On-Line Interactive Systems in Libraries, Now and in the Future

For the sake of clarity one can discuss on-line interactive systems with reference to their technology, their economics, and their application and utilization, even though all these aspects are obviously intertwined and inseparable. They will, however, be considered separately.

On-line, interactive systems were conceived long ago, even before computers. Vannevar Bush's Memex was essentially an on-line interactive system, as were the first teaching machines. What is considered the first computer-based interactive systems were implemented by M. I. T. in its SAGE system which responded to RADAR signals. From a practical, economically feasible point of view, however, certain technological developments were necessary before on-line processing could be widely adopted.¹

TECHNOLOGY

The first computers were serial devices with all information stored on tape. Access to a record involved passing a reel of tape and serially searching the records. Although information could be stored on drums and thus provide very rapid access to records, the capacities of drums were so low and drum storage was so expensive that their use was confined almost entirely to program information, tables, etc. The file of records, except for a few exotic and expensive applications, was not stored on drums.

The only practical, economic storage, therefore, was on tape, and processing, for the most part, had to operate in a batch mode. In the
mid-1950s the RAMAC, the first disk file, was announced. With RAMAC it was possible to access a record directly without having to search serially a whole file. With the disk, there is actually a serial search of a track, but this is so short that for all practical purposes one can speak of direct access to a record. With the disk file, the computer in a sense graduated from the age of the scroll to the age of the leaved book or card file.

With the availability of disks, it became feasible for man to communicate directly with the computer. Where previously most communication with the computer was via the punched card or the punched paper tape, now a number of terminals, primarily typewriter and teletype keyboards, were developed or adapted to communicate with the computer. These terminals, which were a great convenience, because now the computer inputs could be prepared where the data originated, were used for remote job entry and to order printouts from the files. Competing devices, primarily key-to-tape recorders and optical scanners were more economical than the terminals however, and so direct remote job entry for information collection did not find very wide acceptance.

The use of on-line terminals has been popular for text editing which was developed in the early 1960s, especially for those applications that required frequent, rapid and radical changes in text. Even here, the development of such machines as the magnetic tape Selectric typewriter (MT/ST) provided a number of text editing capabilities at such a low cost that many users, and especially libraries—including the Library of Congress—found the MT/ST a very satisfactory and economical device to create records and edit text for computer storage.

Although other devices have been less expensive than on-line terminals for remote job entry and for limited text editing, they have been preferred where large amounts of text must be handled or where complicated text editing is required. For example, when an intellectual activity has to be supported such as doing calculations for the engineer, scientist, statistician, accountant, etc., or a proper entry must be chosen to fulfill an order or make a flight reservation, on-line processing is again the most preferred method. It saves much human labor, increases and speeds up the throughput, and, in many instances, improves the precision to such an extent that commercially competitive industries and professions are adopting on-line methods wherever it is advantageous.

With the development of direct access storage devices (DASD), it became feasible to communicate with the computer in real time. The user could interrogate the files to get an instant response. If the amount of information that was transmitted and displayed was not extensive, then the voice grade telephone lines and the 15-characters-per-second typewriter terminals were adequate. Where, however, the number of characters to be displayed exceeded 200 to 300, then the user became impatient since he could read much faster
than the terminal could type. The display terminal, built around the cathode ray tube (CRT), has therefore become very popular; it can display 500 to 1,000 and more characters instantaneously, and is proving to be, in its basic form at least, inexpensive to manufacture and operate, and hence very competitive with the typewriter terminals.

One major objection librarians have had to the display terminal is that the more available, low cost display units have had limited character sets, and have not provided the upper and lower case and special characters that librarians want. Extended character sets are available with display terminals but they are relatively expensive. It actually does not cost much more to manufacture a terminal with an extended character set; the market has not yet justified the large-scale manufacture of such terminals. As the market develops, especially in the so-called media industries—advertising, printing, etc.—large-scale manufacturing will bring down the cost of extended character set terminals, and the librarians will get the terminals they need at a price that they can afford.

There are other aspects associated with terminals of interest to librarians. The so-called intelligent terminals that permit a certain amount of processing by providing access to special local files, and the buffered terminals that will hold working amounts of inputs and outputs, represent special approaches to develop the most economic configurations for a variety of applications. In other words, the terminals that will be used will range from the so-called dumb ones, like the touch tone telephone, to the intelligent ones which are essentially small computers.

The use of terminals initiated the need for communication links. Although many data processing people were apprehensive that the communications industry would become a major bottleneck for data processing, this has not proved to be the case. Attempts are currently being made to reduce communication costs, and there is every indication that they will succeed. Hardware, in fact, is probably the smallest obstacle to the development of interactive library systems.

