# CONTENTS

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction to Data Processing; Introductory Remarks. Edward M. Heiliger</td>
<td>5</td>
</tr>
<tr>
<td>The Hardware of Data Processing. C. D. Gull</td>
<td>6</td>
</tr>
<tr>
<td>Computerized Cataloging: The Computerized Catalog at Florida Atlantic University. Jean Perreault</td>
<td>20</td>
</tr>
<tr>
<td>Automatic Classification and Indexing, for Libraries? Donald V. Black</td>
<td>35</td>
</tr>
<tr>
<td>Computerized Serial Records. Don S. Culbertson</td>
<td>53</td>
</tr>
<tr>
<td>Computerized Circulation Work: A Case Study of the 357 Data Collection System. Ralph E. McCoy</td>
<td>59</td>
</tr>
<tr>
<td>Data Processing Aids in Acquisitions Work. Louis A. Schultheiss</td>
<td>66</td>
</tr>
<tr>
<td>Dissemination of Information. I. A. Warheit</td>
<td>73</td>
</tr>
<tr>
<td>The Library of Congress Project. Gilbert W. King</td>
<td>90</td>
</tr>
<tr>
<td>The MEDLARS Project at the National Library of Medicine. Charles J. Austin</td>
<td>94</td>
</tr>
<tr>
<td>The Machine and the Librarian. Ralph H. Parker</td>
<td>100</td>
</tr>
<tr>
<td>Statement of Types of Classification Available to New Academic Libraries.</td>
<td>104</td>
</tr>
<tr>
<td>Nominees, 1964/65</td>
<td>111</td>
</tr>
<tr>
<td>Second International Study Conference on Classification Research: Conclusions and Recommendations</td>
<td>113</td>
</tr>
<tr>
<td>Variant Pricing of Serial Publications. Peter Gellatly</td>
<td>117</td>
</tr>
<tr>
<td>The Advantages and Disadvantages of a Classified Periodicals Collection. Joseph C. Borden</td>
<td>122</td>
</tr>
</tbody>
</table>
EDITORIAL BOARD

Editor: ......................................................... Esther J. Piercy

Assistant Editors:
Richard M. Dougherty ........ for Acquisition Section
Paul S. Dunkin
for Cataloging and Classification Section
Stephen W. Ford ................. for Serials Section
Allen B. Veaner ............... for Copying Methods Section

Editorial Advisers:
Maurice F. Tauber (for Technical Services)
Doris Ransom (for Regional Groups)

Managing Editor: ......................... Ray O. Hummel, Jr.

Business Assistant:
David Turiel (for Advertising)

Circulation Manager: ...................... Mrs. Elizabeth Rodell

Library Resources and Technical Services, the quarterly official publication of the Resources and Technical Services Division of the American Library Association is published at 2901 Byrdhill Road, Richmond, Va. 23205. Editorial Office: Processing Division, Enoch Pratt Free Library, 400 Cathedral St., Baltimore, Md. 21201. Circulation and Business Office: 50 E. Huron St., Chicago, Ill. 60611. Subscription Price: to members of the ALA Resources and Technical Services Division paying ALA dues of $6.00 or more, $2.00 per year, included in the membership dues; to members paying less than $6.00, and to nonmembers, $5.00 per year, single copies $1.25, orders of five or more copies (same issue or assorted), $1.00 each.

"Second-class postage paid at Richmond, Va., and at additional mailing offices."

LRTS is indexed in Library Literature and in Library Science Abstracts. Its reviews are included in the Book Review Digest.

Editors: Material published in LRTS is not copyrighted. When reprinting the courtesy of citation to the original publication is requested. Publication in LRTS does not imply official endorsement by the Resources and Technical Services Division nor by ALA, and the assumption of editorial responsibility is not to be construed necessarily as endorsement of the opinions expressed by individual contributors.
INTRODUCTION TO DATA PROCESSING*

Introductory Remarks

EDWARD M. HEILIGER
Director of Library and Information Retrieval Services
Florida Atlantic University
Boca Raton, Florida

This meeting is the result of a cooperative effort by the Reference Services Division and the Resources and Technical Services Division of ALA. A committee of members of the two divisions planned this Pre-Conference Institute. These members were: Melvin J. Voigt, Henry J. Dubester, Donald V. Black, Robert Thomas, Jesse H. Shera, Maurice F. Tauber, Robert E. Kingery, and Edward Heiliger. Donald Wright and Elizabeth Rodell of the ALA staff, Ralph H. Parker and Frederick L. Arnold, Jr., also assisted the Committee. The guidelines laid down for the Committee stated that: (1) the Institute should be for general librarians, not specialists; (2) the Institute should be an introduction to data processing; (3) the program should be aimed at librarians from all types of libraries; (4) there should be some talk about how the new hardware is being used; and (5) education for this new approach should be discussed. Except for this last point, this has been adhered to. The Library Education Division is devoting a session to this subject at its meeting in St. Louis on Wednesday morning. The speakers for our Institute understand that their remarks must be directed to those of you who are coming without benefit of acquaintance with his new field. I hope they can overcome their knowledge-ability, and get through to you in a meaningful way. The speakers, with the exception of the two on Saturday, are all professional librarians. This, in itself, is some indication that the library profession is awakening to the possibilities of the new machinery. Some of these librarians have had many years of experience in this new area; others have had less, but all have become involved. All have something to say to you from their experience.

*Editor's note: The ten papers which follow are revised versions of those presented at the ALA Pre-Conference Institute held at the University of Missouri, Columbia, June 24-27, 1964. The Institute was jointly sponsored by the Reference Services and the Resources and Technical Services divisions of ALA in cooperation with the University of Missouri. One talk is missing: Joseph Becker outlined in an informal manner the potential possibilities of the use of automation in libraries. He stressed the fact that machines are becoming more and more flexible and adaptable at the same time that they are shrinking in size, the introduction of transistors effecting the most striking improvements. What he said is generally covered by the book, Information Storage and Retrieval, by him and Robert Hayes (Wiley 1963, $11.95).

Mr. Heiliger acted as Chairman of the Planning Committee and also of the Conference; he has also assisted in editing the papers.

Volume 9, Number 1, Winter 1965
The Hardware of Data Processing*  

C. D. GULL  
Professor of Library Science  
Indiana University, Bloomington

MR. BECKER HAS TOLD YOU that the first thing to start with is the punched card, historically, and perhaps in your operation. Here is a plain punched card (Illustration 1); the little rectangular holes or the circular holes have been punched in with a very accurate punch. Mr. Becker has already spoken to you of the binary system. This card is a representation of the binary system because at any one of those coordinate positions, the intersection of a horizontal or a vertical axis, a punched hole has the equivalent of a “yes” value in the binary system, and the absence of a hole has the equivalent of a “no” value in that system. These coordinates permit you to place against each of the punching positions a variety of values. In this particular slide you will notice that one of the horizontal rows is labeled “six,” for example, and the vertical columns are numbered at the bottom from one to eighty. Each punching position can be identified in more than one way, and you can build this variety up into letters and symbols as well as numerals. There is no text across the top of that card.

By putting a card through one of the devices known as the Interpreter (Illustration 2) it is possible to read the holes and to print across the top of the card the text which is punched into the card. The cards are fed up into the top and come out in the lower pocket, and the cards are then humanly readable. The text on an interpreted card is printed across the top to correspond with punches for the letters of the alphabet, the numerals, and the punctuation marks in the body of the card. The registry of a punched character and its printing involved is not exact, because in the earlier days when interpreters were first put out, they weren’t able to print close enough to have a one-to-one relationship between columns and characters. The primitive stage has been overcome in the 026 printing punch which punches and prints simultaneously. (Illustration 3) The young lady has alphabetic and numeric keyboards under her right hand as well as a number of control switches. She is feeding the cards in from the upper right, they pass across in front of her to the left, and they are stacked in the upper left position. The machine has the capability of duplicating columns from a previous card,

* (Editor’s note: Mr. Gull spoke informally, explaining the 46 slides which he showed and which pictured the equipment and methods of using it. It was with great reluctance that he permitted us to publish the transcription of his remarks since he feels them to be inadequate without the illustrations which we, because of cost and space limitations, could not publish in full. The Editors, however, consider his comments a useful pulling together of information. The illustrations are available in the publications listed at the end of the discussion.)
or, with the aid of a program control card in the center of the machine, can tabulate by columns, duplicate cards, or leave cards unduplicated. The 026 is a much more sophisticated device than a simple typewriter, with much greater flexibility. The 026 printing punch is an input device. It is the station at which your written text, cataloging, order slips, etc., are transformed into binary codes in the form of holes in cards.

The Flexowriter made by Friden is another input device, a typewriter which produces a paper tape with the punched holes in it. Paper tape in one sense is merely a number of cards strung together, or, if you prefer, cards are simply paper tape cut apart. The text, however, is not visible on the paper tape, and consequently you do have a problem in reading and correcting it which you do not encounter with punched cards. Paper tape has certain advantages, and you will want to get your engineers to advise you which of these two input devices you will want to choose in your particular application. Paper tape is the input chosen, for example, for Index Medicus and the MEFLARS Project at the NLM.

The 026 printing punch prints the characters in the same column as the corresponding holes. Each character is made up of a number of little dots which are imprinted on top of the card as it goes through the printing punch. The punched and printed values are the same for numerals, because you chose that value for the row. But when you want to make the alphabet out of this card, you find that the original card has only the capacity for the ten decimal numerals. Two additional horizontal rows, sometimes called the eleven and the twelve punches, were added. By grouping the twelve punch with one to nine, you can build up A through I in the alphabet; grouped, you see, A through I with the twelve punch and a sequence of numbers. There are some other combinations which provide punctuation symbols. The manufacturer has simply taken the various positions on the card and built them up to form his standard alphabet. Other manufacturers use different standard combinations, and mathematicians and others have a lot of fun making many codes out of the various combinations.

The normal feed of a punched card is to put the nine edge (the bottom edge) through a reading machine, thus reading one to eighty characters at a time and correspondingly feeding impulses out to various mechanisms, particularly to a printing device. Feeding on the nine edge normally means that you want multiple character reading from each card. There are a few card readers which will feed from the left edge and proceed from column one to eighty. These are character by character reading or printing devices.

The card movement is from the right to the left, and a metal contact roller below receives successive electric impulses. So long as the paper intervenes, the electrical impulse never reaches the brush; but when it reaches a hole, the impulse goes on through and can be carried, to all intents and purposes, wherever you wish within the equipment. The timing of the impulses is significant. They are intermittent, and they are available only as each of the reading positions passes by, so there are a...
dozen impulses for the movement of one card. You will recall that the hole has the positive value of “yes” attached to it, and the absence of the hole, or the state of being insulated, has the “no” value. The reading of a card can be described as a logical operation. The hole can be called “matching,” and the insulated state is “not matching,” or “rejection.”

The next piece of equipment is a sorter. A deck of cards is fed into the right-hand side of the machine, and the cards move toward the left and drop into pockets according to the punches which are read in a single column on a single pass (or sort) of the deck of cards. The blades direct the cards into the pockets according to the timed impulses as they pass through particular holes. One of the pockets is the reject pocket because there may be no hole in a particular column. There are twelve pockets, all of which are matched to some different positive value, and the thirteenth pocket for the reject situation. Since you can change the columns across the card and read different ones in succession and since you can specify that you may wish to read only certain rows within those columns, you have considerable flexibility in sorting operations. The physical arrangement of this equipment requires the operator to feed cards into one feed and to withdraw the cards from thirteen pockets, then repeat the process many times until the required order is achieved. The operator controls the order of the return of the cards to the feeding station. Ordinarily we are trying to arrange a deck of cards into a certain order. We may wish to have them in numerical order, in which case one pass in a column will accomplish the work, for the number of columns equals the maximum number of digits in the arithmetical field. If we have to sort alphabetically, we have to pass the card through twice on each column.

The two main operations here which the operator has to observe are the “greater than” and the “less than” situations. Thus we add two more operations to our understanding. The operator must be careful of the direction in which he picks up which packets of cards to reintroduce them into the feed. The memory of how to do this procedure often rests in the operator’s mind, or he follows the directions of another person or which he has worked out for himself to solve his particular sorting problem.

Some sorters are much more sophisticated; they can count holes in many columns. They have multiple brushes and little visible counters from which you can take statistical information. They can also note similarities in numbers in different fields on the face of the same card, in the fixed field situation, or in the free field situation, and then direct certain cards into certain pockets. These units are often called statistical sorters, and they are sometimes employed in non-conventional information systems for the purpose of retrieving information, for asking questions and producing answers in terms of the accession number or the call number of a document.

Sorters are used to sequence cards to give a useful arrangement. They are used for the arrangement of cards prior to interfiling them into a catalog, for example. In the operations the sorter does the matching:
that is, the sorter makes a decision, thus relieving the librarian of that job. The operator, however, retains the decision of greater than or less than in the ordering process.

After you have put cards into order, you are in a position to do something with them. We now consider the piece of equipment commonly called the tabulator. There are a number of different models, and they are used for several purposes; but librarians are primarily interested in them as printing devices. One deck feeds into the machine at the top on the nine edge so that all eighty characters are read at once and then they are stacked up at the bottom. All or a portion of each card can be printed at a time. The printing speeds are about 150 lines a minute. Until 1950 this speed was marvelous and we certainly enjoyed the sound of this machinery clanking along rhythmically because the sound meant that we were getting printed output, but today it is a very slow speed. Since these primarily are accounting machines, they will add and subtract numbers; furthermore they will give you minor, intermediate, and major totals which means that they will group certain types of information and then do additions afterward.

There are other devices in the punch card line which will perform multiplication and division, as well, and produce new punched cards. These are various card calculating devices. You may take the new cards and put them back through tabulators for printing out, for example. The arithmetical capability of these printing machines is primarily of use to librarians in order department work, for personnel records, etc.

Illustration 4 shows how tabulators are wired. As the card is being read, an impulse passes through the hole, over to an external area and through a control panel; then the impulse goes back into the tabulator to actuate a solenoid to strike a piece of type on an elevated typebar; the impression will go through the ribbon and put a character on a piece of paper. This slide shows why the machine clanked along, because when all eighty of these typebars are in proper position, the eighty solenoids for the hammers all hit at once, and a whole line is printed at a time.

The plug board is a very interesting part of the equipment. It will permit the operator to break a circuit or to continue it, and may permit him to switch a particular impulse to a new location or perhaps into multiple locations. The basic principle was very difficult to unearth in the punched card literature of the 1930's. It is similar to travelling to the fork of a road, where you can choose either fork and proceed to another fork and repeat the process. The plugboard offers great flexibility in directing an impulse wherever you want it to go.

The real disadvantage of tabulators for librarians is that they offer only a very limited set of characters. Some of them are as limited as thirty-nine characters in the set, that is twenty-six letters, ten numerals and three punctuation marks; others have sixty-four characters in a set. This limited quantity is in upper case characters only; generally, there are no lower case characters and very little variety in symbols. While we think of our alphabet of twenty-six letters as being a restricted one, yet
capable of writing everything, in practice we use much larger alphabets. The character set used by Chemical Abstracts contains nearly a thousand characters. This particular type of printing restriction has been very serious indeed, and is, I think, responsible for the lack of enthusiasm with which librarians view the possibility of printing with the tabulator.

Illustration 5 shows the plugboard which goes with several of these devices. Each one of the little wires can be removed and placed somewhere else. The operator can wire a bewildering variety of programs into a tabulator with a plugboard for routing the impulses. This characteristic of the plugboard must be understood clearly. Each new job means that the plugboard has to be rewired by hand and checked for accuracy. This rewiring is not required with computers, because computer programs are written and placed in the computer’s memory to accomplish the control of the operations.

A little more advanced device for printing, is a print wheel. Instead of having the type bars move up and down, the bar, now wrapped around in a circle, is a wheel which is rotated into position. The type is struck against the ribbon and paper by a small hammer. This arrangement affords a little greater speed and a little larger set of characters.

One deck of cards is often not sufficient, so you need extra decks. The reproducer is the machine by which you accomplish this job. You feed the deck you want copied into one of the feeds and blank cards in the other, and you get out two identical decks. You can operate with both decks, put the second deck in a new order, for example. But this device doesn’t print; it only reproduces. If you want to read what is on the card, you have to take the new deck to the interpreter before you can use them manually and visually.

