

A Computer Produced Shelf List

The Harvard University library has developed a system for converting its sheaf shelf lists to machine readable form and for using the computer to produce a new type of three-part shelf list which will provide readers with a new approach to the library's collections and the library staff with a more accurate and efficient inventory and classification tool.

THE 2,225,000 VOLUMES in the Widener library are represented by an estimated 1,600,000 entries in what is probably the only remaining sheaf shelf list in any large American library. A few of the two hundred classes in the Harvard classification scheme are on cards, but the major portion of the shelf list continues to be maintained on handwritten or typewritten sheets in looseleaf binders. This type of list is difficult to maintain, and errors tend to accumulate each time the contents of a full page have to be transcribed to several fresh pages to permit the addition of new entries. The disadvantages of maintaining a shelf list in this antiquated and inefficient form are obvious, but there are two important advantages that are worth citing—it takes less space than a card list, and it is easier to consult.

The shelf list was designed to serve two essential purposes. It is an inventory record of what is in the library, and it is an indispensable tool for the assigning of class numbers to new books. The first function is no longer as important as it once was, because the library has now virtually ceased taking inventory or reading shelves from the shelf list.

In the absence of a classified catalog the shelf list has long been used by librarians and a few other sophisticated library users to provide a means of systematically surveying the library's holdings on a particular subject. When perusing the shelf list one sees all the titles that have been classified in a given segment, and not merely those which happen to be on the shelves or whose spine labels are legible. In addition, one can take in at a glance the essential bibliographical description of a book—its author, title, place and date of publication.

It is not surprising that this potential usefulness of the shelf list to readers has been generally unrealized. The shelf list was never meant for the use of readers—it exists in only one copy, and that copy has always been located in staff areas that are relatively inaccessible to the public. While readers have always been permitted to use the shelf list, they have never been encouraged to do so, and the staff has made no special effort to explain how it could be used. A card shelf list is not easy to consult because the cards can be viewed only one at a time; the sheaf shelf list in Widener is for the most part handwritten and difficult to read. To use a shelf list intelligently and efficiently a reader needs access to the

Mr. De Gennaro is Assistant University Librarian in Harvard University.

classification schedules to serve as a guide, index, or table of contents of the list. The classification schedules used in Widener have never been published and are available only in typescript form in the catalog department—one floor above the room in which the shelf list is located.

Conversion of the shelf list to standard library cards has been considered and rejected on a number of occasions in the past on the grounds that the expense involved would not be worth the advantage gained—and the advantage was always thought of in terms of the present limited function of the shelf list as a classifier's tool. Computer technology now makes it possible to enlarge the concept and to expand the uses of the shelf list while improving techniques of maintaining it and making it available. This consideration has supported the idea that it would be desirable to convert the shelf list to machine readable form, and a plan for doing this has been formulated and is being tested.

The plan calls for converting the shelf list class by class and using a computer to produce a new kind of three-part list for each class. The first part consists of a printout of the classification schedule, which will serve as an index or key to the second part. The second part is a listing of the titles in the class in call number or classification sequence with class headings interspersed at the appropriate places in the list. These two parts together serve as a kind of classified catalog or browsing guide to the particular subject area covered by the class.

Part three is a printout in alphabetical sequence by main entry of the entries in the list and is obtained by a programmed computer sort. This provides a new approach to the books in a classification and serves as a brief finding list for books in a subject area. By using this list the specialized reader might avoid much of the frustration of consulting a six million card catalog for the call

numbers of the books he would like to consult. The entries in the Widener shelf list are frequently much less complete than those in the public catalog, but in most cases they will be adequate as a quick reference or finding entry. Readers would use the public catalog when they needed more complete bibliographical entries. It should be emphasized that these class lists will supplement rather than replace the subject approach of the public catalog.

Once they have been converted into machine readable form, the shelf lists can be produced in multiple copies at a reasonable cost, or if a demand for them materializes, they can be photographically reduced and published in small editions. Copies of the lists of a particular class will be located and prominently displayed in the stack with the class in order to provide a guide to intelligent and systematic browsing. Professors and graduate students could obtain copies of the lists for classes of particular interest to them, and could use these to do a good deal of preliminary bibliographic work and to familiarize themselves with the holdings of the library in their own fields. Copies of the lists could be made available to scholars or libraries away from Cambridge; in fact, a significant demand might materialize for lists in subject areas where the Widener library is particularly strong, such as American history, Slavic, and the various other literature classes.

Obviously these lists will have their deficiencies. No classification is perfect, and books are not always placed where the reader would expect to find them or brought together as would be desirable. Rare books in the Houghton library, infrequently used books in the New England deposit library, and related works in other Harvard collections will not be listed; important material contained in journals will not be analyzed. There are errors and inconsistencies in the shelf

list, most of which will not be eliminated in the process of keypunching.