Even the software problems which affect library automation and also affect batch systems, actually present no real hindrances. The main problem for library automation is the organization and handling of a variety of very large files. In addition, library bibliographic files are somewhat different from the more usual commercial and name (so-called people) files. Library records are long; the average MARC record is of the order of 600 bytes to which has to be added the local library’s data including costs, holdings, etc. There are many records. There are 16 million records in a National Union Catalog, and, the Library of Congress needs a trillion bit file to store such a file. Although such files exist and are being used very successfully, they are slow and hence not well adapted for interactive processing. In a typical situation, it now takes about five seconds—an intolerable length of time for a computer—to get the
first record. Actually this is not very bad since by use of lookahead, overlap, buffering and taking advantage of the slowness of the human, the getting of subsequent records is faster and does not necessarily hold up processing.

The lessened reliance on physical movement and the greater reliance on electronic switching is so speeding up the accessing of magnetic records that fetching a record will not be the bottleneck it has been. There is also the exciting potential of holographic memories, but that is farther in the future. Massive storage devices and digital communications are being improved very rapidly. In an industry which has had a rapid technological growth, these two areas are currently having the most rapid growth and development.

Computer designers are very much concerned with the balance between the various working parts of a system. For example, getting the input/output (I/O) functions in balance with the internal processing can be a problem. Wherever this type of physical bottleneck develops, one can be sure that a great deal of effort will be devoted by manufacturers and data processing personnel to solving the problem. Putting together the physical components of a system is not the job for an amateur. There have been several computer configurations assembled by library systems people on their own which have left a lot to be desired.

With the coming of time sharing, the most difficult problem was what data processing personnel call resource allocation or resource management. This also involves communications control, i.e., controlling the traffic generated by a number of different users, each doing different tasks from different terminals. As a result, operating systems and supervisors or monitors began to be developed. Since different users accessed a number of different files and there was a growing desire to separate application programs from data base management, major efforts were devoted to building independent data management systems. These developments all represent a currently strong movement towards making application programs independent of the host. For library interactive systems, data base/data communications (DB/DC) systems are needed that will work efficiently in a conversational mode and on unformatted, variable length records. Such programs are currently being developed using a number of different approaches; as a result there is some confusion and chaos. Standard, supported programs are however now available and users do not have to compromise or make do with supervisors and monitors that are barely adequate or prohibitively expensive.5

With on-line processing, data base integrity becomes very critical. Where there is updating in place, the files have no old master on which one can fall back in case of failure. The file, therefore, must be protected while processing goes on in core. Input from a terminal cannot be held to the quality standards normally expected from keypunch. There are no verify operations from the terminal and bad input data can damage the data base. If the originator inputs the data and scans it on the display terminal before storing it, the inputs are
usually of very high quality. Nevertheless there is a great need for very good editing and input validation. Data bases also have to be protected from faulty application programs which can occur when existing programs are modified.

For all the above and a number of other security reasons, it is extremely important to have powerful DB/DC systems. The development of such programs is a highly specialized task. One must caution library systems personnel against preparing homemade, special DB/DC programs. This is an expensive waste of time which really does not aid the library, even though it may satisfy the ego of the system designer. Essentially, what I am trying to emphasize is that library people should not try to act as engineers or programmer specialists building type I programs, but should devote their time to application programs and problems that are peculiar to libraries.

One could go on to analyze a number of different aspects of data base management, data communication, and operating systems, but these are really not any impediment to the adoption and utilization of interactive library systems. There are a number of fairly satisfactory software programs for all of these problems and a large number of people are working on improving or replacing these programs with better systems.

File and index organization is much more specialized and pertinent to the success of library programs than the DB/DC programs. Although computer professionals and the computer industry are devoting an appreciable effort to the organizing and accessing of large files, they often overlook or ignore the special problems of library files. Library files are different and are used differently from the more usual data bases found in business, industry and science. Librarians and the library systems personnel must, therefore, devote some thought and effort to the organization of their files.

Librarians are beginning to appreciate the economics of book storage where little-used materials are kept in less accessible but more economical storage facilities. However the cost of retrieving and returning the book from such storage is high compared to that of retrieving the more frequently used materials which are kept in the most expensive storage area close to the library user.