The most sophisticated device in the punched card line, is the collator. It feeds two decks simultaneously; usually these decks are in the same order, but there are some situations where you may ignore that requirement. Instead of comparing only a column at a time, as the sorter does, these may compare from eight to sixteen columns of alphanumeric information rather than just numeric data. In other words, the collator is a word-comparing machine, whereas the simple sorter is a character-comparing machine. We need to know what can be done with cards on a word-by-word basis. The collator permits us to select or reject cards by the matching operation; this is a form of retrieving information. The collator permits us to compare values by the greater than or less than operations, as the equal or unequal comparison. It enables us to accomplish a sequence check; we can put some cards into a collator and find out if they are in the right order and check them for ascending and descending order before using them further.

All of us have had to interfile cards; the ones we take off the sorter, for example, which were in alphabetic order, have to be interfiled somewhere with another deck. The collator permits us to do our filing or merging, and this can be done on the basis of a common number or on the lack of a common number. The collator can also be used for re-
trieval operations, by putting two decks in, each of which represents a single subject heading, for example, and looking for common numbers in other fields of those decks. If a common number is found, the corresponding cards are selected out and they are presumed to contain the answer to the retrieval question. This matching is the logical operation called “and,” logical product, or logical conjunction. “And” is one of the new operations that you are to remember, then.

If two decks are merged by the collator, in the absence of numbers or by ignoring the numbers, making one deck out of two, the operation is called logical “or,” logical alternation or logical sum, or as some prefer to call it, disjunction. The important point about these operations is that they are controlled by the equipment on the card-by-card level. The operator makes the choice of logical operation at the group level, but the individual decisions are made by the machine on the individual punched cards, and you work only with the results. This description is equivalent to saying, “I will go to a subject heading in the dictionary catalog for retrieval purposes and pull out five hundred cards on that subject. Since I want to consider those cards which are related to a second subject, I have to look over all five hundred cards one by one; but with punched cards and a collator, I can turn that job of scanning over to the collator, and it will select all those cards which show the relationship between the two subjects.” The collator has relieved the human of this particular decision problem. Now the operations AND and OR can also be accomplished with sorters, but humans control the operations with those devices.

At this stage in the punched card art, which was roughly in the 1940’s, all of the ideas for computers were available in physical form. There were input, processing, and output of data. The punched card was a form of memory, in the data recorded thereon. The control procedures were largely external, and the humans had to establish them. Therefore the memory of these procedures rested in the human mind or in manuals, with all their deficiencies. Because operators had to feed cards and take them out of pockets and put them back in again, human intervention was very frequent. Every situation which requires human intervention introduces high error rates. Our technology has gone beyond this stage and so we must explore the advantages computers offer.

It required some very ingenious people, von Neumann and others, to conceive the idea of putting the instructions, the procedural control, and successive amounts of data into an internal semi-permanent but yet erasable memory within a computer. The data establish the conditions which exist, and these conditions then control the subsequent operations which are performed on the data. You can express it very simply this way: if X exists, then do Y; if X does not exist, do Z, a different operation. Now, today, the external memory is kept principally in magnetic tape form, although you can use punched cards and paper tapes for this purpose if you wish. The power of computers, which we are coming to, is derived from their logical capabilities, MATCH, AND, OR, and
NOT, GREATER THAN and LESS THAN, and from the number of commands which you can give to the computer in its machine language.

Mr. Becker already mentioned the number of parallel operations which can be accomplished in computers, the amount of two-way communication which exists among the components of computers, the number and variety of peripheral devices which can be hitched on to computers. He mentioned their operating speeds, and then finally the question of how much software is available from each manufacturer for his particular computer. We can’t generalize about the seventy models of computers that are currently available in this country; I'm not going to try to do so. The important point about computers is that human intervention is very much reduced for the amount of work accomplished.

Magnetic tape codes can be developed on magnetic tape so that you can see them. The magnetic tape is the same kind of tape in principle that you use for tape recorders at home; a mylar ribbon with finely divided iron oxide on one surface. There are a number of tape codes and attempts at standardization in spite of the proprietary interests involved here. There can be 200, 556, or 800 characters to the inch, so that the packing densities of these tapes is much greater than anything on punched cards.

Illustration 6 shows a 1401 computer, a character by character machine. The manual controls and the indicator lights are shown here on the panel, and the cabinet contains the memory, the central processor, and the power supply. Since the memory here is relatively small, the computer requires small programs and can process only small amounts of data at a time, because the memory holds both the program and the data in the same physical location. The control consoles are different on the models available in this country, but in general they have start and stop buttons, switches to establish certain conditions that are required for operations, and lights to show conditions and errors. There will be buttons by which to override the errors and try the system again, that is, to force something through the system.

Information must be fed into the computer just as into the punched card devices which were its predecessors. The 1402 card reader punch is an input-output device. The stacker holds about 3000 cards and they are fed in rapidly to introduce information into the central processor of the computer. It can also be used as an output device by routing the impulses out of the computer to punch cards, with which you can operate punch card devices or use the cards manually if you interpret them.

Two kinds of things can be put in here, the programs to control the operations, and the data; you can also take out data and modified programs. The production of modified programs is a significant capability of all computers. There are programming routines known as editors and compilers which enable you to accomplish a pretty sophisticated job of changing your programs. The final program is sometimes pretty far removed from the work of the original programmer after it has been edited, compiled, and assembled on the computer; it hardly will recognize its intellectual parent, if you want to put it that way.
Illustration 7 is a picture of the magnetic cores and Mr. Becker described them so fully to you, that I’ll only say that you test the state of the core with a diagonal sense wire and read that state out and do something with the answer you found for that particular position. A number of positions taken together give you a character or word. These cores are rather tiny, about the size of a sequin, and many thousands of them are used in each computer memory. They are assembled into a core plane, and the planes are assembled into blocks. The result is a core memory in block form, with thousands of wires leading out from it. For easy understanding you can consider that part of the memory containing the program is static during the running of an operation, but the remainder is very alive; it is handling the data very rapidly. The active part processes data at greater speeds than does the human brain.

The punched card devices are, in effect, so far as sorters and collators are concerned, extensions of human hands and arms for the manipulation of $3 \times 5$ cards, but they are so slow that some other equipment had to be developed to overcome this slowness. The magnetic tape handlers (Illustration 8) are one such development. A full reel of tape is put on one side of a unit and a take-up reel on the other, and the tape is moved forward and backward under the control of the computer program. The tapes contain programs and data. Usually tape handlers are used in groups, standing side by side. Not only do these units surpass humans in sorting things out and interfilining, they also read and write at the same time. As the tape moves, the magnetic reading head takes the impulses and puts them somewhere in the computer. After the impulses are processed to the computer’s satisfaction, they are read onto another tape. The magnetic tapes are erasable and can be used thousands of times. As to tape speeds, the tapes may move 75 to 125 inches per second; the units are remarkable instruments for acceleration and deceleration; they start a tape from scratch and stop it very rapidly. The read-write speeds exceed anything the human can undertake; they range from 20,000 to 100,000 characters per second. Although this high speed is precisely why magnetic tape is used as an input-output medium to a computer, modern computers are still largely tape-bound for most operations. The machine processes are retarded by the read-write tape speeds, because the internal processing is so much faster.

The data and program information flow to and from the memory, buffers, printers, punches, processors, and tape handlers. There is continuous flow of impulses within the computer, but the physical motion is intermittent. There is a considerable flow of impulses to and from them for every physical motion observed. Information is transferred from one tape to another so that extracting or interfilining means writing on a fresh, reusable tape. It is not necessary to cut a tape apart, add a strip, and seal the parts together to interfile something. Information is taken off one tape and written on another one to expand (or contract) a tape.

Some device is required for printing out the results. One such is a 1403 printer. The impulses are obtained by reading a magnetic tape.
the central processor arranges the characters according to specifications and sends them to the printer. The 1403 uses a print chain; it looks something like the timing chain in an automobile, but with type in place of teeth. There are 132 printing positions across the paper for each line. The print chain is divided into five sections of forty-eight characters each, and it is in constant motion during printing. The printing speed is about 1100 lines a minute for numerical information and about 600 lines a minute for alphanumerical information. In operation these printers sound like a hail storm, because each character is struck individually in the very brief period when a line is available and when each character arrives in the proper position. The 132 characters are struck at different instants for a line, instead of together in one blow. At these speeds the paper has to be fed from endless, perforated, folded piles.

This print chain has twenty-six capital letters, ten numerals, and twelve symbols. At the insistence of documentalists, a few chains have been made with upper and lower case characters. The effect of this change, against a fixed rotating speed and a fixed number of positions on a chain, is to reduce the printing speed, because the character wanted doesn’t come into printing position as frequently as it does in five sets of 48 characters each. The price of typographic excellence is to reduce the printing speed by about 50%. Chains can be changed in about two minutes, however. There are other types of computer printers; some of them print line by line; some use cathode ray tubes and xerography to print on the paper at rates as high as 5000 lines per minute.

The most significant technical development of the Index Medicus or MEDLARS project was GRACE, Graphic Arts Composing Equipment, or the Photon 900 as it is called commercially. It was designed to provide a greater set of characters, 226 characters in the set, at 330 characters per second exposed to negative film. The GRACE type of printing will be used for the August issue of Index Medicus, because the machine is now in the National Library of Medicine for its acceptance trials. The entire July issue is being run through as part of the acceptance trials. It’s turning out very nice copy, upper and lower case characters, bold face italics, and Greek symbols, etc.

Paper tape is one form of input. Some computers can provide a paper tape output as well as using a paper tape input, and this output can be used to actuate typewriters and some other printing devices.

The 1401 is one of the simple small computers, a character by character machine. Most of the larger computers operate with computer words rather than with characters. Computer words are not quite English dictionary words, but they have a larger number of bits than characters. The word size may range from 20 to 42 bits, and the words are handled effectively as units in the machine. These other computer models are larger, in performance and physical size, than the 1401. This condition is not a contradiction of what Mr. Becker said about computers growing smaller and smaller in recent years. In general, the physical size of modern computers varies directly with capacity, but even the largest
computers today are smaller than the earliest large computers. Some of the larger computers can handle as many as 66 tape handlers at once. It would take a regiment of librarians to compete with these computers. The memories are larger too. The common sizes of magnetic core storage for these word devices are 8,000, 16,000, 32,000, 64,000, and 128,000 word memories, for example. There are also thin film memories.

We haven't mentioned the problem of order of records which is a really difficult problem in librarianship. All of these magnetic tape devices have a linear or sequential pattern on the tape, reading from one end of a scroll to the other, reading forward or backing up if you wish. Some people don't like linear scanning because of the amount of time consumed or the amount of processing required. They would rather have a computer analog of the dictionary catalog, or the inverted file which is broken up into a discrete order. With this order they can go to a physical location and find a piece of information.

Magnetic disc storage is made, in effect, of constantly rotating phonograph records with magnetic characters on the top and bottom of each disc. There is an access arm which can move vertically up and down the stack of disks and horizontally in toward the center of each of those discs. It has a read-write head. If you tell it that all the information on a certain problem is to be found in a certain location, it will shift to that location and read that information into the central processor. This action is called "random access" and for librarians' purposes that's a very poor choice of words. Each location is known exactly, just as surely as you know your house address. Your house address may not have any particular relationship to any sequence of numbers and words, but you know where it is, and the postman knows where it is. Random access is a direct addressing form of storing data. Random access provides a way of getting at what you want without going through all the tapes. It has other advantages as well. A similar device with much less capacity is the magnetic drum in which the magnetic characters are put on the surface of the drum, and the drum rotates at constant speed.

There are additional peripheral devices which offer a variety of remote input and output stations which can be connected to computers by direct wire or by radio or you can send tapes and cards through the mail.

**PUBLICATIONS FROM WHICH SLIDES WERE TAKEN**


*Volume 9, Number 1, Winter 1965*
Wiring for Printing

The Internal and External Wiring to Cause Printing

Figure 4

407 Plugboard

Figure 5
1401 Printer

Figure 6
Magnetic Cores

Figure 7
Tape Handlers

Figure 8
STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION
(Act of October 23, 1962; Section 4369, Title 39, United States Code)

1. DATE OF FILING
   September 21, 1964

2. TITLE OF PUBLICATION
   Library Resources and Technical Services

3. FREQUENCY OF ISSUE
   Quarterly

4. LOCATION OF KNOWN OFFICE OF PUBLICATION (Street, city, county, state, zip code)
   2901 Byrdhill Road, Richmond, Virginia 23225

5. LOCATION OF THE HEADQUARTERS OR GENERAL BUSINESS OFFICES OF THE
   PUBLISHERS (Not printers)
   50 East Huron St., Chicago, Illinois 60611

6. NAMES AND ADDRESSES OF PUBLISHER, EDITOR, AND MANAGING EDITOR
   PUBLISHER (Name and address)
   American Library Association, 50 East Huron St., Chicago, Illinois 60611
   EDITOR (Name and address)
   Miss Esther J. Piercy, Enoch Pratt Free Library, Baltimore, Maryland 21201
   MANAGING EDITOR (Name and address)
   Ray O. Hummel, Jr., Virginia State Library, Richmond, Va.

7. OWNER (If owned by a corporation, its name and address must be stated and also immediately
   thereunder the names and addresses of stockholders owning or holding 1 percent or more of
   total amount of stock. If not owned by a corporation, the names and addresses of the individual
   owners must be given. If owned by a partnership or other unincorporated firm, its name and
   address, as well as that of each individual must be given.)

   NAME
   American Library Association
   (No stockholders—non-profit organization)

   ADDRESS
   50 East Huron St., Chicago, Illinois 60611

8. KNOWN BONDHOLDERS, MORTGAGEES, AND OTHER SECURITY HOLDERS
   OWNING OR HOLDING 1 PERCENT OR MORE OF TOTAL AMOUNT OF BONDS,
   MORTGAGES OR OTHER SECURITIES (If there are none, so state)

   NAME
   None

   ADDRESS
   

10. THIS ITEM MUST BE COMPLETED FOR ALL PUBLICATIONS EXCEPT THOSE
    WHICH DO NOT CARRY ADVERTISING OTHER THAN THE PUBLISHER'S OWN
    AND WHICH ARE NAMED IN SECTIONS 132.231, 132.232, AND 132.233, POSTAL
    MANUAL (Sections 4355a, 4355b, and 4356 of Title 39, United States Code)

<table>
<thead>
<tr>
<th>A. TOTAL NO. COPIES PRINTED (Net Press Run)</th>
<th>Average No. Copies Each Issue During Preceding 12 Months</th>
<th>Single Issue Nearest To Filing Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8,604</td>
<td>8,295</td>
</tr>
<tr>
<td>B. PAID CIRCULATION</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. TO TERM SUBSCRIBERS BY MAIL, CARRIER DELIVERY OR BY OTHER MEANS.</td>
<td>7,066</td>
<td>7,147</td>
</tr>
<tr>
<td>2. SALES THROUGH AGENTS, NEWS DEALERS, OR OTHERWISE.</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>C. FREE DISTRIBUTION (including samples) BY MAIL, CARRIER DELIVERY, OR BY OTHER MEANS.</td>
<td>57</td>
<td>76</td>
</tr>
<tr>
<td>D. TOTAL NO. OF COPIES DISTRIBUTED (Sum of lines B1, B2 and C)</td>
<td>7,123</td>
<td>7,233</td>
</tr>
</tbody>
</table>

(Signature of editor, publisher, business manager, or owner)

I certify that the Statements made by me above are correct and complete.

(Signed) LeRoy J. Gaertner,
Comptroller

Volume 9, Number 1, Winter 1965
THE ORIGIN of any event or process is not a simple thing, even though we can view the concatenation of forces as simple when we are concerned with the result by itself; but such a view, of an effect with no consideration of the causes, would be no explanation at all. These causes, though they all have simultaneous effect in our resultant procedures, must be examined separately for a clear insight into the essential newness of our attempt at the Library of Florida Atlantic University, together with a realization of the continuity of our solution with that resulting from the traditional theories of the catalog.

Therefore, before I launch into a full-scale description of our procedures, there are three introductory aspects to be examined, plus a couple of digressions.

Introduction (a)

The first basic introductory aspect I want to examine is that of the image, “the bibliographic string.” The image implies seriality of elements, a seriality imposed and not necessarily already contained in the data to be cataloged. (This seriality might be defined as: The making explicit of what is implicit in the data.) The “string” must be “strung out” rather than left in its tangled original state: order must be imposed, and this order is that of the traditional theory of descriptive cataloging. All cataloging, as it has been developed from the late Renaissance up to the most current practices, has been based on some string-concept or other, that is, on some procrustean order imposed on the data.

