notis: an integrated computer system for a large research library

james s. aagaard
head, information systems development office
northwestern university library
evanston, illinois

The Northwestern On-line Total Integrated System (NOTIS) began operation eight years ago with a pioneering circulation control system which featured self-service book charges. Technical processing functions were added to the system in 1971. After several years of successful operation an extensive redesign was started to take advantage of what had been learned and of new computer technology. The new version of the system (NOTIS 3) became operational for Northwestern in the spring of 1977; the libraries of the Garrett Evangelical and Seabury Western theological seminaries had used it for a year prior to that time.

It is the purpose of this paper to present an overall description of NOTIS 3, particularly with respect to technical service functions. The basic design of the circulation part of the system was described in 1972[1] and has not changed, although the physical file structure has been brought up-to-date and different terminals are now being used. Also, a self-service circulation inquiry function has been added, which has proven very popular with library users.

Northwestern University has received no outside support for the development or operation of the NOTIS system, and therefore one of the most fundamental considerations was economy. We believed that the state of the art in library automation was a long way from producing a sufficiently general system that would serve adequately the needs of a small public library, a large university library, and everything in between or beyond. Thus the original design, although for an integrated system supporting all phases of technical processing, included only the capabilities needed by the Northwestern library. Subsequent improvements and reductions in cost of computer hardware allowed for a more general viewpoint in the design of NOTIS 3, and we believe the current system would be applicable to a wide variety of libraries. It is still an integrated system, however, and it probably would not prove economical to implement it as only an acquisitions system or only a serials system. As with most computer systems, cost savings can be achieved only when the same data in machine-readable form can be used for many different purposes.

File Structure

The two basic principles which guided the design of the file structure for NOTIS 3 were full MARC compatibility and simplicity. Commercial data base management systems were eliminated on the basis of cost, processing overhead, failure to accommodate variable length fields, and difficulty of recovery after a system "crash." In the original NOTIS system all information related to a title was converted into a MARC-like form and stored in a single physical record. We discovered that this often resulted in unwieldy records, sometimes longer than 5,000 characters, and that many requests did not need the full record. An analysis of the problem led to the structure used in NOTIS 3, in which there are three types of records; bibliographic, holdings, and order. These may be stored in three separate files, in a single file, or in two files (with bibliographic and holdings records in one file and records in the other). The choice of one of these options is dependent on the size of the files and on the number of institutions using the system.

Bibliographic records are essentially exactly as defined in the MARC formats, except for the addition of a few provisional fields which are used when full cataloging information is not available immediately. The system control identifier assigned to each bibliographic record is composed of three letters and four digits. This combination will allow nearly 140 million unique identifiers and is easier to use than a longer all-numeric identifier.

Bibliographic records are found to average

nearly 500 characters in length; however once cataloging is completed they rarely need update. This offers the possibility of providing a two-level storage structure for them, consisting of an inactive file and an active file. Records in the inactive file are packed tightly together and cannot be updated, but they can be displayed. Furthermore, when a terminal operator attempts to update such a record, it is automatically restored to the active file in updated form. Thus the existence of the inactive file is nearly transparent to the user of the system. It is expected that records will be transferred from the active file to the inactive file two to four times a year; the criteria for selecting likely candidates are still under evaluation. At the present time the Northwestern library has nearly 300,000 bibliographic records available on-line; about 90 percent of them are in the inactive file. This represents a saving of about 130 million characters of direct-access disk storage.

The use of the inactive bibliographic file in the NOTIS system is optional; it appears to be worthwhile when the number of bibliographic records is greater than about 50,000. As the number approaches a half million or more, more drastic approaches may be needed. One which is being considered is to remove many of the note and added entry fields from the bibliographic record in the on-line inactive file and transfer them to an off-line inactive file. Still to be found is a way to minimize the inconvenience when such a record must be updated.