However, the librarian is reluctant to organize all his records similarly, feeling that the information seeker will not understand and hence will be less tolerant of impediments and delays in finding information. Once having found the books he wants, the library patron is more apt to be tolerant about the delivery of the older, more exotic and hence less accessible items. Much of the present effort in the data processing community is to organize files based on usage patterns. Although the librarian may accept this concept for his materials, he is less apt to accept this for his subjects and authors. He therefore has no hesitancy about putting in separate storage certain information such as the data associated with ordering and receiving an item—vendor, cost and the like—which he knows he will hardly ever use after the
transactions are completed. However, subject and author information, even though it may be associated with little-used materials, may be consulted urgently and frequently both by library patrons seeking information and librarians using the records as authorities for processing new acquisitions. Data processing people as a rule do not fully appreciate this and do not really understand why the file organization developed for, let us say, a large insurance company is not ideally suited for a file of library records of a similar size.

The index organization of the more heavily posted areas—the Bible, Shakespeare, the U. S. Government—also present special problems. In the past when using manual files, reliance has been on human discrimination in wading through these interminable indexes. Librarians have ignored this problem partially because they had no way of knowing how people used such indexes. Now with computer systems having statistical modules, librarians will be better able to know how people use the catalogs and other finding tools and hence will be in a better position to do something about catalog organization.

At this point data processing is just beginning to attack the whole problem of subject cataloging and indexing. Most library data processing systems, certainly those in academic and public libraries, are merely accepting the subject organization which has traditionally been used in library catalogs. There are only a very few systems that not only allow for several levels of subject indexing but also for completely different vocabularies. There are many difficulties in accomplishing this, an example is the National Library of Medicine’s problems trying to interface the MeSH headings for pharmaceutical and organic compounds with the much larger and more specific vocabularies used by Chemical Abstracts and the Food and Drug Administration.

Even with the dual vocabularies—detailed descriptors as well as standard library subject headings—that we used in the library management system installed in the library of the IBM Advanced Systems Development Division, Los Gatos Laboratory, and even though full file review is available for all access points and Boolean operators can be used in searching the descriptor indexes, and permutation is applied to titles and corporate authors, still the research personnel are not fully satisfied with the searching capabilities of the system. They want the full panoply of Boolean operators extended not only to full subject headings, but also the masking and permutation capabilities extended to the subject headings so that the individual elements making up a subject heading can be searched like descriptors. Then the system will help them make hierarchical searches, provide automatic cross reference control, etc.

But this is a large area needing specific discussion and is not specifically associated with on-line interactive systems. What is important for interactive systems is that no matter what type of indexing or labeling used and no matter what search strategies employed, there are going to be some heavily
posted items which will be time-consuming to access, no matter how fast the hardware might be. Some thought must therefore be given to the design of accessing or addressing methods to overcome, or at least mitigate, this difficulty. Probably some form of a hierarchical index sequential method as presently applied in one library on-line system will have to be employed.

Of more immediate consequence and promising greater benefits for organizing the main bibliographic records is the file organization used in the Los Gatos Library management system mentioned above. Although all library records of an item were combined into a single logical record, the physical storage and display of the record was segmented into separate physical parts. The system designers decided to establish a hierarchy of physical storage since only a small part of the record—author, title, publisher, date, physical location and availability—was used for finding purposes; and the rest of the stored information, both bibliographic and commercial, was rarely used, and then primarily for processing purposes. In normal searching of the files, only a two-line record was displayed and the normal terminal could display up to four such records simultaneously; if the total record of an item were desired, then the rest of the record was fetched from a lower storage and only a single record was displayed, usually filling the whole screen. At Los Gatos, it is strongly suspected that not only is this more efficient for storage and machine processing, but it increases the possibility that the use of the catalog record as a means of finding information will succeed. The library catalog, after all, is built for two purposes: one as a control of the collection and the other as a finding tool. Too often the control functions complicate or at least obscure the purely identifying and finding elements of the catalog record and hence make the library catalog difficult to use, especially for the public.

Once it is recognized that the usage patterns of library files is different from the more usual files used in other applications, one can make more intelligent decisions about organizing the records. For example, in most situations where a trade-off between processing and storage is possible, it is more economical and more efficient to save processing time at the expense of storage costs. Library usage seems to dictate just the opposite. With only very tiny segments of the file accessed at any one time and with very large parts of the file used extremely infrequently, it is more economical and basically more efficient to conserve storage space as much as possible even at the expense of increased processing costs. This may involve the extensive use of pointers and codes and tables to compress records and avoid redundancies, or the physical segmenting of records to avoid transporting more than is necessary.