And this basic order has had as its natural concomitant the articulation that determined the seriality of the order being imposed. That is, for there to be order there must be parts contiguous to parts; and for there to be seriality there must be a pre-determined before and after. The parts of the bibliographic string can thus be conveniently conceptualized as “1, 2, 3, ... n.”

Each of these articulations can, in terms of the original figure, be visualized as a “knot” in the string, as a signal identifiable because of a particular kind of indentation, type-face, color, punctuation, or associated number. In other words, each such “knot” bears a particular function,
and these functions are generically and specifically differentiated, and are recognized without their differentiation being in every case made perfectly explicit. We can think, for instance, of non-filing initial words; our attempt has been to render each such function entirely explicit.

Introduction (b)

The second basic introductory aspect I want to recount is the historical one. The proximate origin of our ideas and procedures has been the investigations carried out at the Chicago campus of the University of Illinois. Among the principal protagonists in that project were C. D. Gull, Don S. Culbertson, Louis A. Schultheiss, and Edward Heiliger. The Chicago campus investigation’s foremost outcome was the publication of *Advanced Data Processing in the University Library*, the main point of which (over-simplified, of course) was that computerization of the catalog (and, in the broader view, of the library in large part) was both possible and (possibly) economical.

Then, during consultations between the F A U (Florida Atlantic University) staff and Louis Schultheiss, several modifications were arrived at in the provisional design worked out at the Chicago campus. In particular, it became evident that economy of space was a crucial consideration, and that the serialities traditionally generated in a bibliographic string may be more complex than necessary for their proper functioning. Thereafter, our whole effort was expended to make the knots in the string both unique and unambiguous, yet to let their position on the string be flexible enough (a mixed metaphor?) to conserve all the space that was thought possible. Some of these modes of flexibility will be more precisely explored later.

Further consultation, this time with Fred Kilgour of the Yale Medical School Library, resulted (a) in further modifications of our provisional design, in order to make our system as compatible as practicable with that of the Columbia-Harvard-Yale Medical School Library project; and (b) in a large measure of agreement on the limited number of characters that can, economically and without intellectual loss, produce the catalog we have come to expect for a respectable research collection. (The print chain implied in all of this will also be more fully described later).

Ruling over our historical origins, thus, has been cooperation. We hope that this guiding concept will become that of the computerization of libraries all over the country.

Introduction (c)

The third basic aspect I want to examine is the theoretic one. We may say that one of the basic questions upon which the computerized catalog (like every one) is based is: How can the intellectual decisions the cataloger makes become embodied in a catalog? But then we must first ask: What is a catalog, as against an inventory? or as against a bibliography? Let us take a (perhaps illegitimate) shortcut here and not too closely
examine these different kinds of lists (for they all are lists), as worthy of question as these others may be. Instead, I would offer as a definition (that is to say: a setting up of limits) that a catalog is a list, from the internal contents of each item within which there are generated other (quasi-) items. In less abstract (= more library-oriented) terms I mean: There is more than one access-vista to the same document in a catalog.

The catalog thus creates a record of the document in order to “enter” the document in the catalog, but this document is to be entered in the catalog (and the user can enter into it through the catalog) by a possible variety of entries other than the first one. Thus is logically generated the main entry and the various secondary entries. All of these entries must not only be embodied in their appropriate places in the catalog, but must be traceable back to their decision-source: the tracings.

In the traditional cataloging operation, the official catalog is the record of the decisions of the cataloger: this record, before it is filed for possible later reference, is turned into main and secondary entries by typists, and these embodiments of the cataloger’s decisions are placed in order and inserted into the already serialized file, by clerks.

But, as noted before, the embodiments that arise from the official catalog do so implicitly, except where there are explicit tracings. And even in such case, proof-reading of the typed cards is no guarantee at all that the embodiments are equally correctly typed, or have been filed correctly.

Correctness means conformity to rule; and the most logical means to guarantee strict correctness is to generate all the secondary entries at the same time that the main entry is generated, rather than all the others from the one (as with traditional official catalog generating the public catalog). In the traditional catalog, however correct the results may be, in a sense they can be said to be “accidentally” correct. Taken in the precise sense, accidental means that which does not invariably follow from the nature of the substance; the “substance” in this case is the cataloger’s decisions. And it is thus clear that a system which could “naturally” generate its results with full explicitness, a system in which every quasi-item was generated by “chromosomes” imbedded in its own substance (the knots in the bibliographic string)—would be by that much preferable. (This could be true of an un-computerized system as well, but I doubt that it could be accomplished outside the computer.) This very built-in explicitness is what the FAU system aims to achieve (as do those of the Columbia-Harvard-Yale Medical Schools, Ontario New Libraries Project, and others).

**Digression 1**

Let us digress momentarily. Even in a computerized system, something akin to the “collective consciousness” (whose presupposition and occasional lapse result in either correct or incorrect filing in the traditional library) is attempted, for instance in setting up a table of non-filing initial articles. Such a word is the German “Der” which can be a truly non-filing
article (for instance, Der Mann . . .) or a filing article in the genitive (for instance, Der Männer . . .); or the Dutch "De," which, though non-filing, can easily be confused with the Latin preposition "De" which does file.

This sort of ambiguity, the source of filing errors in traditional filing, could equally become a cause of error in a system relying on tables that must be used to look up decisions, since in the computer there can be no discrimination between identicals. It seems better to us to carry explicitness all the way, to indicate every time, automatically (by "programming the cataloger") that here "Der" is non-filing, there it is not. The means by which this is accomplished will be examined later.

Digression 2

Another digression: the lecture on the I B M 357 Data-Collection circulation system will show that because a process in its traditional shape is easier to diagram, we cannot immediately conclude that the more-complex-to-diagram automated system is therefore really more difficult to handle or to operate. The reason is, once more, that the traditional system bears along with it all too much implicitness.

Description of the System: Input

And now, to begin to understand our actual cataloging process at F A U, let us take the sample Catalog-Input Record and fill it in.

Shown here in Figure "a" is an imaginary title-page. The author is of course not to be expected to be given in full or in the correct spelling, but by the aid of reference sources, we determine the correct full form of his name, and print it into Area 10 of Figure "b."* We have encountered the first instance of the normally-tangled state of the bibliographic string, upon which we must impose our own order.

The next instance of a tangle to overcome is the title, which (we shall assume) conflicts with several other editions of the same work; so we determine its original form from reference sources and print it into Area 22, surrounding it with brackets and indicating language and version. The publisher's title is then printed into Area 23, excluding extraneous elements, and then the imprint and collation are printed into Area 31, in pre-determined rather than given order. Into the imprint area is also printed the series note, in the form in which it has been established rather than that given.

The applicable subject headings are printed into Area 70. Then, in Area 76, we trace the name of the person responsible for the literary shape of the work. There is no need here for a title-area tracing (as will be explained later), so we print the call-number into Area 80, and then the location symbol and modified Luhn-number, which takes the first four letters of the main entry, three significant letters from the publisher's

* Editor's note: Mr. Perreault presented many detailed figures illustrating the various steps in preparing information for the computerized catalog; unfortunately we could not publish them all, but anyone really converting to this form of cataloging can probably secure copies by writing him directly.
title, the letter that stands for the month, and the day in that month that
the book is either ordered or cataloged.

The expanded-collation code is now printed into the columns at the
right end of the form. The main entry is personal, so we put a check in 81.
The work, a reprint, was originally published at Leipzig, in 1891, by the
publisher Breitkopf und Härtel, so we look up the code numbers in the
book we have produced for that purpose and print into 81:6—19 the
code numbers 51, 20, 267, and 140000. Then we print the original copy-
right date into 81:24—27, check 81:80, 81, and 33 since the book is a
translation, a critical edition, and a reprint. We also check 81:38 because
our examination of the book shows a listing of editions and translations
of the author’s works. The book is a reprint by our definitions, so we go
on to column 82 and print in the code number for the reprinting pub-
lisher (Musica Press) and the date of the reprinting.

Now we examine the body of the book, noticing that it has an index,
a glossary, and a facsimile of a page of the original issue. So we check
82:35, 36, and 28, and since the facsimile is in the form of a plate not in-
cluded in the pagination, 82:17. It also includes a recording of the organ
music of the author in a pocket in the back, for which we check 85:26 and
trace (in Area 76) the performer.

Description of the System: Output

Now that you have become somewhat more familiar with the mode of
input of the catalog data, let us examine some of the rules, purposes, and
results of this data as it is keypunched, transformed into a magnetic-tape
record, and finally printed out in book form. (The discourse will be made
more meaningful to the reader if he refers to the copy of the Input Rec-
ord Form—Figure b.)

The main entry, whatever its type, is printed into Area 10. If it is a
personal main entry, only the first two lines may be used (76 spaces),
and the appropriate check is made in Area 81 (space 1). If the main
entry is any of the other types, one of the other spaces in Area 81 (spaces
2—5) is checked, and the whole Area 10 may be filled in.

If there are elements in the main entry that would cause misfiling,
they are blocked out of the sort-tag by symbols used for that purpose
alone; the symbol “less-than” (\<) means “What follows, up to the filing
symbol (if one follows), is to be excluded from the sort-tag”; the symbol
“more-than” (\>) means “The sort-tag begins here (if there were a non-
filing symbol at the beginning of the area) or resumes (if the preceding
non-filing symbol were later than the beginning of the area).”

The third and fourth lines of Area 10 have the same number of spaces
as do the two lines of Area 22. This latter is for the conventional title,
and can be used only when the main entry is personal. (The computer is
informed of this by the check in 81:1.) Thus, the total number of spaces
available for main entry plus conventional title (when present) is 148.

The title of the work, when it is not superseded by a title main entry,
is printed into Area 25, using the same non-filing symbols when necessary, and indicating the end of the title-to-be-traced by a non-filing symbol. Considerable pruning in this Area is sometimes necessary to make the operation of the computer economical and rapid. Only the essential elements are retained, and these we construe to include: (a) the actual title, (b) the alternative or sub-title when necessary to explain the actual title, (c) the person or organization responsible for the work's bibliographic form, such as editor or translator if this has not been stated earlier (for instance in the main entry or as part of the actual title), and (d) the number of the edition.

The imprint, collation, and certain essential notes are printed into Area 31. The most common acceptable note is that for a series, if it is to be traced (the process of decision whether or not to trace will be examined later). If the series note is to be traced in title form, punctuation that would make it look like an author-title series is removed and the note is preceded by an equal-sign; if it is one which cannot be treated as a title-form tracing, and accordingly must go to the main entry catalog, it is preceded by a per-cent-sign.

The point of all this is that no further tracing effort is necessary beyond printing the note into its proper place and adding the proper tracing symbol. (Note that again the non-filing symbols are often called for.) Other permissible notes are for the original title of a translation, change of serial title, and the like.

Tracings are printed into Area 70, 76, and 78. Since our catalog is to be divided (Author, Subject, and Title) rather than in the dictionary form, we do not adhere to the L C tripartition into Arabic numbers, Roman numerals, and parentheses; instead we retain all Arabic-number tracings in Area 70, but divide Roman-numeral tracings between Areas 76 and 78, and, when series must be traced in one of these Areas rather than in Area 31 with the functional symbol, they too must be divided between Areas 76 and 78.

The call number is printed into Area 80, as is the symbol for location (when appropriate), and the modified Luhn-number which serves to tie together the deck of key-punch cards.

Into Areas 81, 82 and 85 are printed data which in traditional cataloging might have been part of the collation. (The retrieval aspects of these bodies of information will be discussed later.)

When the bibliographic string is available in a form closely akin to our own decision-pattern, the only indications necessary to make the data computer-assimilable are the "knots." This outside-available form of information is the L C printed card; all that is needed to make this information equally intelligible to the keypuncher is to associate the proper functional area-number with each section of printed data, besidesexcising those elements which would not be allowable under our policies of retention of essentials only, or would overflow the allowable total length of any Area. (See Figure c)

Volume 9, Number 1, Winter 1965
When the Library of Congress has cataloged the work and the printed card itself cannot be supplied, a xerographic copy of the appropriate entry in the National Union Catalog is made through a special mask. Then the same process is performed as when a card can be supplied.

Digression 3: The Print Chain

To digress a moment from the process, the basic theory behind the construction of a catalog as such is that it is to provide multiple access-vistas to one document among a great many; it is thus (at least primarily) a finding tool. Yet what is being found is bibliographic data, and its organization is along the line of principles devised for solving bibliographical problems. There are thus poles between which we might well vacillate, or which might generate tension. These poles are (on the one hand) machine economy in its widest sense, and (on the other) “bibliographical integrity” (a phrase of Fred Kilgour’s). For our computerization to be a success it must be economical, it must not cost more per unit of work done than the traditional process does; and its general form must be happily compatible with the capabilities of computers. For our computerization to be a success it must also satisfy all of the bibliographic/finding-tool requirements put upon the traditional catalog; we must keep faith with the development of the tradition of the catalog, from the late Renaissance up to yesterday.

Part of our commitment to this tradition is the shape and function of the large elements of the catalog; but another (and equally crucial) part is the shape and function of its smallest elements, the letters and symbols available for the embodiment of the intellectual decisions of the cataloger, through the agency of which the user can apprehend the content of the collection and the individuality of the works entered in it.

The letters and symbols available must be various enough to embody all of the material which a large research collection must contain, that is, a great many languages other than English. They must therefore be far more than the 48 characters normally supplied by I B M. (Indeed, these 48 are not even adequate to catalog an entirely English-language research collection.) The normal I B M print chain includes no lower-case letters; present are only the period, comma, slash, hyphen, equal-sign, apostrophe, parentheses, and plus-sign as punctuation; no diacritical symbols are present; numbers (zero—nine) are of course supplied. We have assembled a new 88-character print chain from the various I B M catalogs, and with this design have achieved a measure of agreement with the Medical School Libraries of Columbia, Harvard, and Yale (a union-catalog project), and with Ontario University. These institutions, planning for collections of different character from ours, have so far added six characters to the original 88; besides, there are present on the chain (which has 120 spaces for symbols) a variety of symbols for the use of the programmers and for scientific-computation use of the computer.

The additions to the 48-character chain therefore are: lower-case
Description of the System: Retrieval Aspects (Expanded-Collation Code)

To return to our processes: The information printed into Areas 81, 82, and 85 is not actually part of our conception of a book-catalog, but rather is intended as a means of document retrieval. The basic concept was arrived at during our consultation with Louis Schultheiss and went something like this: When there are generically similar but specifically differentiated elements none of which can occur simultaneously, they may legitimately be input under a single functional number with a modifying number used to show the differentiation. In particular, the main entry can be either personal, corporate, uniform, anonymous-classic, or title; each of these could have been allocated a separately-coded area, but a great deal of space would have been invariably left vacant, since only one such area could be used for any particular work. Again, in the area for imprint, etc., there could be four solutions: (a) a separate field could be left for each element (place, publisher, date, pagination, other collational items, series note, etc.) with at least an occasional instance of truncation of one or more of these internal elements; (b) an open-field situation as against the fixed-field arrangement just outlined, with one maximum for the whole area, into which each item is printed, with its own sub-area number; (c) an open-field situation with an accompanying field of yes/no checks to indicate the characteristics of the internal elements; and (d) translation of the total verbal content of the area into numeric codes, to be re-translated into verbal output upon print-instruction.

Whatever the advantages or disadvantages of the other possible solutions, we at F A U have chosen solution (c), and the accompanying field of yes/no checks is what we have expanded into our retrieval-collation code. In it are recorded the decisions as to type of main entry; as to presence of a large variety of collational elements for all sorts of learning resources (not books alone), for descriptions of the work in terms of its literary origin and provenance; and for encoding the numerical codes for imprint retrieval.

With this device we can retrieve the relatively few documents which have a large variety of desired characteristics many of which would not be available through even the most thoroughgoing traditional cataloging. For instance, we could extract a list (of either call-numbers or the whole catalog entry) of those works bearing not just one subject-heading but a large number, for instance "A" or "B" plus "C" or "D" or "E" plus "F", besides specifying that we want only those published in a particular city or cities, country or countries, between certain dates, including an index and a bibliography that goes beyond (foot-) notes, and certain types of illustrative materials. Except for the subject-heading search, even one such criterion is all too much for a traditional library; we propose to

Volume 9, Number 1, Winter 1965
perform such searches on a more or less routine basis (though to do so economically we must use batch-processing or random-access devices).