Holdings records are subdivided into two types; copy holdings and volume holdings. The internal control identifier for a copy holdings record is the identifier of the associated bibliographic record, followed by a two-character code indicating the institution, followed by a zero. This allows a number of institutions to share the same bibliographic records. The copy holdings record contains one field for each copy of an item held by the institution; the field contains the location (such as a branch library) and call number of that copy. There is also provision for note fields at the title level and the copy level. A copy control number is automatically assigned to each copy; this may or may not be the same as the copy number assigned by the cataloger, depending on the practices of the institution. The present system is limited to 255 copies of a title per institution.

The volume holdings records are optional; their primary purpose is to record bound holdings of serials. They are used in other cases where the holdings are too complex to be recorded in the copy holdings record as a part of the call number field. Each copy may have a separate volume holdings record; the internal identification is the same as the copy holdings record except that the zero is replaced by the copy control number.

The average size of a copy holdings record is about 100 characters. Volume holdings records also average about 100 characters; however there are only about one-sixth as many. Thus the average amount of holdings information for each title is about 116 characters. These records are subject to more frequent update than the bibliographic records as the result of the acquisition of added copies or volumes, and therefore are maintained in active storage.

The third type of record is the order record. There may be many orders for a given title, up to a current limit of 255, and each order may be for as many as 12 copies (for practical reasons, serial orders are usually limited to one copy per order). The internal control number for an order is the institution code, followed by the bibliographic record identifier, followed by a system assigned sequential number. The sequential number followed by the bibliographic identifier is used as an external order number.

The order record contains all of the usual information relating to an order, such as vendor, price, items ordered, items received, invoice amount, currency code, fund code, and shipping information. It is designed to facilitate the link to fund accounting and invoice payment modules in the future. For serials the order record serves as the "check-in" record, and always contains the current unbound holdings.

In addition to information in the order record which pertains to the order as a whole, there is provision for several types of fields. These are for the scope of the order, receipts, payments, and notes. Each type of field can have an "action date" assigned by the terminal operator; in the case of the last receipt statement the action date is automatically provided by the system but can be overridden. The action date generates a notification (produced daily on a batch basis) when it expires; a terminal operator then inspects the record to determine what action is needed. This frequently takes the form of a claim notice, which is triggered by the entry of a special note field into the record.

We find that order records average about 200

characters in length. Once a year, records for orders completed more than one year prior to that time are purged from the active file and cumulated on microfiche. Also, since orders for serials may continue indefinitely, such records may be partially purged of old payment fields.

The structure of the order number permits the system to locate the corresponding bibliographic and holdings records when the order number is known. To permit access in the other direction, the copy holdings record includes a linkage field which contains the number of each order for each copy, along with a one-character order status code. Starting with a display of the holdings information, this usually permits the selection of the desired order record on the first attempt. The linkage is simple enough that it is not susceptible to significant difficulties due to system "crashes" during its update.

Bibliographic Access

The NOTIS 3 system employs a unique method to provide author/title access to the bibliographic files. Index entries, the length of one line on a display terminal, are created from each record. In most cases each entry consists of three segments, the first approximately 48 characters long, the second 17 characters long, and the third 4 characters long. The exact lengths of the first and second segments will vary to make optimum use of the available space.

An entry will be created using the main entry, series entries, and each added entry (except subjects) as the first segment, deriving the second segment from the title field. One index entry is created using the title as the first segment and the main entry as the second. In each case the third segment contains the year of publication for monographs and the first four characters of the place of publication for serials.

These index entries are sorted in alphabetical order and stored on direct-access storage devices in a compressed form. A terminal operator can enter as many characters of an author or title as desired, and the index will be displayed, 16 lines at a time, starting at that location. This provides a browsable access method, which is much closer to the traditional card catalog than other computer systems. If further information is needed, the terminal operator can display brief bibliographic data for any index entry at the bottom of the screen without losing the index display. Once the proper record is selected, display of the full bibliographic record or of brief bibliographic data plus holdings may be obtained.