Many of the problems described are still anticipatory. Except for the rising quantity of MARC records—which few libraries are really prepared to store—librarians do not yet have very large files, nor do they make heavy demands on them. It is going to be a while before the National Union Catalog is stored in the computer or before all of Chemical Abstracts is available in
digital form. In the meantime, technology is not going to stand still. It seems rather pointless for the librarian to spend time worrying about solving tomorrow's problems with today's tools, especially when he is not sure what tomorrow's problems will be.

This all means that essentially there are no serious technological difficulties which hinder the development of on-line, interactive library systems. There is an awareness that, as interactive systems develop and as people learn how they work and what their potentialities are, systems will grow and change and library processing and library usage will change. Nevertheless there is a good understanding of what needs to be done now to get started, become operational, and to do today's jobs. It means that efforts should be expended to produce application programs and operate them not only to do productive work but also to explore the improvements that can be made. To quote from the report Libraries and Information Technology, of the National Academy of Sciences to the Council on Library Resources, "the primary bar to development of national computer-based library and information systems is no longer basically a technology-feasibility problem. Rather it is the combination of complex institutional and organizational human-related problems and the inadequate economic/value system associated with these activities. National leadership to solve these problems has not emerged."

ECONOMICS

The "inadequate economic/value system" that the National Academy of Sciences considered "the primary bar to development of national computer-based library and information systems" encompasses the totality which they call the "economics of information." The economic justification of information in general and libraries in particular goes far beyond the purview of this article which assumes that data processing has been accepted or at least is being considered for adoption in libraries. However, a number of interactive systems are being looked at and evaluations are being made about the utility and value of on-line interactive systems.

An on-line system requires a communications link, a terminal, storage for its records in a more expensive device than a reel of tape, and controls for the flow of messages which makes it more expensive than an equivalent batch system. The basic question is therefore, "Is it worthwhile to have an on-line system?" The answer is, "Only if the total on-line operation is more economical, and/or if it provides added services such as producing a better product, or doing things more quickly." One should not minimize the convenience factor. People are willing to pay a great deal for things which minimize work, save time and in general increase throughput. But basically, for an on-line system to have wider use in the library, it must be more economical than a batch system.
The hardware and software of an on-line system cost more than a batch system. Ultimately the difference will not be as great as it is at present, since a good on-line system does not require the large and expensive support mechanism that all manual and batch systems require. These include card files, printed lists and directories, guidebooks and other masses of printed paper. A good on-line system can be truly paperless, and the maintenance of paper in the library—that is the working paper and not the bibliographic collection—is an extremely expensive business. It is reported that the Library of Congress maintains over 1,200 files. It is quite conceivable, although this has not been worked out as yet, that the larger displaceable costs of an on-line system will make it competitive with a batch system. Currently, for example, an on-line circulation control system in a large library can be cheaper than a batch system because of the elimination of the massive printouts of circulation lists.

What is incontrovertible is that the on-line system saves labor compared to the usual batch and manual system. At the lowest level, just bringing the needed information to the worker rather than having the worker physically go and get it can be a very important laborsaver. Studies at the Library of Congress have shown that 60 percent of a cataloger’s time is spent walking from his desk to the various files and catalogs, opening the trays or bound volumes, transcribing the information and carrying it back to his desk.8

Repeated copying of information is another wasteful operation. At the National Library of Medicine analysts select and index the journal articles and prepare worksheets for the keypunch operators. These clerks transcribe the worksheets and prepare machinable inputs. When the information flows into the computer it is edited and the detected errors, whether originally committed by the indexer or introduced by the transcription clerk, are detected and the record is recycled, usually all the way back to the indexer.9

Recycling costs are very high. Even with the MARC program and its use of the MT/ST, which has reduced the error factor at the Library of Congress and has speeded up cataloging, the amount of recycling is still very high and most of this is due to error detection in machine editing and proofreading. On the average it takes three inputs to produce two records at the Library of Congress. Where editing and proofreading are done on-line, the original material is still in hand, and the decisions are still fresh in the librarian’s mind, and there is no intervening clerk to introduce additional errors, the amount of labor required is greatly reduced and the time needed for the preparation of the record is shortened.

Similarly an on-line circulation control system can be a laborsaver. Since each transaction in an on-line circulation control system is essentially an inquiry to validate the legitimacy of the transaction and since all the elements involved in the transaction are present—the borrower, the book and the librarian—the snags and conflicts can be readily resolved. In a batch system, the snags show up long after the transaction has been completed and the
loan has been made. To resolve the problem at that time is often a long and tedious business. In one university library, the on-line circulation control system made it possible to reduce the library’s circulation staff from ten to five and even the remaining five had more time to provide other reader services.