Work Flow

Let us now follow the outlines of the work flow. The card accompanying the book, or the xerographic copy of the N U C entry, or the original-cataloging transmittal sheet, is compared to the author and title authority files (to be explained later) to see if any new decisions or alterations are necessary. The cataloger then connects functional numbers to the appropriate data (we assume the presence of the L C card) and fills in the collation code. The books are sent to the shelf while clerks complete the operation by filling in the numerical codes for city, country, and publisher, and prepare subject-authority references whenever the need for new ones arises.

The input record thus prepared is keypunched; a deck of punched cards thus arises, from which a magnetic-tape record is created. This magnetic-tape record is not, as compared with those input just before and after it, in any kind of order; the order of input to the master tape is entirely random. From this random-order tape we will create the three tapes from which the three parts of the catalog will be printed. When the master tape has been created from the punched cards, it is read through the computer onto the three “write” tapes, each to contain the data selected from the master tape for the particular catalog function. These functions are made known to the computer by the functional area-numbers and other symbols, such as the series-tracing equal-or-per-cent-signs.

For example, an entry with a personal main entry has (as do all entries) an Area 10, and when the master tape is read, this functional number means to the computer: “Take Area 10, followed (if present) by Area 22, and then by Areas 23, 31, and 80, to the author tape.” If there is an Area 22 (which could occur only with a personal main entry, since the computer would not even bother to read Area 22 if the main entry were of any other type), the computer gets the message: “Take Areas 22, 10, 23, 31, and 80 to the title tape.” If Area 22 is present, it is normal for there to be no need for tracing the publisher’s title (Area 23) since a conventional title by definition calls for see-reference from each publisher’s-title variant in the collection. But if this Area 22 were one which could not become part of a conventional-title see-reference structure (for instance, Works, dramatic. Selections. English. Smith), it might be necessary to trace the publisher’s title. This is done by printing simply “23” into Area 78. With no Area 22 present, the data printed into Area 23 generates the write instruction “Take Areas 23 (up to the non-initial non-filing symbol), 10, 23, 31, and 80 to the title tape.”

Area 31 is then read and written, but does not normally generate tracings; but if there is present a series note which has been established as traceable, the appropriate symbol placed before it will cause tracing. If the series note to be traced is in title form (which we prefer, as we do
for serial main entries), it is preceded by the equal-sign, which gives the write instruction, “Take the series note, followed by Areas 10, 22, 23, 31 (without the note being traced), and 80 to the title tape.” If the series note to be traced is unavoidably in author-title form, it is preceded by the per-cent. sign, which gives the write instruction, “Take the series note (broken into its author and title components), followed by Areas 10, 22, 23, 31 (without the note being traced), and 80 to the author tape.” If the series note is in author-title form with the series-author indicated by his, her, its, or their, it is again preceded by the per-cent. sign, which gives the write instruction, “Take Area 10, followed by the series note less the initial underlined word, and Areas 22, 23, 31 (without the series note being traced), and 80 to the author tape.”

Area 70 is then read through; the separate subject-headings traced there are each ended with a record-mark (except the last), and each generates a separate write instruction, “Take this segment of Area 70, followed by Areas 10, 22, 23, 31, and 80, to the subject tape.” If the work is autobiographical, the number “10” in Area 70 as a separate segment generates a write instruction to use the data in Area 10 twice, first as subject and then as author.

The manipulations performed on the data in Areas 76 and 78 is similar to that done in Area 70, except that the destination of the secondary entry is in each case a different tape. Shorthand like the use of “10” in Area 70 is used here whenever possible to make operation and input equally economical.

**Filing Rules**

The three tapes from which to print the catalog have been thus generated as automatically as practicable from the one block of “official” input data. Each tape is now in raw order, and must be put into alphabetical order before printing. This brings up the problem of filing rules.

As mentioned before, we consider a divided rather than a dictionary catalog appropriate because of the predictable difficulty involved in programming a body of rules as complex as those for library filing. Secondly, we have committed ourselves wholeheartedly to the idea that filing is to be interpreted as a function of the symbols whereby the data is communicated, rather than by general principles of organization which bear no relation to the symbolization which is the basis for any collation sequence.

Thus, in the traditional library catalog, not only are all of the Saints named John in a group before the Popes named John, followed by all the kings similarly named (all of which contravenes the idea of strictly alphabetical filing), but among the kings the organization is not by the symbols following the name, but by the country governed. This is possible only by the elevation of general principles of hierarchies of ranks above the principle of collation order determined by symbols.

Our order of collation, on the contrary, is set up thus: double-blank, single-blank, A through Z, zero through nine. Double-blanks are gener-
ated by a period, comma, parenthesis, or bracket standing next to a single-
blank. Thus “London, Jack” would come before “London during the
great fire,” which in turn would come before “London’s historic houses.”
A collation sequence which ignored all punctuation would instead file
these three entries “London during the great fire,” “London, Jack,” and
“London’s historic houses.” Our order of titular names will be strictly in
accordance with the symbols used, so that “John, duke of Gaunt” will
come before “John, King of England,” and “Charles II, King of England”
would come (probably) directly before “Charles II, King of France,”
with both Charles I’s preceding them, rather than the traditional anti-
symbolic order.

**Work Flow: Summary 1**

As a first summary of the flow of the material, observe the outline
(figure d) showing that the data is input by a cataloger onto Input-
Record Forms, keypunched into decks of cards, read onto a random-order
master tape, distributed onto the three catalog-production tapes by the
functional numbers, sorted one at a time, and printed on the I B M 1403
printer with its new 88-character chain. It is then reduced by a camera to
a size easier to use than the huge full-size I B M sheets, and again printed
(in multiple copies) on an offset press.

**Control Documents**

But into this complicated but straightforward process must be intro-
duced the controls necessary to lead the user from subject to subject,
from variant author to established author, and from variant titles to
conventional title. These controls must also keep the cataloger from
accidentally using any of the variant forms that are supposed to be re-
placed by established forms.

There have been established, therefore, four types of control docu-
ments.

Type 1 is used to refer from variant to established subject (x-refer-
ences and their inversions, see-references), and from one subject to
related subjects (sa-references). The information input here generates
references only in the actual book-catalog, since we have in the L C Sub-
ject Guide a record of our decisions.

Type 2 is used to refer from variant authors to established ones, from
variant title-form series to established author-title-form series, from title-
position anonymous classics to author-position ones, and of course from
author to author as sa-references. These references are printed out in the
book-catalog but also generate a booklet for the cataloger listing all
variants under their established forms, plus each of the variants referring
to the established forms.

Type 3 is used to refer from variant titles to established ones, from
variant author-title-form series to established title-form series, and from
title to title as sa-references. These references are like Type 2 in that they
print out both in the book-catalog as see-references and form an authority file for the cataloger containing sa-, x-, and see-references.

Type 4 is used for reference from variant titles of a single work to a conventionalized form. It also carries standard subject-headings for that work, which generate blanket references in the subject catalog to the appropriate places in the author and title catalogs where all the editions will be found.

Each of these types of control documents (except Type 1) thus serves a double function: they create a computer-assimilable record of a decision for the use of the cataloger, and a guide for the user to the right place to find what he is after—all automatically from the very document on which the cataloger records his decision.

Work Flow: Summary 2 (with Controls)

It can be seen, then, that before the computer-print step in our earlier flow of materials can be made, we must merge in these separately-produced control tapes. The author-catalog printing then results from the sort of parallel flow shown in figure e. There is, of course, a similar operation in parallel for the production of each of the other tapes (figure f).

Conclusion

Moving from detail to a broader view, we find, first, a hope that by these means we can radically release a number of typists and filing-clerks from humdrum and consequently error-laden tasks, laying these drudgeries instead upon a machine that can produce entirely faithful embodiments of our decisions and file them by inflexible programs rather than vague remembrance of principles, without ever tiring or relaxing. Second, we retain all that the traditional catalog has to offer except unlimited space. We must conserve space for the computer to operate really efficiently; but we have actually added information we could not hope to find in the traditional catalog and have added this information in a way that makes possible a use of the computer to search the tape-stored catalog in a way that can be a great and fruitful aid to reference and research (rather than saving information just for its own sake, never to be accessible in any truly helpful way).

Traditionally, what we want in a library catalog is accurate information handily arranged and pleasantly intelligible, capable of flexibility to harmonize with each sort of document, and able to be kept almost instantaneously current. The card-catalog does these things, sometimes, or at least some of them. The book-catalog does them, too, and opens the way to new and challengingly promising functions. And (we sincerely believe) it can even beat the card-catalog on its own field, in almost all of the areas of comparison: accuracy, handiness, intelligibility, flexibility, currency.

Volume 9, Number 1, Winter 1965
Automatic Classification and Indexing, for Libraries?

DONALD V. BLACK
Consultant on Automation and Information Science, The Library University of California, Los Angeles

THE TITLE of this paper ends with a question mark to indicate that I am not suggesting that automatic classification and indexing is a fait accompli. Probably I will raise more questions than I answer.

The first thing that any speaker ought to do is to be sure that his audience understands the title of his talk. There is a certain amount of imprecision in my title today, and I am going to exploit this by making up a definition to suit my purposes. I plan to use the two terms “classification” and “indexing” somewhat interchangeably, contrary to normal library practice. For one (“classification”) means the placing of written works, of whatever nature, into subject classes, whether or not these subject classes have any systematic order, i.e., generic-specific relationships. By “indexing” I mean the assignment of subject terms to any kind of written material. Notice the difference. “Classification” is concerned with subject headings, and “indexing” is concerned with subject terms. What I intend to do, then, is to use the word “classification” in dealing with typical library-type materials. These would be conventional books and other printed materials, which would normally go into a library of the traditional type. “Indexing” I intend to use in connection with the type of printed materials we now tend to call “technical report literature,” or other such written communications, which are not so widely held in conventional libraries and for which subject headings in the usual, traditional library practice are very rarely used. Now, in my opinion, the methods which I intend to discuss are the same for both automatic classification and automatic indexing. Thus, it is of relatively small importance whether at one moment I talk about classification or about indexing, since the processes which I am going to discuss are equally applicable to the one or to the other, in my opinion.

At this point, it is perhaps useful that I should explain my personal philosophy of librarianship. There seem to be, currently, two camps

* Opinions stated herein are those of the author and do not necessarily represent those of the UCLA Library Administration. Mr. Black is now Head, Technical Services, University of California, Santa Cruz.

Volume 9, Number 1, Winter 1965 • 35 •
among librarians. One is perhaps best characterized as being completely bedazzled by the glittering chrome on the computing machinery: the flashing lights, the whirling tapes, the stream of punched cards flowing in and out, etc. The other camp is frequently characterized as hiding behind a wall built of moldy, old tomes, where their councils meet to decry the advent of mechanization and to pass from hand to hand a favorite volume, so that all may feel the pages, admire the binding, smell the ink on the page. From time to time, one of their number will mount a stack of books to look over the battlements, to view the advancing column of computers, and to shout imprecations against those who seem to be guiding the oncoming machines. Librarians seem to regard the computer as either an instrument for their oppression or as one for their liberation. Now, I have never cared for any system of logic which allows for only two states. Thus, I hope that I am representative of some third school, or middle ground, or what you will, which can see the computer—or any device for the mechanization of processes—as a liberating instrument on one hand, as well as an instrument for oppression on the other, and can, hopefully, maximize the liberation and minimize any oppression. This is a tall order for anyone, and I am not sure that I am man enough to do anything to further this somewhat pious hope. The situation, though, is often not as bad as it seems. What may seem to be oppression may be, in reality, liberation, if we are but honest with ourselves.

Now, to the actual discussion at hand. Probably the first hint of a system of automatic classification or indexing was contained in an article by H. P. Luhn in the *IBM Journal of Research and Development*, October 1957. This was entitled “A Statistical Approach to Mechanized Encoding and Searching of Literary Information.” Luhn followed this germinal article by many more over the years in one place or another; however, his basic technique of using statistical methods did not change. He is responsible, however, for the modern day technique of permutation indexing, although it was a librarian, Crestadoro, who first suggested this back in the 19th century. Evidently the basic idea of Key-Word-In-Context, as the permutation indexing systems have become known, occurred to a number of individuals at about the same time. At least one of them, Herbert Ohlman, then of the System Development Corporation, was aware of Crestadoro’s initial idea behind the permutation indexing technique. Evidently Luhn, however, was not aware of any predecessor in this area, and to him and the IBM Corporation must go full credit for developing the KWIC system into what it has become today.

Phyllis Baxendale, also of IBM, suggested some techniques which could produce an automatic index of journal articles, provided that the complete text were available for machine input. These differed from Luhn’s techniques in being based on the prepositional phase rather than a statistical count of the frequency of words used.

Both of these systems, developed by IBM personnel, can be categorized as, basically, operating on the materials at hand, without a large
body of computer-stored information to aid them. The Key-Word-In-Context System and also Baxendale’s “phrase system” both operate using a relatively small store of words in computer memory. In the KWIC system, the only words which must be stored are those which are deemed non-significant and which will not be used as key words. In the Baxendale system, the only words which need to be stored are the prepositions and all words which are to be excluded from further consideration.

Obviously, what I am saying here is an over-simplification, and I trust that the experts will not take me to task for neglecting to mention some of the finer points. However, it does not seem necessary, in this sort of introductory discussion, to go into these details. Furthermore, I am personally of the opinion that the Key-Word-In-Context approach has been developed about as far as it can go, and that while it can, indeed, be useful in many library functions (which I will not discuss further here) we can look to it for no more new or startling developments, and its future exploitation will bring us no nearer our far-distant goal of perfect control over printed materials.

The next point in our history which I want to discuss is the article by Don R. Swanson in the October 21, 1960 issue of Science which was entitled “Searching Natural Language Text by Computer.” This article reported on work which began considerably before the date of publication and was carried on for some time thereafter. Before proceeding, I want to make it clear that I consider myself a disciple of Swanson. I worked on this early project with Dr. Swanson and others, and very recently I have tried to implement a system which is based upon his ideas.

For those of you who are not familiar with this early experiment, perhaps it would be worthwhile briefly to recapitulate it, since it does form the basis for what I believe to be a workable system of automatic classification and/or indexing. This early experiment by Swanson pitted human indexing and retrieval against machine retrieval, using material in nuclear physics as an experimental library. Note that there was, initially, no special attempt to produce indexing or classification by means of the computer in this first experiment. The computer was to have available to it the full text of the experimental library. Thus, it was felt that no indexing, as such, was necessary.

A group of individuals, all of whom held Ph.D. degrees in nuclear physics, was assigned a number of issues of the journal Physical Review which was to form the experimental library. There were 96 articles in the field of low-energy nuclear physics which were used as the primary experimental body of materials. Each physicist produced a series of questions which were made up from the articles which he was assigned to read. These were examination type questions. I will quote one or two of them here for illustration. For example, “What is the best available value for the electric quadrupole moment of the deuteron?” Another: “What nuclear reactions are sensitive to the spin and parity of mesons, and hence are useful in measuring those quantities?” One last example: “Are
bevatron neutron beams monoenergetic?" The articles which elicited these questions were termed "source articles."

The entire group of physicists then examined the experimental library with all the questions in hand, and decided which articles had pertinence to each question. This pertinence was rated on a scale of one to ten. (There is a technical difference between pertinence and relevance, which it does not seem necessary to observe here. Suffice it to say that sometimes an item may be relevant but not pertinent. Ingenious readers can construct their own examples.) In the first experiment, a completely separate group of nuclear physics articles, taken from Physical Review and some other physics journals, was examined, and on the basis of this a system of subject headings was created. This was not a classed system but followed (more or less) Cutter's rules for alphabetic subject headings.\(^4\)

It must be emphasized that the questions which were to be asked of the system as a test of retrieval effectiveness were not used in preparing the initial subject heading list. In this respect, the experiment certainly follows normal library practice. That is, the catalogers have no idea who is going to use the products of their efforts, how these patrons will approach the catalog, or in what way they will express their needs and wants. There is some question, of course, whether or not these exam-type questions would ever be asked of an information system in the real world. Personally, I do not think so, although there may be a small percentage of the total list of questions which could occur in a real-life situation.