This type of index requires somewhat more storage space than search-code type indexes, but is easier to maintain on a real-time basis than inverted word indexes. The average number of index entries per bibliographic record has been found to be about 2.5, so that the index for the 300,000 record file mentioned earlier has over 700,000 entries. With compression, and allowing some free space to facilitate real-time additions, it requires about 50 million characters of storage. Since the entries are entirely in upper case, a further reduction could be achieved by using only 6 bits for each character instead of 8 bits.

A subject access capability is still in the preliminary design phase, but initially it appears that a different approach will be used.

Application to Technical Processing

Usage of the NOTIS 3 system begins at the point of considering the acquisition of a new item, with the search of the existing data base. In some cases search of the card catalog also may be necessary. If it is decided to acquire the item, bibliographic and holdings records are created for it, and at least provisional author and title information entered.

The integrated system is most effective if the full bibliographic information can be obtained at the earliest possible stage of processing. Northwestern maintains the entire MARC file off-line (on tapes) with LC card number, ISBN/ISSN, and search code indexes on disk. If an item is expected to be in MARC scope, the terminal operator requests a search of the MARC index; these searches are performed once a day on a batch basis. Records which are found generate selection requests which will eventually cause the complete MARC record to be transferred from the tape files into the on-line file; this may occur the same day or as much as several weeks later, depending on the age of the MARC record and the processing schedule for the tapes. If the MARC record is not found in the initial search, the request is kept on file and matched against each new weekly MARC tape as it is received. For items which are not in MARC scope, other sources, such as the National Union Catalog, are used and the data entered through a terminal.

Once the bibliographic record is created, with at least provisional information, a terminal operator

can create an order record. This process will cause a purchase order to be generated during the daily batch processing.

When the item is received from the vendor, its receipt is recorded in the order record and a paper temporary catalog card and a cataloging work sheet are produced. The work sheet is used to complete the cataloging process, and the final step is to request production of catalog cards (done once a week), book labeling materials, and punched book cards for circulation and inventory.

All bibliographic information except that which is transferred from MARC records is entered on-line. The system is designed to take full advantage of the capabilities of modern display terminals; for example parts of the display which the operator is not authorized to change are protected, and an entire screen may be entered or changed before transmission to the computer is necessary. To further aid the operator, the bibliographic display uses LC-assigned mnemonic field tags and indicators rather than the numeric forms. All indicators and delimiters are checked by the computer programs for validity in the particular fields in which they appear, and all single-character codes in the fixed fields are checked against tables of valid codes. MARC books and serial formats are now available; others are in the process of implementation at this writing.

Costs

The NOTIS system at Northwestern operates on an IBM/370 Model 138 computer (with 512K bytes of processor storage) which is shared by other administrative functions. Under circumstances such as these, charges to the user may vary considerably, depending on the policies adopted for allocating costs. At the present time charges are being made on a transaction basis, even though it is known that the system resources required for a given transaction are subject to wide variations. The present rate is $.04 per transaction, and technical services usage is about 90,000 transactions or $3,600.00 per month. In addition there is a monthly charge of $5.00 per million characters of on-line disk storage. To this must be added the cost of terminals in the library and batch processing charges (about $2,500.00 per month). This covers all processing for a library which acquires about 50,000 volumes per year, representing about 35,000 new titles, and has nearly 20,000 active serial records. Over 300,000 catalog cards are produced each year.

Future Plans

It seems that the idea of closing the card catalog and making the machine-readable data base available directly to the library patron is high on everyone's list of priorities, and Northwestern is no exception. Other projects involve the implementation of the accounting and invoice processing modules mentioned earlier. We look forward to a computer-to-computer connection of some sort which will provide access to the MARC file (at least) without the necessity of our maintaining it locally. With the size of our processing load, we see very little economic advantage in transferring it to some remote network computer. Finally, we see the need for additional capabilities for authority control.

Reference

1. James S. Aagaard, "An Interactive Computer-Based Circulation System: Design and Development," *Journal of Library Automation*, vol. 5, n. 1, pp. 3-11; March 1972.

Reprinted from Illinois Libraries

Volume 60, Number 4; April 1978