There is one area where on-line systems can add to labor costs. A good on-line system forces the librarian to validate variable information. For example, when order librarians order a book, the system forces them to check the author and title even though in their own mind they are positive that the author and title are not in the library’s records. The system also forces them to look at the completed order to be sure that everything is correct. Some are so confident, and some library’s operations are so uncomplicated or at such a low level that such safety controls may be unnecessary and a hindrance to rapid and efficient processing. Some small systems, therefore, dispense with such controls. However, such safety controls are worth their cost, especially in large systems, even though the number of errors caught may be very small.

Since machine costs are decreasing rapidly and consistently and labor costs are rising rapidly and consistently, savings are ultimately going to be made with the system that will save labor even with an added machine cost. In one study of a university library, it was shown that the projected rising labor costs, which included salary rates and increased processing requirements, would provide the necessary displaceable costs for a total library on-line system within five years. The costs of the system were set at today’s costs and today’s efficiencies. There was no consideration of any potential improvement in the system or the lowering of machine costs. In this instance, although there would be added costs for a number of years, the total system costs of the on-line system would in five years be less than the projected current system.

One can speak blithely of future costs, but the credibility of such figures seems to rest mainly not on any empirical evidence but rather on the basic attitudes of the individuals concerned. A few basic points should be made to support the contention that the developing on-line, interactive systems will be cheaper per operation while manual and machine batch systems will be more expensive.

One argument has already been made: that labor costs are rising steadily. In libraries they are rising at the rate of 10 percent per year. Data processing machine costs are dropping. Typical of any new technology that is past recouping the initial development costs, the cost decreases have been spectacular and will undoubtedly slow up in time. The rising operating costs are posing a real threat to libraries, for society is beginning to show signs of unwillingness to support libraries. When properly applied, machine systems can effectively perform many of the library’s tasks and perform them more cheaply.
A few examples will be given of what is meant by "properly applied." There is a learning curve when a new system is installed. Once a new system becomes operational, one learns not only how to improve the system but how to improve one's skill in using it. Typically, in one interactive library system the operators could chain commands and institute default operations. The order librarian wants to examine an author entry to determine the correctness of the order information. The step-by-step procedure would be to first select the function to be performed, in this case, file review. From the file review list the index to be reviewed will be selected. In this case it is the author list. From this list the segment that contains the author's name will be examined. All this would involve a minimum of four steps with three displays. The first three steps can be eliminated along with two of the displays simply by chaining the commands which would bring forth the last two displays. Skillful experienced order librarians can chain a dozen or more commands—each command is a single letter or digit and involves only one keystroke; three or four commands can be routinely chained.

In addition to this command chaining, the operator can accept "by default" prestored information. For example, all the standard data stored in the file of the selected vendor can be accepted. Normal orders are, for example, put on an open or blanket order; ship best way; payment will be made in dollars upon receipt of invoice; and first claims are made in six weeks. None of this need be keyed in with the individual order since it is prerecorded in the vendor file.¹¹ The skill in using command chaining and default options depends on the experience and the ingenuity of the operator. The novice will have to proceed step by step. The experienced, clever librarian will be able to ignore and skip many operational steps and thus produce more with less work.

Systems also become more efficient as they become more integrated. Operating integrated systems should really be considered as experimental since they have not been applied as yet to the total system for which they have been designed. However, based on experience with such an experimental integrated system, certain conclusions are obvious. A single record can be put to a number of uses and produce a variety of products. The information keyed in to order a book is later used without additional keying by the cataloger to prepare the catalog record, and from that is produced the pocket and spine labels and the circulation card. It is this multiple use of a single input which can make the difference between a profitable and an unprofitable system.

Integration and multiple use of single inputs is possible not only with on-line systems but, to a certain degree, with batch systems. Actually, transferring a record from application to application in a batch operation, especially when the file is stored on tape and is accessible only in a serial mode, is so clumsy, expensive and impractical that it is better to recreate the record as needed for each separate application.
The design of the Library Management System in Los Gatos was to be the fourth generation of that library’s system. The third generation, a partially integrated batch system, was characterized by one of the systems people as a series of high speed highway segments connected by bottlenecks. To achieve any real breakthrough in efficiency, the new system had to work in a totally integrated manner; all information had to be stored as a single logical file accessible to everyone, according to his needs. To achieve this mechanically, the file had to be on-line in a direct access storage device and accessible from terminals.

Not only is there economy in the production of records but, in having a single logical record, there is economy in utilizing records. Today in machine batch systems and manual systems, when a patron or librarian needs information about a book, he often has to utilize a number of different tapes or go to a number of different places: the catalog, the book shelves, the circulation file, the on-order file, the in-process file, or a branch or departmental library. This weary traveling from department to department to find various files is a fairly common complaint.