The reasoning behind this experiment was that if the computer had the complete text available to it, no classification or indexing would be necessary for the documents; and that the computer, having the full text available to it for search, could retrieve material more successfully than a human who would have only a catalog to search. As you know, this hypothesis was tested, and under the conditions of the experiment it was found to be true. The human retrievers found only a fraction of the material which the computer was able to retrieve. Interestingly enough, it was also discovered that neither system, either separately or jointly, retrieved every document known to be pertinent to a given test question. The group of people who actually performed the retrieval experiments was non-homogeneous. There were mathematicians, physicists, computer personnel, and librarians. None of these people had seen the test questions prior to performing the retrieval experiments; they were not informed of the results until after their work had been done.

At the end of the first experiments, the results looked something like this: For the conventional subject heading search, of the maximum percentage of source documents (those documents which elicited the questions in the first place) the maximum percentage retrieved was 86%. The best computer search retrieved 84% of the source documents. For each question, the average number of relevant documents known to be in the experimental library was 6.9, and the average number retrieved in the conventional search was 1.2, while the number of irrelevant docu-
ments retrieved averaged 2.6. (See Table I.) Thus, for approximately every one relevant document that was retrieved, two irrelevant documents were retrieved in this conventional subject heading search. (N. B. All figures exclude the source documents, except where they are specifically mentioned.) Results of this first experiment were reported in the above mentioned article but no widely-available report, at least in the library literature, was made on the second phase of the studies.21

One of the criticisms of the first study was that the conventional subject headings had been made up without any idea of even the type of question that might be asked. Therefore, in the second phase the entire list of questions was reviewed, and new subject headings were created. Then the entire experimental library was re-indexed using the new list

**TABLE I**

RETRIEVAL RESULTS ON NUCLEAR PHYSICS LITERATURE
A COMPARISON OF HUMAN SEARCH VS. COMPUTER SEARCH VS. TITLE SEARCH

| R= Number of relevant documents retrieved |
| I= Number of irrelevant documents retrieved |
| SD= Source Documents retrieved |
| C= Conventional search, Phase I |
| C'= Conventional search, Phase II |
| A= Computer search aided by Thesaurus |
| T= Title search |
| NLS= Computer search with cutoff to retrieve 100% of Source Documents |

Note: all results are averages per question

*Volume 9, Number 1, Winter 1965*
of subject headings. There was considerable time span involved in these activities, so that probably there was very little, if any, carry-over from one experiment to another. Also, the indexer was not a physicist, and thus little carry-over normally would be found.

On the basis of the second experiment, it was discovered that the review of the materials had improved the retrieval effectiveness in that an average of 1.3 relevant documents—as opposed to 1.5 irrelevant documents—was found for each question; and the number of source documents retrieved improved also to 54% as opposed to 36% in the first experiment. The magnitude of the improvement is really most noted in the fact that the number of irrelevant documents per relevant document was reduced by 50%, approximately, although the actual number of relevant documents retrieved was not increased in any significant degree. By allowing more irrelevant retrieval, the retrieval effectiveness of the conventional search was improved up to a maximum percentage of 56% of the relevant information known to be available in the library, and 88% of the source documents. At the same time, however, the percentage of irrelevant retrieval went up from 54% to 77%. For the computer to retrieve 100% of the source documents, it was necessary that 62% irrelevant material be retrieved, and an overall total of only 53% of relevant information known to be in the library was obtained. As you see, even the computer wasn't too successful!

At a subsequent period, the titles of the articles in the test "library" were reviewed, and retrieval effectiveness of the titles alone was tested against the questions. It was found that the titles alone retrieved 93% of the source documents. However, since these source documents had suggested the questions to begin with, the language of the title may well have had some bearing on the language used to form the question. Thus, source document retrieval was not considered as indicative of the effectiveness of titles in this respect. However, the titles alone did produce 33% relevant retrieval* (non-source documents) at a penalty of only 50% irrelevant retrieval.** We will have more to say about the titles presently.

The members of Group Two, mentioned above, performed retrieval experiments on the Experimental Library. All the members of this group, except the librarian, constructed search questions which were presented to the computer. The first experiment depended solely on the knowledge of physics of the men involved and in their knowledge of the manipulation of retrieval questions within the computer.

Using the full text of the Experimental Library, a "Thesaurus" was constructed by the computer which sought to link together synonyms, near synonyms, word associations, etc., which might be of aid to the retrievers in their experiment. This was produced in printout form on the computer and was available to each individual for the second retrieval

* That is, 33% of the known relevant documents were retrieved.
** That is, 50% of the documents retrieved were judged irrelevant.
experiment. These individuals were not aware of their successes or failures in the first experiment.

Rather than give any more details along these lines, I will only add that every effort was made to insure the experiment's being as objective as possible, and anything which could conceivably influence the results was firmly controlled.

After the second computer search, which was aided by the Thesaurus, and as part of the second phase of the overall experiment, a detailed examination was made of the results of the two experiments. That is all of the initial experiments, both by humans directly and by use of the computer, were studied to discover, if possible, the reasons that retrieval effectiveness was, in general, so poor. Despite the fact that all of the individuals who were involved in the computer retrieval experiment were competent and held advanced degrees (some in nuclear physics), it was discovered that the search instructions which were formulated for the computer exhibited elementary oversights and very few exhibited any great degree of ingenuity or insight. It was decided that search instructions could probably be formulated as well by the computer itself; at least the computer would exhibit greater consistency than had the human beings. We will return to this point later.

Some other interesting facts emerged from the first phase experiments. One was that if words occurring in an article were actually within a sentence span of three or four sentences of one another, they were much more likely to be related conceptually (within the context of the article) than if their proximity were greater than this three or four sentence span. The proponents of coordinate indexing techniques have used mere co-occurrence in many of their systems for retrieval; and co-occurrence is the weakest form of syntactic coupling. Such devices as "role indicators" or "links" have been developed to strengthen the coupling between terms which are to be searched for retrieval on a coordinate basis, but coordinate indexing still loses a great deal of information because it lacks the ability to indicate syntactical specification. There have been recent experiments which seem to indicate that "roles" and "links" do not really provide any useful measure of syntactical specification. However, proximity does seem, in some cases at any rate, to provide a practical substitute for syntactical specification.

Another interesting fact, which seemed to be indicated from these experiments, was that the frequency of occurrence of a natural language term within a document of whatever kind is not necessarily related to its relevance, or lack thereof, for storage or retrieval. This assertion has not been as strongly indicated in some later experiments as it was in the experiment involving nuclear physics literature. Yet it would seem to cast some doubt on automatic indexing techniques which use frequency of occurrence as their basis for the decision that a given term is representative of the content of a document—that a term is a "keyword" because of its frequency of occurrence.

Volume 9, Number 1, Winter 1965...
We must now move along to consider the present state of human classification and indexing. You will recall, I hope, that I mentioned above that at one point in phase two of the Swanson experiment, titles alone were used to retrieve documents to answer the examination questions of the experiment. Titles alone retrieved 93% of the source documents but only 33% of non-source documents known to be relevant to any given question. Now, a number of people have been interested in the use of titles as indicators of subject content. As far back as 1960, a group at IBM ran an experiment comparing four types of lexical indicators of content. An article describing those experiments at IBM appeared in the journal Human Factors in August 1960, under the title of "A Re-Evaluation of Machine Generated Abstracts." I should like to quote the summary to that article: "Twenty-five subjects were divided into five groups matched for their reading speed and scores on a sample criterion test. Each group was given a different kind of lexical indicator of content for a set of seventy-five documents, namely: titles, three types of abstracts, and complete texts. In addition, they were given an examination composed of seventy-two short-answer type questions, which were derived from fifteen out of the seventy-five documents. They attempted to answer the questions and evaluate the relevance of each document in answering the examination." This experiment seemed to indicate that the task of determining whether documents are relevant or not to some given purpose can be performed by any one of the four types of lexical indicators with about equal results.

A subsequent experiment was reported by Resnick in Science for October 6, 1961, under the title "Relative Effectiveness of Document Titles and Abstracts for Determining Relevance of Documents." This experiment reported a test undertaken as part of a system of Selective Dissemination of Information, wherein individuals were asked to determine the relevance of documents to their interests on the basis of titles, and on the basis of abstracts. The results of this test seemed to indicate that there was no significant difference between the usefulness of titles and the usefulness of abstracts for such a purpose.

Some question may be raised about both of these tests. The numbers of people involved were small: in the first case, only twenty-five individuals and only seventy-two articles; in the second case, 400 documents and 400 individuals were used in the test. However, the 400 documents were separated into two groups of 200 each. I do not believe the numbers involved in the first test, since they concern human activity, are really significant, statistically speaking, despite all the statistical formulae to the contrary. Of course, the same criticism can be levied against the Swanson experiments involving nuclear physics articles. The primary objection to the second experiment revolves around the fact that the notices contained the author's name in all cases, and, quite likely, many individuals selected items of interest on the basis of authorship rather than anything else. Subsequently, Rath et al. devised another experiment to overcome
this difficulty.\textsuperscript{22} It seemed to show, also, that titles were just about as effective in determining content as abstracts or even full text.

Maizell, in 1960, published an article entitled “Value of Titles for Indexing Purposes.”\textsuperscript{11} He found that the titles contained about 58-70\% of the terms under which they were indexed.

Even though these experiments may all have flaws, and I do not say that they do, they provide us with indications of trends which may be further studied and verified.

In October 1962, an article entitled “Machine-Like Indexing by People” was published in American Documentation.\textsuperscript{12} This report, by Montgomery and Swanson, studied what was actually being done in a system which used human indexers. Let me emphasize what Swanson and Montgomery themselves emphasized in their introduction but which was largely overlooked by some librarians. The study was not concerned with how a given number of journal articles could have been indexed, or should have been indexed, but only with the end-product, namely a printed index. Nor was the study concerned with what the individuals actually did in the preparation of their output, but only with what was produced!

This experiment did not involve evaluation of retrieval effectiveness of any system, but rather it was based on the question, and I quote, “To what extent can the human indexing operations that take place in an existing system be simulated by machines?” Thus, the question of whether or not a given system is effective for retrieval is by-passed. Briefly stated, the study examined the September 1960 issue of Index Medicus. The system of subject headings which Index Medicus uses is similar to the subject headings used in regular library practice in that it assigns journal articles, in this case, to pre-established subject categories, just as most libraries assign books to pre-established subject headings. The study was undertaken to see how many of the assignments could have been made from an examination of the titles alone. None of the results were intended to show, nor was it stated, that the people that do the indexing for Index Medicus actually examined only the titles. Some 4,770 entries were studied. Matches could occur, or were considered to occur, in two ways: one, the exact equivalent words were to be found in a title, or two, some synonym of the subject heading was found in the title. (This is a great oversimplification; however the complete details are readily available in reference 12.) In a machine procedure, of course, such a table of synonyms could be created and stored in machine memory, so that the searching process would be automatic. Matches were found in 4,093 entries. There were 1,499 doubtful, and the remaining 528 did not contain words which conceivably could be identified by machine procedures as being synonymous in any way with the subject headings.

The results of this exercise suggested that it would be interesting to examine some special bibliographies which had been compiled in the UCLA Bio-Medical Library, in response to requests for information.
sample of 83 special bibliographies was examined. Once again, the examination was not to determine the actual relevance of the articles listed in each bibliography, or how they were actually produced, but to determine if they could have been produced by a computer from an examination of the titles alone, using some sort of a machine dictionary or thesaurus as a subject authority list, cross-reference guide, and synonym finder. Of the total of 3,145 bibliographic citations evaluated, only 5.8% did not contain at least one term which corresponded to the terms or concepts expressed by the requester in his question.

O'Connor recently published in *American Documentation*, April 1964, the results of a study of the correlation of index headings and title words in three medical indexing systems. O'Connor had raised some doubts concerning the validity of the Montgomery-Swanson study at the annual meeting of the American Documentation Institute in November 1961. In the interim period, he had been studying the problem of correlation of headings and title words, and this recent article represents his current viewpoint on the matter.

I will not go into the details of this study, except to indicate that the three systems studied by O'Connor were: (1) volume 2 of the *Index-Handbook of Cardiovascular Agents*; (2) documents indexed in 1955 for the Merck, Sharp and Dohme punched-card retrieval system; and (3) the 1961 National Institutes of Health Research Grants Index. Not a very great number of subject heading-title pairs were used by O'Connor. He recognizes the limitations inherent in such a small sample, but he feels that the small sample was large enough to permit some generalization.

He found that for the first system there was a 92% positive relation, for the second system there was a 54% positive relation, and for the final system there was a 26% positive relation. This is not especially encouraging if one hopes to use titles as a means of automatic indexing or abstracting of printed materials. However, O'Connor does seem to make some mistakes; at least, it would appear to be the case from the samples which were printed in *American Documentation*. I will mention only one. Under the heading Nephritis the title “The Role of the Kidney in Protein Metabolism” is not considered to have any correlation with the heading. However, the word Nephritis comes from the Greek word meaning kidney, and it would seem that a set of rules could be constructed which would tie the two words together. There are a number of other instances which appear to me to indicate the possibility of automatic processing, were a machine thesaurus of reasonable complexity and quality available to the computer.

Another interesting piece of work contrasting human concept indexing with machine indexing is reported in the April 1964 issue of *American Documentation*. M. J. Ruhl, under the title “Chemical Documents and Their Titles: Human Concept Indexing versus KWIC-Machine Indexing,” reports a study which compared the indexing of identical documents appearing in *Chemical Titles* and in the *Chemical Abstracts*...
Subject Index. *(Chemical Titles* is a publication made up solely of a key-word-in-context index prepared by machine, plus author index.) This study showed that more than half of the titles included all concepts, or their equivalents, as indexed by *Chemical Abstracts*. This is a particularly interesting and important study, since for many years the *Chemical Abstracts* Subject Index has been the standard against which others were compared. The work reported in this study was actually done from September through December 1961. Since that time, several changes have been made in *Chemical Titles* format, editing rules, and word omission rules. It would appear from the changes which have been made that were a study made now, the title might be even better in indexing every concept than it was at that time.

It is only fair to point out, of course, that the human indexing done on these documents was done from the abstract and not the full text of the article. It may well be that the indexers relied too heavily on the document titles for their indexing, neglecting the abstract text; or the abstractors relied too heavily on the titles in preparing the abstracts. The general point of the article was that authors and editors must be consciously aware of the increasing importance of the title, and, indeed, this is a good point.

At one time I was in charge of the Physics Library at the University of California, Los Angeles. Intrigued by some of the earlier title studies, I selected all of the significant additions to the Physics Library collection made over the period of a year and analyzed the titles of these materials to see whether or not the subject headings which were assigned by the catalogers could have been assigned by a machine process were a reasonable thesaurus available to a computing machine.

The Physics Library was producing a monthly list of new accessions. It was from a year's cumulation of these lists that the titles which I examined were taken. To be included on the list an accession had to be considered a significant addition to the collection. Such things as added copies of titles already in the Library, older materials of little interest other than historic, and elementary physics textbooks of a high school or junior college level were excluded. Also excluded were certain items which were purchased solely for the use of the library staff of the Physics Library, i.e., materials which were to aid the library staff and were, typically, not to be used by the staff and students of the Physics Department.

Before describing Table II which illustrates my findings, I must define one term: the term *conceptual equivalent*. I have chosen to use this term and to create a definition of it for use in connection with Table II. Others have used varying terms to mean, more or less, the same thing. For example, O'Connor used the term "synonym inclusion," and his conditions for such were as follows: (1) a single word heading is identical with a title word: e.g., aedes and "aedes;" (2) a single word heading is an inflexional variant of a title word: e.g., aggression and "aggressivity;"

*Volume 9, Number 1, Winter 1965*
(3) A single word heading is (reported by an authoritative dictionary to be) a synonym of a title expression: e.g., aging and “senescence,” (4) a word which is part of a multi-word heading, and occurs in no other heading, is related to the title by synonym inclusion: e.g., tranquilizing agents and “tranquilizing” or a synonym; (5) each word in a multi-word heading without unique heading words—in the sense of number (4)—is related to the title by synonym inclusion: e.g., stomach neoplasms and “stomach neoplasm.”

I have used the term “concept equivalent” to mean the type of synonymous equivalents which one would find in a subject heading authority file: e.g., the term NUCLEONS as listed in a subject heading authority might say “see PARTICLES (NUCLEAR PHYSICS).” A term such as RADIATIVE TRANSFER might show “see RADIATION.” In addition, there are such equivalents as “conference” and “congress” which, for all practical purposes, mean the same thing in most printed proceedings. The term “anthology” is clearly equivalent to the two-term phrase COLLECTED WORKS which is a part of a form subdivision. A phrase such as “information for the engineer” can be considered to be equivalent to the heading ENGINEERING, and the phrase “men of science” can be considered equivalent to the heading SCIENTISTS.

