand and operate; it provides a form of coordinate retrieval; and it offers a number of clerical shortcuts (random filing of cards, single typing of bibliographic citation, a means of quickly reproducing citations to make up bibliographies) that result in a great saving of clerical time.

It is not, however, in its present form, suitable for large collections. Unlike Uniterm, which for a given search considers only those documents entered on the Uniterm cards chosen, the system described here considers every indexed document in the collection during every search. As the collection grows, so does the file of Keysort cards, and the time and labor involved in needling becomes correspondingly greater.[9] The system

[8] The needling procedure is fully described in several places; for example, by Robert S. Casey and James W. Perry, "Elementary Manipulations of Hand-Sorted Punched Cards," in Robert S. Casey and others, eds., *Punched Cards: Their Applications to Science and Industry*, 2d ed. (New York; Reinhold, 1958), p. 12-29.

[9] One device adopted to reduce needling time was to use only shallow punches until they were exhausted, then assigned code numbers requiring deep punches. Thus, numbers 1/1 through 1/95 were assigned, then 2/1 through 2/95, and so on until 22/95. A deep punch, of course, requires two passes, a shallow punch only one.

might be adapted to larger collections by color-coding the cards to represent large subject areas, but for the present it seems advisable to limit the size of collections for which the system is used to ten thousand documents.

False drops have presented some problems. If one is needling, for instance, for terms coded 3/45 and 7/68, cards coded 3/68 and 7/45 will drop out also. In practice this has not been found to be a serious hindrance. Rarely are more than ten items turned up by a search, particularly if one has been careful to define the subject carefully, and it is a simple matter to scan the titles and reject the unsuitable cards. Most queries are so specific that several terms are required to describe them adequately, the multiple passes needed lower the number of documents yielded by a search and provide a cross-check to lower the number of false drops.[10]

This retrieval system is recommended for libraries having small special collections and has as its main advantages coordinate retrieval and low expenditure of clerical time. ■■

NOTE: The authors would like to express their thanks to McBee Systems for their help with those aspects of this article specifically pertaining to their equipment.

[10] The problem of false drops is considered at length by A. K. Soper, "Some Observations on the Use of Punched Cards for a Personal Information File," *Aslib Proceedings,* VII (1955), 251-58.

Annual CRL Index

THE ANNUAL INDEX for *CRL* and its *ACRL News* issues will be published this year in the December *ACRL News* issue. Since the news issues of *CRL* are not at present available on subscriptions, subscribers copies' of title page and index for volume XXVII will be available on request after December 10, from the ACRL office, ALA Headquarters, 50 E. Huron St., Chicago 60611.