Another requirement for a system to be “properly applied” is that there be enough work for it. There has to be a minimum amount of work—a single threshold if you will—to make the system viable. The threshold depends not only on the complexity of the task, but also on quantity. If an expensive machine is idle most of the time, it would be better to have a human do the work even though he may be less efficient. At least he is more versatile and can use his idle time for a variety of tasks. Too often a system is used for a few, often trivial tasks; machines are then misapplied. It is senseless to use an automobile to travel a few hundred yards to perform a simple errand, although it is often done because of laziness, personal convenience or unthinking habit. It is wrong to build an interactive on-line system just to do simple lookup like identifying an author or an entry. If this lookup is part of a larger more complicated process, then it is entirely valid economically. It is only when lookup becomes a more complicated search that the interactive system comes into its own.

Two growth factors greatly affect the threshold level of a system: (1) the increased growth of the operation itself and (2) the decreased cost of the hardware. Originally data processing—and today interactive systems—was adopted primarily by growing and expanding operations. A stagnant or shrinking operation seldom feels any need for adopting new methods and techniques unless, of course, it becomes evident that the stagnation is due to the present inefficient methods. Also, a growing operation will grow into the economic capacity of a machine installation even though it may actually start below the economic threshold.

As Chief of the Library Branch of the Technical Information Division of the U.S. Atomic Energy Commission (AEC) in Oak Ridge, the quantity of
documents I had to process was increasing rapidly and Civil Service would not give me more personnel as the workload increased. Finally, in desperation, I asked for enough money to install some unit record equipment to do the work. I could not reduce staff and the machines would not be utilized for more than an hour or so a day, but I would never ask for any more people. I made a very rash promise twenty years ago, and today, although the amount of library processing put out by the AEC in Oak Ridge has increased manyfold, they have not had to increase manpower and the machine capacities have been more than able to keep pace with the growth in demand.

As evidence of the decreased cost of the hardware, today, a small computer—and by this I do not mean the little minicomputer but the small general purpose machine that is capable of performing all the basic data processing jobs in the library—rents for the salary equivalent of one to two professional personnel. Seven years ago when library automation first got underway, the least expensive hardware configuration capable of doing the library's work cost ten to fifteen times as much. One should not forget that today's small computer is at least the equal in power, capacity and performance of that much more expensive machine. In addition to these small stand-alone machines, the increased availability of time-sharing facilities is also lowering the economic threshold to where the smaller libraries can afford to make use of data processing. One also should remember the tremendous potential saving that is possible by using shared data bases. Until on-line systems are utilized, such sharing cannot be fully exploited.

Looking at the total operating costs of the library, it appears that data processing, when properly applied, will either reduce costs or slow down the present rapid rise in processing costs. On-line interactive systems will be the preferred method for a number of applications both because of the economics of the situation and because of the inherent needs of the application.

APPLICATIONS

One could spend an appreciable amount of time defining interactive systems based on level of communication. There is simple remote job entry or one-way communication. Typical examples are placing a hold on a book or recommending a new acquisition. At a higher level is two-way communication in a prescribed format. Examples of this might be placing an order with a vendor with a system check for duplicates or checking an authority file for correctness of entry. At the highest level a system might provide fully unstructured conversation.

Based on available descriptions of both library and non-library systems, one can extract those elements of an application which strongly influence a user to select on-line, interactive processing. Some of these elements have been
mentioned already, but they bear repeating. One is the need to transport the information directly to the user so that he will not have to spend time and energy going to the various sources. Librarians have recognized this problem in designing their workrooms and in locating the official catalog, the shelflist, the serials check-in files, etc. It is an accepted idea that library processing people must go to the various catalogs and files. Many workers develop their own little card files, indexes, authority lists and other working tools for convenience, to save time, to avoid walking to distant cabinets and shelves, and to avoid waiting in a queue while someone else is using the file or book. When the information is brought to the individual immediately as needed, and there is never a conflict or out-of-file situation as exists in manual files, then the productivity of the workers is greatly increased. Some have argued that physically going for information is a good distraction and breaks the monotony of the job. But such arguments are mostly rationalizations, certainly in the library.