The other studies of titles were concerned with journal articles. Subject headings used in indexes and abstracts for journals differ, in many respects, from the subject headings used in an academic library. For example, there are many “form” headings used in library catalogs, and they are relatively rare in periodical indexing. DISSERTATIONS, ACADEMIC which is then subdivided by the institution and the department, and the subdivisions such as ADDRESSES, ESSAYS, AND LECTURES or COLLECTED WORKS are examples of headings used at UCLA. These are all commonly found in regular library cataloging or classification, but they are not generally a part of the problem of producing subject indexes to journal articles. Another phase of library cataloging which is not found in the normal journal index is the corporate author. In some cases, the combination of corporate author and title may well indicate something of significance in the automatic production of subject headings. For example, the proceedings of a conference on vacuum technology would form a corporate entry in a regular library catalog, and the title alone, which might be “Proceedings” or “Transactions,” normally would be meaningless in an automatic system. However, the combination of the corporate author plus title could well be used to indicate the exact subject heading, in this case VACUUM-CONGRESSES, which is given to such material.

In Table II, materials are divided into five classes. The first class (Equal) is made up of items for which words in the title were exactly matched by the subject headings chosen by the cataloger, or were conceptually equivalent (as defined above) to the subject heading or headings chosen. In many cases there were multiple subject headings, and to

- 46 -

Library Resources & Technical Services
be included in this class, all subject headings had to be matched exactly by words in the title.

The second category of Table II is No Match, which is used for those title-subject heading pairs which are not clearly enough connected, so that a simple process utilizing computer look-up in a machine thesaurus probably could never produce automatic subject headings assignments.

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of Titles</th>
<th>% of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Exact or equivalent</td>
<td>147</td>
<td>36</td>
</tr>
<tr>
<td>2. No match</td>
<td>72</td>
<td>15</td>
</tr>
<tr>
<td>3. Probable match</td>
<td>151</td>
<td>37</td>
</tr>
<tr>
<td>4. Special problems</td>
<td>19</td>
<td>5</td>
</tr>
<tr>
<td>5. No headings</td>
<td>21</td>
<td>7</td>
</tr>
</tbody>
</table>

Total number of Titles ............ 410

1 For explanation of categories, see discussion in text.
2 Percentage figures are rounded off.

The third category is Probable Match, which is used to indicate that: (1) either the subject headings actually assigned are not exact matches, but are still probably "conceptually equivalent"; or (2) that where there were multiple subject headings assigned to a given title, only one or two, but not all, of the headings matched. Many of the items in this category actually had perfect matches of perhaps one or two subject headings. Yet there were more subject headings assigned, and the remainder did not contain matching words or matching concepts.

The two final categories are really equivalent to the second type, that is, No Match. I have arbitrarily divided these two into (4) Special Problems, and (5) No Assignment Made. Special Problems indicates items which were too difficult conceivably to be assigned by machine processes but which had elements of further interest in them for detailed analysis.

Let us consider the title "Statistical Theory and Methodology in Engineering." This title was given two headings: (1) MATHEMATICAL STATISTICS, and (2) EXPERIMENTAL DESIGN. Now a knowledgeable human can see easily the relationship between these subject headings and the words in the title; yet looking at this from the other side, it seems doubtful that even a very knowledgeable human would have been willing to suggest these two assignments without, at least, looking a little further than just the title. At any rate, these particular items were interesting enough to me to be put into a special category for further study and analysis.

The final group, those with "No Headings Assigned," are all dissertations for the doctoral degree in physics. Each of these was assigned a form heading, i.e., DISSERTATIONS, ACADEMIC—UCLA—PHYSICS by
the Catalog Department, and sent to the Physics Library for purposes of subject heading assignment at a later date. These items were all so difficult (i.e. esoteric) no regular subject headings were assigned by the time they were entered on their respective New Booklists. Subsequently, they were given some sort of subject heading, but the headings given were not necessarily those used by the Library of Congress or the UCLA Cataloging Department, and they were assigned in the Physics Library and not by the regular catalogers. Thus, they were not included in this study.

What does all of this mean in terms of automatic indexing as it now stands? We must ask ourselves, first of all, what is it that we are attempting to produce when we classify or index? We are attempting to produce a product which is a representation of a body of knowledge. We must have this representation, abbreviated though it may be, since we cannot examine each and every item in our collections. This latter, of course, would be the only sure way of determining whether or not each and every document did or did not contain something relevant to a given individual's need for information. Even in one's own personal collection of books and journals, it is not possible to examine each and every item every time one needs to find some bit of knowledge. Thus, our products are for the purpose of presenting a searchable representation of our library collections.

Enough research has been performed to indicate that any system of subject headings, coordinate index terms, machine retrieval, whatever, which continually makes its specifications narrower, in order to eliminate irrelevant material, will be subject to the error of over-specification. This error will cause relevant material, to a greater or lesser extent, to be missed. The problem, then, is not to find a solution which will be theoretically perfect and which will prevent all irrelevant retrieval as well as produce all relevant materials in any given collection. For such a solution exists only as a rather fantastic dream. Rather, our goals should be to produce a system which will minimize irrelevant retrieval and maximize relevant retrieval to the greatest extent practical. The emphasis should be on the word practical in all cases.

Actually, we are operating very much in the dark. Justification of any system must ultimately be based on successful retrieval, but how can success be evaluated within an existing collection of any size which is in operation on a daily basis? It cannot. Success can only be evaluated in terms of a closed system, that is, a system wherein sufficient knowledge is available of the entire contents of the Library, so that evaluation may be made of various techniques. And, unfortunately, there is no way of knowing whether we have retrieved all pertinent or relevant materials in a real situation. It is impossible to know, since it is impossible to be familiar with every item in a collection.

We might now ask, why should anyone be interested in automatic classification and indexing? There are many reasons. For example, many, many libraries have large arrearages in their catalog departments. This would seem to indicate that they do not have enough catalogers to pro-
cess the materials which are coming into the system. Also, in all systems utilizing human beings as classifiers and indexers, the success with which information can later be retrieved depends critically on the care and effort expended in analyzing the materials as they are entered into the system. That human beings do not perform optimally at all times is well known, but frequently forgotten. Indeed, the level of human performance is quite variable from one period to another. This point is covered in a recent article by DeLucia which appeared in the April 1964 issue of *American Documentation* and was entitled "Index-Abstract Evaluation and Design." We frequently tend to forget that human beings have daily cycles of efficiency, that individual classifiers have varying abilities depending on the subject matter they are indexing, and that outside pressures frequently cause a decrease in efficiency. In comparing machine processes with human manual manipulations, there are some who invariably show the human system in its best light, as if every individual involved were a veritable paragon of intelligence and virtue; one who has never made a mistake in his life, knows his job and subject matter perfectly, has learned (and remembers) every policy memo sent to him (and to everyone else in the organization), has always shown acute insight as well as considerable foresight, and can work tirelessly, at peak efficiency, 365 days of the year, at least 28 hours a day. The fact: people come and go, are sick, die, are replaced by new employees, must be trained, do not always work at peak efficiency, take vacations and coffee breaks, etc., may walk off with large quantities of the institution's supplies and equipment, may suffer memory failures—in other words, they are human. None of this seems to bother these foes of mechanization.

Experiments have been performed that seem to show that automatic indexing or classification can produce usable, reliable, and, above all, practical results, provided that some measure of machine readable and interpretable input is available. A machine system for classification or indexing depends on the ingenuity and resourcefulness expended upon the development of the system, as well as the continued proper operation of the machinery that is to be a part of the system. Mechanical and electronic reliability is now specifiable beforehand (in terms of mean hours before failure), whereas human reliability cannot be predicted. Provided that overall effectiveness is nearly equal, the system that depends less on the human element would clearly seem to be more desirable from a standpoint of reliability and efficiency, and perhaps even from a standpoint of economics as well.

It is perhaps a truism that input is more important than output, for if there is failure at the output end, i.e. in retrieval, one can always try again. But if there is failure at the input end, the material is most likely irretrievably lost. In a system utilizing human classifiers and indexers, overall performance of the system, particularly for retrieval, is a direct function of input quality. The quality of input is, in turn, a direct function of four elements concerning personnel: (1) availability of qualified
individuals; (2) their knowledge of the system by which they will index, i.e., the system of classification or subject headings; (3) quality of their training; and (4) continuing coordination and auditing of their work.

In a mechanized system, only one of the above four items is a factor, i.e. the continuing coordination and auditing of the system performance. The other three factors are taken care of once and for all at the beginning, during the establishment of the system. Subject heading authority guides, the knowledge of many individuals, cross references, etc., can all be placed in a machine memory, and the system tested again and again, each time perfecting it a little more, so that ultimately a practical system is obtained.

To conclude, I should like to summarize a system which I believe can be used to provide automatic classification and indexing.\(^\text{19}\) It depends on more than titles alone. What I would add as machine readable input is as follows: table of contents; an index contained in the book itself (if any); introductory paragraphs which describe the contents of the book; a short indicative abstract (where available) would also be useful. Such material could be put into machine readable form (where it is not now available) by clerical personnel, following a consistent set of rules as to the elements of the materials to be used and in what order they are to be followed.

This system works by means of a machine thesaurus which contains a vocabulary of words weighted for retrieval importance. The computer looks up in the thesaurus every word and contiguous word pair of each sentence of input. (In this sentence contiguous word pairs would be: “every word,” “contiguous word,” “word pair,” etc.) The thesaurus contains cross references, and all words are grouped into synonym classes. The computer keeps track of the location of each word, and thus proximity and pairing factors can be calculated. (Pairing factors refer to combinations of words which make up subject concepts. For example, the terms nuclear, power, and propulsion when joined together as in nuclear power, or nuclear propulsion, mean something quite different from nuclear alone, power alone, and propulsion alone.)

Weights and pairing factors are put into the thesaurus by knowledgeable human beings. That is, certain words in certain subjects are more important than other words. This is readily recognized. Thus, in determining subject headings, weights and pairing factors are very important, but they are created, for storage in the thesaurus, by humans.

For each subject to be included in the automatic system, a separate thesaurus must be constructed. This is similar to creating a special subject heading list for any given subject. As a matter of fact, subject heading lists provide a good basis for the beginning of a thesaurus. However, it is suggested that processing of natural language text from various subject disciplines may well provide new terms not now in existing subject heading lists.

The whole process is perfected by running it parallel with human classifiers and indexers through a pilot period when both systems process the same materials. As machine errors are discovered, it is hoped com-
puter programming rules can be modified and changes made in the thesaurus so that the next trial, using different materials, will then result in a better process. This can be continued for a considerable length of time, ultimately, one would expect, resulting in a practical machine system for classification and indexing.

It is not clear, at this point, how much machine-readable input is necessary to produce usable automatic classification and indexing. We have mentioned the numerous tests and experiments which looked at titles alone as the potential machine input. It would seem that titles alone are not enough, but how much more we must have in each and every subject discipline which is covered by a large academic library, no one can say. Actually, no one can really say how much is needed in any subject discipline at the present time, although one has reason to believe that in the sciences it may be that, with carefully constructed titles, very little more than the titles and the table of contents would be adequate. This, of course, would mean that the input costs could be quite low.

Certainly, the categorizing of printed materials is a profound matter. No one is suggesting seriously that all humans will be relegated to a state of useless bystanding. There are obviously many, many books and other printed materials which are exceedingly difficult to classify or index because of their esoteric contents or the opacity of their writing. In a system utilizing automatic indexing or classifying, there will always be a need for knowledgeable humans to take the difficult materials which the machine is unable to deal with. Rather than making catalogers feel inferior to a machine, if they are given only the most difficult materials to catalog, it should make them feel superior since the machine process cannot do the job.

How close are we now to an automatic system? Not many months ago a Washington, D. C., firm, Information Systems, Inc., announced the availability of a system of automatic cataloging and classification employing the IBM 1410 Computer. I have seen a sample output of this process, and it appears quite usable.

If we are truly interested in better and better library systems to serve our varied clientele more usefully, and if a machine system of cataloging, classification, and indexing can make a library more usable, then no one should oppose it. If a machine system cannot do this, then it should be discarded. At this point we do not know which will prove superior, but we ought not to base our decision solely on obvious cost factors.

I suspect that there are some who will label me a misanthrope. I shall deny that, although I must confess that there are certain properties of machines which I find appealing. Machines believe everything one tells them, literally. They never misinterpret one's meaning. They rarely gossip, nor do they impute sinister motives to one's every action or word. Let me hasten to add that I do not wish to imply that humans do those things either. But they possess the means to do so, and machines do not. I suppose that the moral of this tale is that, if you have any choice in the
matter, you should choose humans for your friends and machines for your enemies.

REFERENCES


* Library Resources & Technical Services
I SPENT OVER a week looking for a story concerning a humorous aspect of serials, but I finally came to the conclusion that serials aren't funny. I think most librarians will agree with me.

A friend of mine who sells computers was once talking of the "good old days" in his sales career when he sold encyclopedias door to door. He then made a sale, delivered the set of books, pocketed his commission, and never saw the customer again. Now, as a computer salesman, he had to work with the customer's employees after the sale so they would be trained before the machine was installed. He had to see that they got off on the right foot after delivery time. He had to keep training new customers' employees. He had to keep software up-to-date even on out-of-date computers. He said that acquiring a customer now meant that he would become that customer's nursemaid for life.

A similar situation exists in libraries in considering monographs and serials. A monograph is a one shot deal. Someone wants a particular item, we identify it, put it on order, catalog it after it arrives, and then, as far as technical processes is concerned, forget it.

Serials, of course, are like the computer. Once a serial title is on order, we become its slave. Even at the time of initial order we have to decide whether to start with the current issue or the first issue of the current volume. Do we want backfiles at once? Later? Do we want microfilm or the original paper? Once the thing arrives, then binding rears its ugly head. Every year subscription renewals have to be made. Changes in title and in frequency, mergers and splits, non-arrivals and subsequent claims have to be watched for like a hawk looking for his lunch.

Well, you know all the problems. Serials are just a mess!

It is precisely because of these problems that some sort of systematized solution has been sought for serials. In Joe Becker's recent article in the *ALA Bulletin,* he briefly covers the history of serials record keeping. Except for a handful of libraries and a lot of wishful thinkers, serials record keeping stopped at the Kardex or Acme type file in either a centralized or non-centralized system. The usefulness of these files is usually dictated by their size. A small file of, say, less than 1,000 titles, usually has all the information about each of the titles located on one record. Check-


*53*
ing-in, binding, finances, subscriptions, holdings, claims, etc., are in one place. The amount of floor space this record covers is less than that covered by a desk, and one person can easily maintain the file.

As the file grows in number of titles, it usually grows disproportionately in bulk, too. The choice is either to put fewer items into a tray at a time (when more items per tray are really necessary) or to take some of the information out of the file and store it in a second (or third or fourth) file. The file is now decentralized, more people can work at the records, and material in the file is not damaged by overcrowding. The file, of course, takes up more floor space, more physical movement is required in using the main part of the file, more knowledge and experience is needed in using all parts of the file, and certain pieces of information have to be expensively repeated on each of the different records. This, then, is the situation in which most of us find ourselves. We are sure things will get worse before they get better.

It is too bad that library solutions to serials problems have to come after the problem is created. The several committees of the professional organizations have suggested ways in which publishers could cease creating problems, but so far, suggested plans for standard format and location of bibliographic data on a serial piece, volume and piece numbering, size and thickness of a single piece, and other quantitative measurements of a periodical subject to variation and of interest to librarians have been ignored by publishers even in the library field. I feel that persons as individualistic as publishers, their layout men, and artists will probably never cooperate in a significant number unless standards were to be tied into their second and third class mailing privilege. I'm not sure that I'd recommend any measure this strong.

Let us keep in the back of our minds, though, the fact that problems with serials are created by the publishers, but for now we will have to find solutions to these problems within the library.