All catalogs are incomplete and incorrect to some extent, however, and it seems clear that these lists would be useful; indeed, they could be expected to contribute to the education of the library's patrons and might even instill in them increased respect and understanding for the bibliographical apparatus that librarians have so laboriously constructed during the last century. It is a well known fact that readers appreciate the privilege of browsing in a stack of classified books, but unfortunately it is also a fact that few readers ever succeed in grasping anything more than the barest outlines of a classification system. A large part of the effort and intelligence that librarians have invested in devising and using classification systems is lost because we have never found a satisfactory way of sharing knowledge of the system with the readers for whom it is designed and maintained. With the new three-part shelf list that has been described, the librarian will equip the reader with a copy of the ground rules that he is using to organize the collections, together with lists in classified and main entry order of the books and journals in the various classes. The effect should be to convert library stack browsing from a frustrating hit-or-miss activity to a systematic and effective intellectual experience. Indeed, it will also provide the possibility of browsing conveniently in a list without actually having to go into the stack at all. This might reduce the objections to shelving infrequently used books by size in storage, because these titles could be retained in the classified shelf list with a symbol showing the actual location of the book.

The preceding paragraphs have emphasized the value of a computer-produced shelf list to the library's readers. The benefits to the catalog department will be equally dramatic. A machine-

produced shelf list will be a vast improvement in accuracy and legibility over the present handwritten sheets, and it will facilitate the process of assigning numbers to new books. The continuing and tedious process of recopying pages that have been filled in the existing lists will of course be eliminated once the initial conversion has been completed. Accurate volume counts of the collection will be automatically obtained and analysis of classes by language and date of publication will also be possible. This latter might be a useful tool for aiding in identifying categories of books which are deteriorating or which might be transferred to the deposit library. Since serial entries will be identifiable as such it will be possible to get printouts of serial titles in the various classes and in the library as a whole once the conversion is completed. It might once again be practicable to undertake a regular inventory because printed lists that could be marked up would materially facilitate such a project.

In the first phase of the project the official copy of the shelf list will be a special printout with several spaces between entries. The procedure for adding a new entry is to copy it by hand into the appropriate place at the time the class number is selected and pre-empted. A card will also be punched and the new cards will be accumulated and periodically merged by machine into the master tape before an updated version of the list is to be printed. At a later phase, random access disc storage might replace the magnetic tapes, and file maintenance would be greatly simplified. Since maintenance of the lists will not be an on-line operation initially, it will be convenient and economical to rent time on equipment in the university's computation center.

The process of producing catalog copy in machine readable form is under study, and when such a system is implemented the shelf list copy might become a by-

product of this process. Also under consideration is the feasibility of generating book cards from the shelf list conversion project in order to computerize the present punched card circulation system. It will be a simple matter to produce monthly accessions lists in classified order from the punched card entries that are created for shelf-listing purposes.

The chief difficulty to be overcome is the actual conversion of an estimated one million six hundred thousand handwritten or typed entries in the shelf list to machine readable form. This is being done by manually keypunching the entries on cards. The process is a tedious one and other possible methods are being studied. The shelf list entries are not as complete as the entries that appear in the public catalog, and nearly 80 per cent of them occupy only one line in the machine printout. The conversion problem is formidable but not nearly as formidable as would be the case if the project were to convert the one million six hundred thousand main entries in the public catalog. These entries are not only more detailed, but they would have to be identified and the entire file would have to be converted before any use could be made of them. The shelf list treats the collection as a series of separate units, each of which is complete and useable by itself. The classes can be converted one at a time, and there is no need to wait until the entire shelf list is finished before realizing full benefits from the work that has been done. Indeed, some of the larger classes, such as American history, could be subdivided and converted a portion at a time. Several other advantages arise from the fact that it is possible to deal with the shelf list one segment at a time—it may be easier to get financial support for a series of separate projects, priority can be given to active and otherwise exceptional classes, and improvements in techniques, format, type fonts, etc. can be introduced as they are developed.

This same "divide and conquer" strategy that is used in shelf list conversion may very well prove to be the answer to the problem of converting, or more properly, re-creating the dictionary catalog in machine readable and hence book form. The prospect of mounting a project to convert a large dictionary catalog by copying main entries in a straight alphabetical sequence is not one that inspires prudent librarians with enthusiasm. If however, instead of focusing on converting the catalog, we focus on converting unit card shelf list entries with tracings and re-creating the catalog by manipulating these entries in the computer we have reduced considerably the hazards of such a project. An analysis of circulation statistics will show that a significant proportion of the total use of a collection is centered in a small number of classes. The most active classes also tend to be the most rapidly growing. Thus, by converting the shelf lists of the several most active classes and creating machine readable entries for current accessions, one can begin to create a machine-produced book catalog of the most used and most interesting segments of a collection. It might be possible to account for a large percentage of the use of a traditional card catalog by converting a relatively small number of shelf lists. The goal should be the eventual conversion of all or nearly all of the shelf lists so that a complete catalog could be constructed, but the effort could be carried on over a number of years.