On the input side, a good on-line system eliminates the intermediary transcriber clerk, usually a keypunch operator or typist. One does not need a separate operation to update the file. The originator of the information—the order librarian, the cataloger, the patron or clerk charging a book or discharging a loan, etc.—in creating the information is immediately creating the record and storing it. No new errors are introduced by a transcription clerk or by a defective work sheet. If any problems develop and if any proofreading or correction is required, this is done immediately by the person who has just created the record and who has all the pertinent information and materials at hand. Furthermore, experience has shown that there are fewer errors when the data is entered by the person familiar with it rather than by a keypunch operator.

Since the system is on-line, the normal editing by the computer can be accomplished immediately and the creator of the record can immediately take the necessary action. This adds up to faster turnaround time, faster and greater throughput. Mention has already been made of the large amount of recycling or looping back which is currently necessary in the MARC program. Figures are not available for the National Library of Medicine Index Medicus program or for other abstracting/indexing services, but all connected with such operations have complained about the burdens of double processing.

This recycling or looping back is due to the fact that the error is detected long after the transaction has been completed. As noted, in an on-line system each transaction is an inquiry which must be answered before the transaction is completed. In a batch system, the inquiry comes after the transaction has been completed and the parties concerned are either not present, as with a book loan, or the material has gone and the person involved is engaged in a new task, as in cataloging and ordering. As a result, more controls have to be built into a batch system than are necessary with an on-line system. In some
libraries the loan clerk may have to verify that the patron is a legitimate borrower and the book may be charged out. The cataloger may check the authority lists for every entry and verify every field, even though subsequently these may be machine-editing functions. The order librarian will carefully fill in all the boxes on the order form even though a large number of them carry default information such as method of payment, method of shipment, claim cycle and the like, simply because she is not absolutely sure what the stored default information is. In a good on-line system the completed order, including the default information which did not have to be manually entered, is displayed for verification and acceptance before final processing.

On-line processing guarantees the currency of records. The moment a record is created and stored, it is available to every inquirer. As soon as a journal issue is checked in, an order recorded, a book cataloged, or a loan discharged, that information is immediately available to everyone. There is no librarian in existence who has not, on many occasions, spent long frustrating hours trying to determine the status of an item. Most of this is due to the fact that the record is not available until long after the event.

Currency of information is of overriding importance in many commercial applications. That is why so many credit and banking applications, insurance processing, airline reservations and the like were pioneers and moved rapidly to on-line systems. Such urgency may not exist in librarianship, but it should if libraries are to have satisfied patrons. Too often the acceptance of delays in making information available and the difficulty in obtaining it reduce the utilization of such information and so reduce the use of the library. Furthermore there is little incentive to maintain such records. A case in point is serials holdings records. When its serials holdings records are first published, it is not unusual for a library to experience a 20 percent increase in the use of periodicals. It is often also startling and amusing to discover how bad the old card file of serial records are and how rapidly they are cleaned up and updated when they become highly visible in published form. In fact, the most difficult problem in building a computerized serials list is cleaning up the old card file. The bad record is usually not due to the fact that the information was not available, but simply because changing the record involved a lot of work and there were higher priority jobs demanding the librarian’s or clerk’s time. So at the most, a note may have been made which might act as a reminder.12

As already mentioned, the ability to obtain current information very quickly in an on-line system does away with the need for maintaining expensive paper files. Yet, since we have always had them, they are given up with great reluctance. When the RAMAC was first introduced and used for inventory control, the information was processed and stored in the machine, but the records that were consulted by the users were printouts. It took some time for people to give up their cherished lists and card files and go directly
to the computer which produced these lists and cards. Too often batch systems are nothing more than a new method for maintaining manual files and sometimes on-line systems are added on top of existing manual systems rather than eliminating the latter. This of course makes for very expensive operations. This is to be expected in new systems because of lack of confidence in the new and familiarity with the old. What is unforgivable is the continuation of the old beyond the minimum conversion period.

An interactive, on-line mode of operation is much more comfortable for the worker than a batch mode because there is no need to break up tasks to fit the batch. If it is better and more convenient, for example, to completely process an item before tackling the next job, even though it requires a variety of tasks, it can be done with the interactive on-line system. The batch system is most efficient when it groups a large number of items for a single task or a small number of operations. This can mean, for example, that the cataloger may have to handle the same book more than once. People find it distracting and confusing to handle many different items in a single batch. There is the problem of broken continuity of thought and the problem of remembering.

One must admit that the efficiency of some catalogers would be improved if they did a little more batching. However, the mere fact that the batch mode is essentially a procrustean bed, inevitably forces the human to accommodate himself to the machine, which can be undesirable. In an interactive system the user can shape the application program to suit his requirements as in command chaining, variable file access, choosing different functions, etc. In essence, the human controls the order and sequence of events, not the machine. The system is said to be user oriented. A batch program, however, is absolutely rigid. Although it may be able to turn out a great variety of individual, customized products, the operational sequences are programmed and unchangeable. A good interactive system should not have any of the dehumanizing aspects that too often characterize our big, industrial society.