Serials seem to be deceptively simple to automate when, in fact, they are more complex than any other phase of library technical services. It is the definition of serials which gives us this sense of false security. We might, for example, define a serial as an item issued in parts, one or more of which must arrive each year. The definitions of what constitutes a serial are almost as varied as the number of librarians making the definitions. However, regularity of arrival (or at least an assumed regularity), is the basis for nearly all such definitions. It is this assumed regularity which presents the problems. We can illustrate this very nicely if we make a list of all the arrival frequencies possible with the serial titles which come under our definition. First, we have once per year in any of the twelve months, then twice per year with the first one arriving in any of the first six months and the second piece arriving six months later. Then an item arriving three times per year with the pieces coming in any month, then four times per year with arrival in any of the twelve months with the added complication of their sometimes being dated by the month
and sometimes by the season. We can carry this through to the newspaper which is issued everyday of the year without fail. Assumed regularity means actually any possible combination of months, weeks, and days. Some kind of a formula can now be set up so that peculiarities of any serial arrival can be represented with a short understandable notation.

This notation of frequency would not do a clerk much good, for a quick glance at the check-in record or possibly even the piece itself would tell her as much in less time. However, this short notation can be punched into paper tape or into a tabulating card with other similar notations concerning payment, subscription, binding, holdings, and claiming procedures. These notations can then be sorted and compared in many different ways, so that one particular serial with certain attributes or all serials with common characteristics can be noted.

One very interesting application of this nature is being used at the University of Illinois, Chicago Undergraduate Division Library. While it is a non-computerized application, I would like to describe it briefly because I think that any library which had access to a card punch and a sorter could benefit from it.

The original purpose of this file was to gather information about the serial titles that was not listed in the Serials Record. As the record was not a centralized one, although it was called a Central Serials Record, it was felt that some kind of intermediate record was necessary. The intermediate record was originally organized around one punched card for each title. An abbreviated title was selected using a small set of abbreviations limited to 95 characters. An eight-character sequence number was also assigned. All titles could then be put into alphabetical order by sorting on this number only.

One column was devoted to method of acquisition. A one code punched in this column meant that the title came as the result of a membership. A two meant subscription, a three was standing order, and so forth. Two columns were devoted to source of acquisition, and one column each was devoted to the month in which the volume number changed, the month the index was published, the month binding was picked up, the year the item was bound, the method of index arrival, the shelving location, the type of binding, and a code denoting dead titles or items not checked in.

The file was used as follows: everything that was known about a serial title was assumed to be in the serials record. All this information was punched into the card. If it turned out that the month the index was published was not in the record, then the column was left blank. Later the whole file was sorted on this particular column. The items which had been left blank would drop out in the reject pocket of the sorter. These cards would then be taken to the shelves and checked against the actual periodicals or they could be checked against Ulrich's or other serial bibliographic sources. In this way the information in the serial record was completed.

* 55 *
Louis Schultheiss, who thought up this whole idea in the first place, now realized that he had all kinds of valuable information captured in a machine-readable form. Why not get some immediate use out of it? The file could be run through the sorter, and the cards for the serials which have indexes which should have arrived during the month dropped out. A quick check against the receipt file or the file drawer where the indexes are actually stored would show which ones had arrived and which ones should be claimed.

Or the file could be sorted on the column denoting month to be sent to the bindery. Only the items to be sent this month would be dropped out with the rest of the file left in order. The cards for the items to be sent during the month can then be sorted again on the column showing shelving location. It would be simple then to hand a clerk a bundle of cards and tell him to go to the Reference Department and take the last volume off the shelf and prepare it for the bindery. He could do the same with the bundles of cards for serials shelved in all other locations in the library.

These are but two examples of how useful a punched card serials file would be to any sized library. Every librarian should be able to think up many other applications.*

Now, why this long-winded description of a non-computerized serials record when this talk has been billed as one dealing with computerized serials records? First, the type of work involved prior to the application is the same in both cases. That for the computerized record has greater scope, and it must be absolutely complete before any of it is useful. Second, the type of applications are the same, although those which are computerized can be of much greater variety and of infinitely greater sophistication. Third, the goals are essentially the same: better service to library users for less cost.

Let us take a look at three typical computer applications to serials problems of libraries. The first is a holdings list. These lists have become very popular as serials records have become both more complex and less accessible to the public. They can range all the way from very restricted one line per title lists to complete ones which have as much information for each item as a library catalog entry. Once the decision to make such a list as this has been made, the scope of the list has to be decided. I believe that this is the most difficult part of the whole procedure. The librarian’s definition of what constitutes a serial must be brought out and re-argued. Should duplicate copies be listed or only the most complete one? What about government document serials? Are continuations fair game for the list? Will dead titles be included, or is this for live ones only? Is a distinction going to be made between a serial which died and one which the library merely stopped subscribing to? In other words, where you draw the line is up to you.

* Mr. Schultheiss has prepared a detailed account of this process as well as its computerized sequel. This paper will be published in a forthcoming issue of LRTS.
Once the scope of the list has been decided to everyone's satisfaction, then the form of the individual parts of the list can be made. Is the title going to be complete in every case or abbreviated in some cases? If it is to be abbreviated, then how short should the abbreviation be? The library of the University of California at San Diego limits its short title to 28 spaces. The University of Illinois at Urbana limits that one to 569 characters or complete title for all practical purposes. The eventual intention is to use a computer to do their abbreviations for them. The University of Illinois in Chicago and Florida Atlantic University use the belt and suspenders philosophy and have both. The short title is limited to 35 spaces, and the full title will be complete in both cases. Whether or not call numbers should be included is another possible bone of contention. It depends on the use of the list, of course.

Lists intended primarily for the use of patrons definitely should have call numbers so that a double table look-up is avoided. How extensive should cross-references be? Again, it depends on the use. (I would tend to be extra generous.) Finally, what will the holdings statements look like? How complete will they be? More blood has probably been let over this problem than any other. Let's assume that both volumes and years will be included. Should parts be listed if the volume is incomplete, or should the volume number be put in brackets or curves to show an incomplete volume? Erring on the side of completeness is best. If material is handled by electronic computer, the computer can take away what is not needed and store it for later use, but it cannot add what is not there.

Finally, the physical format of the list can come in for limited discussion. If the computer has a printer with only upper case and a limited number of special characters available, this is what will be used. If lower case is available, then the list will show greater variety and be easier to read. Some computers have a special feature called "Space Suppress." This, in effect, causes the computer to print certain designated lines two times, one on top of the other. Because of the vibration of the paper, the result looks like boldface. Argonne National Laboratory has a serials list which makes very effective use of the Space Suppress feature. Most important, though, is capturing all that can be captured (lower case, if possible), using what can be used, and saving the rest for future use.

The computerized serials list which is the most attractive of all is that of the Massachusetts Institute of Technology. It is produced using a large scale computer which produces paper tape which is in turn fed into a Photon photographic typesetter. Capitalization is handled through the computer program as the initial input is in only one case.

The second type of computer application is that dealing with a single phase of serials handling. Checking-in, subscription ordering, bindery routines, routing, and so forth can be tackled as a discrete problem. Several libraries are approaching machine handling of serials in this manner. Naturally, librarians want to approach the problems first which irritate them most. And with certain libraries there are only one or two
thorny problems. The rest of the serials routines would be so simple that it would be useless to waste computer time on their operation—the manual method is able to cope quite well. Unless a library is in this type of a position, I do not recommend that a spot approach be made.

The third type of serials application and the one which will become most popular with libraries of any pretension to size is the total system approach. It would be assumed that other phases of library mechanization besides serials would eventually be added or would be in the planning stage at the same time as serials. Except for financial routines, serials can be separated completely from other aspects of library automation. Briefly, the procedure to be followed in a total approach is, first, to make a complete study of the present system. With this background, the demands of a future system can be stated without equivocation. A list must be made showing every piece of information which has to be put into a system in order to achieve the desired results. In order to get this, the layout of the resulting documents must be known down to the last period. Now the computer programmer can be turned loose on the problem. It is up to him to take the data he was told he would have and turn out the documents which are needed.

Once the computer program is written and tested, the full file of serials information must be converted into machine-readable form. Having gone this far, most librarians will probably go the rest of the way, but the conversion will be the most expensive and time-consuming part. It will also be the largest source of errors and will plague the system for months. I would not advise discarding the manual methods too soon; both the new and the old systems should be operated in parallel for some length of time, at least until the operators are convinced that the bugs are out of the new system altogether.

Now the public can be weaned away from the old, barely adequate system to the new one with its infinite expansion and perpetually-available information.

REGIONAL GROUPS

Only two recent meetings have been reported in time for inclusion in this issue.

The Technical Services Section of the Wisconsin Library Association heard Pauline A. Seely (Denver Public Library) discuss "ALA Filing Rules—New and Revised." Elizabeth Rodell, RTSD Executive Secretary, then spoke on "The World of Technical Services."

The New England Technical Services Librarians at their business meeting voted to become a section of the New England Library Association. The group retains its present name.—Doris Ransom, Chairman, Council of Regional Groups
MY ASSIGNMENT is to present a case study of the automated book circulation system now in operation at Southern Illinois University, Carbondale. After almost three years of planning and experimenting, the system went into operation on a limited scale this spring. By means of a step-by-step conversion from manual to automatic operations, we have been able to test the equipment under actual operating conditions, to improve upon certain routines, and to adjust staff assignments with a minimum of difficulty. We expect the circulation process to be completely automated when school starts in the fall.

Before describing the system, let me give you a picture of the circulation problems that led to our efforts at automation.

Southern Illinois University has an enrollment of some 20,000 students, 14,000 of whom are on the Carbondale campus. With the exception of four small specialized libraries, all books on the Carbondale campus are housed, and library services are provided, in a central building—Morris Library. There are approximately 600,000 volumes in the central collection, arranged in four subject divisional libraries and classified according to Dewey. Except for the Rare Book Collection, all books are on open shelves; and a central circulation point on the main floor serves the entire building. The concentration of book circulation at one point was one of the factors that made automation feasible.

When we began the study of our circulation system in the fall of 1961, book circulation from the central library had reached a peak of a thousand volumes a day, and the annual rate of increase for some years had been in excess of the rate of increase in enrollment. The circulation system we had been using effectively for many years was a combination of McBee Key-Sort and Gaylord electric charging. The increased volume of books circulated in recent years had created a burden that the system was unable to support. For one thing, the congestion around the files, i.e. the filing of cards, the needling for overdue books, the answering of inquiries about books charged out, the library clearance of seniors, and the removal of cards in the discharging of books, forced us to move the files from back of the loan desk to another area. Separating the files from service contacts created problems of communication and required additional staff. When the situation became critical in the Fall of 1961, we turned for assistance to the University’s Office of Systems and Procedures.
I am certain that our immediate problems could have been met by a solution less drastic than complete automation, but the existence of a data processing unit on our campus interested in extending its services, and the fact that we were already using IBM facilities for book ordering and might make further use of automation in other library routines, prompted us to consider automation of book circulation. Furthermore, there was the prospect of correlating statistics on book use with student grades, test scores, and other data that would eventually be fed into a campus-wide data collection system.

We were fortunate in having assigned to our library project a young graduate student in management, L. R. DeJarnett, who not only conducted the original study, which he incorporated in a master’s thesis, but who stayed on the University staff to design the system and see it into operation.

The study of the circulation system then existing consisted of gathering statistics and observing patterns of use, preparing flow charts, and, where feasible, conducting time and motion studies of specific operations. In all of these activities Mr. DeJarnett, under the direction of the Head of the Office of Systems and Procedures, R. D. Isbell, worked closely with members of the library staff and with the staff of the Data Processing and Computing Center which would eventually be involved in the operation of the system.

In addition to the congestion around the files and the pile-up of books waiting to be discharged, the study revealed a serious problem in the handling of “snags,” i.e., books which could not be discharged immediately because no charge card could be found. We know snags can be the result of a number of possible errors—incorrect copying of the call number on the McBee card (an error not detected by the circulation clerk at the time of charging), an error in filing the card (sometimes the result of difficult handwriting), or an error in withdrawing the card prematurely. When we consider that some forty student assistants, as well as six full-time clerks, serve at the loan desk during the one hundred hours a week the library is open, we know that the chance of error is magnified. Another trouble spot was the whole complex of pulling cards for overdue books, typing and mailing of notices, and recording payment of fines. One week’s work on overdue books was hardly finished in time to start the process over again.

In addition to studying the existing system, Mr. Isbell and Mr. DeJarnett visited libraries where automated circulation procedures were being used and industries where automated inventory controls seemed to have some bearing on the library problem. They also examined numerous pieces of equipment which might be used in creating an ideal system and sought the advice of experts in the computer industry, particularly the IBM people, whose equipment we decided to use. The system which was ultimately devised, while it borrows ideas from several existing installations, goes beyond those known to us in the extent of its automation.
and in the completeness of the controls which it provides. For example, one of our requirements which complicated the design was that the system should permit making requested books available to readers as promptly as possible through the device of "reserve" and "recall." The circulation study made by George Fry and Associates noted that this requirement distinguished the need of the university library from that of the public library. The use of a "transaction number," common to most automated circulation systems, would not provide positive identification of an item in circulation.

Although the problems encountered in the design of the system were numerous and sometimes complex, the system that evolved was basically very simple. Three items of information are assembled in charging out a book—the borrower's identification number, the call number of the book, and the date due. Together, these items identify each transaction. They are fed into an IBM 357 data collecting unit and form the basis for the tape storage of circulation records.

Let me describe the process of charging and discharging books. The borrower presents his book at the circulation desk, together with a plastic identification card. The identification card contains the borrower's picture and his name, address, and identification number, embossed after the fashion of a gasoline credit card. His identification number has also been punched on the card in machine language. This ID card, in use on the campus for several years, serves many purposes besides library borrowing. It is issued to students by the Student Affairs Office at the time of registration. Faculty cards, issued by the Personnel Office, use the social security number for identification.

Each library book contains in its pocket a master IBM card bearing the call number of the book, expressed both in printed and machine language. The coding was devised so as to provide all of the necessary data on a single line of printing and still fit on a shortened IBM card, utilizing only 47/8 inches of the standard 73/4-inch IBM card. This enables the master book card to be folded when it is inserted into the pocket of a book of less than octavo size. The call number line is divided into six units for convenient reading—the Dewey number, the Cutter number, year, volume, part, and copy. Twenty-five units were provided for the call number, with an additional 16 units to accommodate the borrower's number and due date, a total of 41 units. I shall discuss later how these cards were prepared for a half-million volumes.

The book card and the borrower's card are collected by the circulation clerk who depresses the due date (the third item of information in the transaction) on a manual entry keyboard (IBM 372) which is attached to the basic data collection unit (IBM 357). Due dates will vary, of course, depending upon the nature of the book and the status of the borrower. The book card and borrower's card are inserted into two slots in the 357 unit. This immediately activates a printing key punch (IBM 026) which is located adjacent to the 357 input unit. The key punch pro-

Volume 9, Number 1, Winter 1965 • 61 •
duces two cards, each bearing the three items of information both in printed and machine language. One of these, a pink card, is returned with the master book card to the pocket of the book and represents the borrower’s notice of date due. In addition, this card serves as a valid charge record at the exit check points. The other card (yellow) is stacked in the machine as the Library’s record of the transaction. The identification card is returned to the borrower along with the book, and the clerk is ready to handle another transaction. The entire transaction takes about twenty seconds, only eight of which is machine time.

The IBM 357 system consists of two 357 input stations, back to back, each equipped with a 372 manual entry keyboard for recording date due. These are joined by a single 358 control unit (the “little black box”) and a single 026 key punch which serves two input stations. The control unit not only regulates the timing of the operation, but provides a read-back circuitry that performs an audit of the printing and punching. The system is mounted on a movable dolly. Three such systems with a total of six input stations constitute the final configuration of equipment for automated book charging in Morris Library. With the exception of a slight alteration of the throat of the 357’s to accept embossed cards, no modification of standard hardware was required.

When a book is returned, the clerk needs only to remove the date due card, verify that it is for that book, and send the book to the shelves. At the end of each day (11:00 p.m.) the accumulated transaction cards and return cards are taken to the Data Processing and Computing Center where they are loaded into an IBM 1401 computer unit to update the circulation file on magnetic tape. By 9:00 the following morning a consolidated circulation list, printed in call number order, but giving borrower number and date due for each transaction, is delivered to the library. The computer program arranges the print-out by divisional library so that a copy of that portion of the book charges is supplied to the division for public use. A borrower who fails to find the book he wants on the shelf may check the charge record posted in the divisional library to learn when the book is due to be returned.

Overdue book notices are prepared and addressed daily by the computer, using the circulation record tape, together with a tape file of the names, and addresses of borrowers. This latter file, incidentally, is the basic “tie-in” to the total systems concept, and can relate library usage to such factors as grades, age, housing, etc. Fines for books which are overdue are assessed automatically according to a schedule provided in the programming and from two items of information—borrower status and date due—which were supplied in the original transaction.