This strategy is intriguing, but unfortunately the Widener shelf lists are not of the unit card type, and before the strategy could be used there considerable time and resources would have to be invested in assembling, by one means or another, the information that is contained in a unit record shelf list. The problem of converting the catalog by this method is receiving serious study but the decision has been that the payoff

in computerizing the manuscript shelf lists is significant enough to warrant a reasonable expenditure of effort. It was decided that, initially at least, Widener would content itself with shelf lists that were in the standard upper case machine font. This reduced the time and cost of keypunching very considerably, and it was felt that the appearance and legibility that would be obtained with upper case were commensurate with the level of quality that prevails and is required in the present shelf lists.

Early in the effort it was evident that a small pilot project would give valuable experience which would make it possible to refine estimates of the time and cost involved in converting handwritten shelf lists to machine readable form and would also serve as a test of the formats and programs that would be required to produce the lists. The Crusades classification was selected as the pilot project because it is the smallest class in the Widener stacks and yet it contains an excellent representation of the handwriting, languages, and problems that would be encountered in most of the other classes. There were 1,170 titles (1,400 volumes) which were keypunched at the rate of about thirty entries per hour. Some proofreading was done, but the entries were not machine verified, and it is doubtful that such verification would be worth the effort. Two versions of the classification schedule were made. The existing form of the schedule was reproduced to serve as the contents and index to the list. In addition, each heading in the schedule together with its corresponding call number was keypunched on a card, and these cards were merged by machine with the main deck of cards containing the shelf list entries.

The Harvard College library call numbers lend themselves fairly well to machine coding, since they follow a predictable pattern, usually with a mnemonic letter prefix, a base number, and

one or more dot numbers, *e.g.*, Econ 7042.35.10. The dot numbers are whole numbers, not decimals. The notation can be quite long, however, particularly when the necessary blanks are left to preserve logical order. For instance, Phil 35.9 cannot simply be punched as such, because other prefixes may require five letters, other base numbers up to five digits, and other dot numbers up to four. It is thus necessary to punch the number as Phil [4 spaces] 35 [3 spaces] 9. A particular problem arises with certain dot numbers beginning with zeros, *e.g.*, US 50.02, which is supposed to file ahead of US 50.1. Since zeros follow blanks in machine filing, this presented a complication which was solved by leaving one field blank before the zero dot number. This preserves correct filing order, but necessitated some reprogramming in order to avoid printing two dots, which are not keypunched but are reinserted in the numbers by programing in the printout. Programing steps can also recompact the call numbers so that they will appear in their normal format in the printout. Most of the problems involved in making the call numbers susceptible to machine filing were faced and solved at the time the unit record circulation system was installed in the library.

The call number, card code, volume count, language code, and other necessary fixed fields require over thirty columns of the eighty-column card, but insofar as this information is printed, it is printed only once for each entry. Therefore, a large portion of the 132-position print line seemed likely to be wasted, and it was decided to print two cards on one line and thus utilize all 132 positions. The call number, etc., occupies columns 1-33 on all cards. As formatted for printing, the corresponding information, with blanks needed for visual spacing, occupies the first 38 print positions. Columns 34-80 of the first and

COMPUTER SHELF LIST

(Continued from page 315)

other odd-numbered cards go into print positions 39-85, and columns 34-80 of the second and other even-numbered cards occupy positions 86-132. Normal word division is used at the end of even-numbered cards, while odd-numbered cards normally end in the middle of a word. A dollar sign appears in column 80 of the last card as an end of record code, but it is not printed. Non-printing marks were placed before the title and date of imprint for future uses, such as chronological printouts of certain classifications for special purposes. With a maximum of 94 positions for author, title, date, etc. on each line (minus one for the dollar code), it has been found that a very large proportion of the entries require only two cards and will go on one line. The programs are written in Autocoder. The classified list program, after certain initializing steps such as printing a title page, reads a card, determines whether it is a regular card (for

a book or journal) or a classification heading card, and formats and prints it accordingly. Error checks are provided to point out any call number duplicated or out of sequence, and any cards missing or out of sequence within a set. The classification headings are programmed to overprint so that they stand out on the page like bold face. The alphabetical list is produced by a program which reads the data onto tape and prepares it to be sorted by a canned sort program on the IBM 7094. The list itself is printed from a tape on the IBM 1401.

In order to test formats, printing costs, and user reactions, a small edition of the three-part Crusades shelf list will be produced by offset and distributed to interested faculty members, graduate students, and others. This will mark the end of the pilot project phase and the beginning of routine conversion operations. Two small periodical classes have already been converted, and keypunching of the Africa classification is now in progress. ■■

New and Interesting

A FIVE-YEAR cumulative index for volumes XXI to XXV (January 1960 through November 1964) of *CRL* is being published this autumn. It will be sent free of charge and without special requests to all current members of ACRL. All others may purchase the index for \$1.00 per copy, from ACRL headquarters office, 50 E. Huron St., Chicago 60611. Payment with orders is requested.

The *Supplement to Library Statistics of Colleges and Universities, 1963/64* was compiled by the Library Services Branch of USOE and published by the Library Administration Division of ALA; it provides institutional data on 247 libraries. Copies are available for 75¢ each from the Library Administration Division at ALA headquarters. ■■