At this clinic there is understandably no example of an on-line, interactive reference system. The only ones that exist in libraries are for certain specialties, are experimental or are confined to limited collections. The machine-readable files of large, general collections have not yet been built. These are necessary before truly adequate reference work can be performed. The search and retrieval functions that do exist for general collections are used entirely to support the various technical processing functions in the library.

The major research on interactive, on-line systems involve search and retrieval functions. It seems that this is the main interest of information science; it is certainly the dominant theme in that literature. This clinic restricted itself to the immediate problems of librarians, deferring for possible future consideration all the topics associated with reference work, while
recognizing that much of the original impetus to adopt interactive, on-line systems was to improve search and retrieval. I will, therefore, mention only a single aspect of the search techniques possible with on-line systems since it has a profound effect on library technical processing.

The computer can dynamically organize and structure files. By the use of Boolean operators or multiple access points, it can organize lists and compilations extracted from the data base in many different sequences and with various populations. In a manual system this is not possible since everything is filed in a single sequence. For example, to receive a book, the user must search the outstanding order file. To search the file to answer an inquiry about the item, the user must know the entry which is used to sequence the file. If the original entry is incorrect, which happens frequently, or if the original inquiry does not have all the necessary information, then the user will have difficulty in generating an entry compatible with the file entry. The mark of a good receiving clerk is his ingenuity in solving such snags. A good on-line system provides multiple access points, not only from the individual fields on the order—author, editor, publisher, vendor—but also includes fully permuted titles and corporate authors. The availability of any one of these single clues is sufficient to retrieve the record. In addition, even partial clues using indeterminate search keys such as truncated and compacted terms to find authors, titles, etc., when coupled with the browsing capabilities made possible by the display terminal, greatly improves our searching capabilities beyond anything in the past.

These search strategy capabilities can, in part, exist in a batch system but their utilization is much more restricted and awkward. The user, for example, can submit a query in various different arrangements and then look at the various printouts to see which was the lucky one. Or the user could go through a series of search strategies until he or she is satisfied, but in every case he or she must wait for a completed search before trying a new search strategy. In an on-line system, by seeing what intermediate results are, one can carry out this exploration quickly and efficiently modifying the search as he or she picks up clues, scans lists, etc. There is a proper division between the intellectual processes which the human does and the mechanical processes which the machine does. This is the essence of interactive processing.

Another example is indexing. The present arguments between machine indexing and human indexing should really not take place. Rather, one should examine very closely what aspects of machine indexing can be integrated with human indexing to generate the best index possible. This is a topic which cannot be covered at this clinic. The exploitation of the interplay between what the machine can do and what the human can do is what will really advance the state of library technology. Format recognition as exemplified by the current RECON project will only be really successful in an on-line interactive mode.
Obviously, successful library systems of the future will be hybrid systems, partially manual, partially batch and partially on-line. Data processing people are trying to build the best possible operating system for the library, and they are going to make use of many different technologies and a great variety of systems.

Interactive use of stored communicable and dynamic data bases will undoubtedly have a profound effect on the role of the library in our rapidly changing bibliographic world. For example, on-line systems being developed in business and industry are tending toward the building of large, central, integrated facilities. In the library this will mean everyone having access to large, shared data bases like union catalogs and massive inputs, like MARC outputs, now available only to a very few large libraries. This does not inhibit the development of small, stand-alone systems that perform all the local tasks such as circulation control, periodical check-in, and binding control. Librarians should ask themselves how having what is sometime referred to as a computer utility will affect libraries.

Another subject for speculation is the suspicion that the successful exploitation of retrieval systems and the widespread use of libraries as sources of information will come primarily not because of any improvement in indexing and retrieval methods, but because getting the needed information will be made convenient and easy. When the library terminal becomes as ubiquitous as the telephone, libraries are going to play a much larger role in information processing and transmission.

An even more radical idea is that as interactive systems open the door to new forms of information storage and information transfer, some people are talking about the abolition of all hard copy. Potentially the librarian might have little concern with books, periodicals and other printed matter.

It is tempting to speculate as new tools open up opportunities for new methodologies and services. But in view of the real interests of this clinic’s participants, this presentation has been kept to kept to mundane, immediate and practical topics. Library technology is moving in the direction of on-line, interactive processing and it is important that it be done well.

REFERENCES


8. King, op. cit., p. 80