I have described the basic process. There are a number of refinements, exceptions, and limitations. We have had to make some compromises in the interest of mass processing, but we have not had to adapt ourselves, as Lewis Mumford has feared, “with pathetic docility to the limitation of the machine.” For example, we have recognized the need to provide some means of charging a book manually when the borrower does not
have his ID card. While we expect to be strict in requiring the use of the ID card once the system is in full operation, some exceptions will be made, particularly with faculty. We are determined that the use of the machine will be tempered with human understanding, despite protests from our systems people that even a dean should be turned away if he leaves his ID card in his other pants. The fact that the manual process is more troublesome for the borrower as well as for the library should encourage the use of the ID card.

We also have to make a manual transaction when the master book card is missing or, in the case of a book returned, the date due card is missing. (These IBM cards do make handy book marks.) Manual charges must also be made for certain categories of library material for which no master cards were made. This includes a collection of some 80,000 maps which are seldom circulated outside the building and which would have been difficult to process. We also excluded those government documents classified according to the Superintendent of Documents scheme; after considerable study, our analysts gave up as hopeless the task of fitting the notations of this classification scheme on a single IBM card. Bound volumes of journals are also not provided with master cards because they are not classified and because there is a limitation on out-of-building use.

When there is no master book card, the book is charged out manually on a so-called Universal Charge form, using the embossed portion of the ID card in a manner similar to that employed at gasoline stations with the credit card and with much the same equipment. The call number or other identification of the volume is filled in by hand. The Universal Charge form provides three copies—one copy goes into the book pocket as a date due reminder, a second copy goes to the Data Processing and Computing Center for preparation of a master card and return card, and a third copy is retained by the Library as a temporary record, pending the receipt of the permanent card. The same set of forms is used for charging books without an ID card, the name and address of borrower being entered by hand. We have estimated that the number of manual transactions after the system is in full operation will be considerably less than one per cent of the total. Two-hour reserve books, because of the frequency of turnover, are excluded from the machine operation.

Specially designed "courtesy cards" which can be accepted by the IBM 357 unit will be issued on a term basis to guests, visiting scholars, short-term students, faculty wives, and to record removal of books for such internal processes as binding, recataloging, etc.

The two most frequent questions asked by visitors about our automated circulation system are: How long has it taken you to put the system into operation? What is the cost?

As to time, the design of the system and the development of the equipment took about a year and a half; the coding of the master book cards and the insertion of the cards into a half-million books has taken another year and a half. There was some overlap in the two phases.

The total cost of the project, excluding the salaries of the staff mem-
bers who worked on the design, was approximately $40,000. This included
the purchase of all equipment in the three systems. The operating cost
of machine time at the Data Processing and Computing Center is not
charged to the Library nor is the cost of student and faculty identification
cards. The Library pays for the IBM card stock, courtesy cards, and the
expense entailed in mailing overdue notices. Because of the division of
operating expense between the Library and the Computing Center, it
would be difficult to establish unit cost figures. One full-time file clerk and
the services of a number of student assistants have been eliminated. On
the other hand, additional clerical staff is required at the Computing
Center. There will also be the expense of machine maintenance, this is a
cost factor often overlooked in automation. Accuracy, efficiency, and the
potential for handling increased loads, rather than a significant dollar
savings, are the major advantages of the automated system.

The most time-consuming and costly phase of converting to the new
system was the coding and pocketing of the master book cards. This pro-
cess was greatly simplified, however, when the Systems and Procedures
Office, in cooperation with Science Research Associates, developed a pencil
code sheet which could be optically scanned, thus obviating direct key
punching. Working from the Library's shelf list, trained student assistants
transferred the call numbers to code sheets, eight volumes to a single sheet,
i.e., four on either side. The form allowed for coding multiple editions,
volumes, or copies of the same book by using only one coding entry. The
coded sheets were converted to a magnetic tape record by the Docu-
Tran. The magnetic tapes, in turn, were processed on the IBM 1401 to
create the master cards. A print-out of these punched cards was matched
with the shelf list to discover any errors in coding. Master cards for cur-
rent acquisitions are manually punched from one of the multiple order
slips which is sent to the Computing Center as soon as a call number has
been assigned to the book by the cataloger.

There may be interest in the procedure for making the identification
cards. When a student is enrolled in the University, his picture is taken
by a special camera furnished by the Photo Identification Company of
Chicago. The film is sent to Chicago for development and the prepara-
tion of a laminated card. The card is returned to our Computing Center
where it is punched, then forwarded to the campus Photographic Service
for embossing. Each card is proofread by running it through a 357 unit
and comparing the print-out with the embossing. All this takes a week to
ten days. The card is good for a student's entire college career. It is up-
dated to show payment of fees for the current quarter by means of a
certificate of registration. The identification card and the current certifi-
cate are issued in a two-window plastic envelope suitable for carrying
in one's billfold.

During the course of the recent trial period we have run into numer-
ous problems, most of which have been of a minor nature and were
readily solved. There remain only two unsolved problems of any conse-
access storage. At such time the book circulation record could be transmitted directly from the input station to central storage equipment without benefit of punched card. Conversely, a book could be discharged simply by inserting the master card from the returned book into a discharge unit which would remove the charge from the central data storage and indicate when there was a reserve for the book. In fact the same impulse could generate the printing of a notification form to be mailed to the potential borrower. Furthermore, random access storage would permit a remote control station in the library to furnish immediate information to borrowers on the availability of a book, thus eliminating the need for a print-out of the circulation list. These advances could be carried out, I am informed, as modifications to our present system, once the central storage equipment is installed.

We expect to establish the automated circulation system on the Edwardsville campus when the new library building is completed a year from this Fall.

I can perhaps sum up the advantages of our automated circulation system in this way: (1) it reduces the chance of error in making out charge cards, in filing them, and in discharging books; (2) it saves the time of the borrower who is relieved of filling out a card for every book, (3) it reduces the bottleneck in discharging books and getting them returned to the shelves, (4) it simplifies the routines of sending overdue notices, (5) it simplifies the library clearance of seniors, (6) it lends itself to greatly-increased circulation loads without strain, and (7) it facilitates the compilation of useful statistical and analytical data. And I might add one final advantage. It takes the pressure off the library to pursue impractical schemes in the field of automation, projects that are unsuitable and premature although admittedly fashionable.

In the process of mechanizing circulation routines, both librarians and systems people on our campus have benefited. We have learned to have respect for the ability of the machines to solve our problems, and the computer people have, to their surprise, learned that what appeared at first to be a fairly simple problem, turned out to be highly complex. As Burton W. Adkinson indicated at a recent conference at the University of Illinois, the gap between the computer people and the librarian is closing as the two work together in solving library problems. In our case this gap was bridged effectively by the services of a systems analyst.

* We have since adopted the electrically operated Dashew Datawriter 253.
DATA PROCESSING SYSTEMS for purchasing, distribution, and inventory control have been used successfully by commercial firms for many years, but only a few librarians have as yet adapted the same techniques to solve their own acquisition problems. Many librarians, if not most of them, seem to feel that the purchasing and processing of books, as an over-all operation, is completely and uniquely different from the purchasing and processing of other items, such as machine parts or tomato sauce. To some degree, of course, they are different, and it would be both silly and disastrous to ignore those differences. The physical description of a title being ordered for the library is usually longer and much more complicated than the description required to buy five cases of catsup, and the library order also differs in that most titles are purchased as single units that need not be re-ordered because they do not ordinarily go out of stock. But these are superficial differences and are a matter of degree rather than kind. Nearly all of the requirements of the business system (other than the requirement to show a profit) exist in the library as well: items are described and ordered, bills are paid and financial reports are provided to management, items are received, checked-in, arranged for use, and inventoried at periodic intervals. It seems reasonable to believe that carefully-designed data processing systems can be extremely useful in carrying out these similar operations in large libraries.

One of the great advantages of a data processing system is that data prepared in one format for one operation need not be retyped to produce the document or format required for another operation. Corrections and additions can be made without disturbing other data already in the system, and new documents produced near the end of a cycle can be, to a large extent, by-products of earlier operations. As a result, many repetitive clerical routines can be eliminated, and the amount of revision required to ensure accuracy can be greatly reduced. Members of the professional staff are free to spend more time on truly professional duties.

Several portions of an integrated data processing system designed by the University of Illinois Library, Chicago, and the General Electric Company under a grant from the Council on Library Resources have now been developed in detail and are in the process of being tested under a further grant from the National Science Foundation.

Since the primary objective of the University of Illinois system is the production of book catalogs and printed serials and circulation lists that can be used in various parts of the library or campus, it became apparent
very early in the systems study that a computer, rather than unit-record equipment, would be required to provide the necessary printing capability. The acquisitions program described in this paper is designed for use with an IBM 1401 computer with 8000 positions of memory, four tape drives, and all available advanced programming features. It is not mandatory, however, that the library itself have such a computer, or even that the university have one if arrangements can be made to purchase time from another company or from a data center.

Before going into any discussion of coding, it might be well to indicate how the system assists the staff in carrying out the order function of the library.

At the beginning of the order cycle, bibliographic data in the form of Library of Congress catalog copy plus necessary control codes and order information are introduced to the system on punched cards and converted to magnetic tape. Whenever possible, these bibliographic data consist of the complete LC entry, including the LC call number. A series of computer programs then manipulates the tape record, extracting pertinent sections and arranging it in appropriate formats to produce purchase orders, in-process lists, financial records, book cards, and catalogs. The tape record is added to and modified as necessary, primarily by means of prepunched cards produced automatically as part of the order operation.

During this past year, the Library decided to improve the efficiency of its technical operation by eliminating all local exceptions to Library of Congress cataloging and by changing from the Dewey to the LC classification. This decision, together with the requirements of the machine operation, will shift considerably the workload within the technical services area, and will move many of the duties traditionally associated with cataloging into the area of order verification. The duties themselves do not become less important, but will take place at a different time and will serve a somewhat different purpose than before.

As was the case with the manual system, the order process begins with the delivery of approved order requests to the order section for searching and verification. For the time being, at least, both searching and verification will remain manual operations. An attempt is made to provide a complete LC entry, either by adding to the information on the order request or by providing a proof slip. The bibliographic and order elements are coded for machine identification, and the order is passed on to the Data Processing Division for key-punching and processing.

The order prepared for the vendor consists of computer-printed 3 X 5 slips (see Figure 1), which are returned to the order section with a deck of cards to be used for check-in when the material arrives, and a deck of financial cards that have already been used to encumber funds and will be used again to produce vouchers when payment is finally authorized. The order slips are sorted by vendor number, attached to covering order letters, and mailed. The card decks are placed in tub files and held for later use.
In addition to the order slips, check-in cards, and financial cards, Data Processing produces a new Process Information List, or PIL, every time orders are prepared. This list has two major functions: it serves as an on-order and in-process file, and as a supplement to the catalog. Present plans are to produce new orders and a revised copy of the PIL once each week, although rush orders are produced manually and added to the system as part of the next order cycle (see Figure 2).

Incoming materials and notices pertaining to orders are checked against the PIL. If the item has arrived in acceptable condition, the receipt card is pulled and placed in a new receipts file; if it arrives damaged, the receipt card is pulled and coded to indicate that the material arrived in unacceptable condition. If the dealer cancels the order, or indicates that shipment will be delayed, the receipt card is pulled and coded to indicate these facts. Then, at the established time, the receipt cards removed from the on-order file during the week are gang-punched with the appropriate new status codes plus the current date, and are reintroduced to the system as part of the new order cycle. The new PIL then reflects not only new orders, but changes in the status of earlier orders as well. In addition to the corrected statement in the PIL, three new documents may be produced at this time: a check-in card for the cataloger, a book card, and a pocket label.

---

**EXIAIPIED OF COMPUTER PRODUCED ORDER SLIP (FIGURE 1)**

<table>
<thead>
<tr>
<th>Departmental allocation</th>
<th>Number of copies</th>
<th>Estimated list price</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**LIBRARY RESOURCES & TECHNICAL SERVICES**
<table>
<thead>
<tr>
<th>Title</th>
<th>Author(s)</th>
<th>Publication Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROCESSING INFORMATION LIST</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WINGSEKCI1443 01 000175 753725 8300 1 ORD#</td>
<td>WINGFIELD-STRATFORD, ESME CECIL, 1882-</td>
<td>KING CHARLES AND KING PYM, 1637-1643. LONDON, HOLLIS AND CARTER, 1949.</td>
</tr>
<tr>
<td>WINSF SC1443 01 000350 105075 8400 1 ORD#</td>
<td>WINSOR, FREDERICK.</td>
<td>SPACE CHILD'S MOTHER GOOSE. NEW YORK, SIMON AND SCHUSTER, 1958.</td>
</tr>
<tr>
<td>WRGBF VS1443 01 000750 105075 8300 1 ORD#</td>
<td>WRIGHT, FRANCES.</td>
<td>VIEWS OF SOCIETY AND MANNERS IN AMERICA. CAMBRIDGE, MASSACHUSETTS, BELKNAP PRESS, 6/14/1963.</td>
</tr>
<tr>
<td>WYTRJAAP1403 01 000650 105075 3930 1 ORD#</td>
<td>WYTRWAL, JOSEPH ANTHONY, 1924-</td>
<td>AMERICAS POLISH HERITAGE, A SOCIAL HISTORY OF THE POLES IN AMERICA. DETROIT, MICHIGAN, ENDURANCE PRESS, 1981.</td>
</tr>
<tr>
<td>ZARIO CA1443 01 001690 105075 3500 1 ORD#</td>
<td>ZARISKI, OSCAR.</td>
<td>COMMUTATIVE ALGEBRA. BY OSCAR ZARISKI AND PIERRE SAMUEL. PRINCETON, NEW JERSEY, VAN NOSTRAND, 1958, 1960.</td>
</tr>
</tbody>
</table>

SAMPLE PAGE OF PROCESSING INFORMATION LIST (Figure 2)
The financial operation is very much like the check-in operation. The Financial Clerk matches incoming invoices against the PIL to see whether or not the item being billed has arrived and has been accepted. If so, the financial card is pulled from the tub file, checked for accuracy, has changes indicated, and is then held with the invoice. When all items on the invoice have been accounted for or a partial payment is authorized, the cards are sent to Data Processing to have the invoice number gang-punched into the cards plus the manual punching of changes or corrections. After the production of vouchers, the cards are merged with the financial cards for new orders and the payment is reflected in the new financial summary for the week.

At this point the acquisitions operation *per se* has been completed, and the material is now in the hands of the cataloger. If the complete LC entry and call number were available at the beginning of the order cycle, formal cataloging is limited to a simple matching operation to be sure that the catalog entry fits the edition actually supplied. If part of the catalog data is lacking, or if the copy originally supplied requires modification or additions, the cataloger fills out a code sheet to be sent to the Data Processing Division for key-punching and processing together with the cataloger's receipt card, which has now been punched to show that cataloging has been completed. The next issue of the PIL will show that cataloging is complete, and that the entry can be transferred from the PIL to the catalog when the next catalog supplement is printed.

This completes the description of the order program cycle. For the remainder of the time, I would like to discuss the layout of the cards used to introduce data into the system and some of the decisions that had to be made before these layouts could be finally established.

The individual punched cards are laid out in such a manner that the first 18 columns are used for control information, and the remaining 62 columns for bibliographic or order data. The first 12 of the control group are used for a sequence number, or Luhn number, which is assigned to the title at the time of coding and serves as an identification number for the title until acquisitions and cataloging have been completed. Since this number is based on the author's name and the first significant words of the title, it is also useful as a means of arranging entries in the PIL without going into filing routines (Figure 3).

The next three columns of the control group are used for bibliographic codes. These are required because, as part of its decision to accept Library of Congress copy without exceptions, the Library refused to limit the length of any part of any catalog entry by setting fixed fields for specific kinds of bibliographic data, even though such a decision would have simplified some of the problems of the Data Processing Division.

The problem of using fixed length or variable length fields for catalog data is roughly analogous to a situation requiring the storage of successive shipments of different machine parts packaged in identical boxes. If the storekeeper does not know in advance exactly how many boxes of each part will be received in future shipments, he has two choices for the
arrangement of his shelves: (1) he can set aside a group of shelves for each part, allowing for the accommodation of a maximum number of each kind, risking wasting shelves or the more serious problem of not having enough space to store some shipments; or (2) he can take as much space as he needs for each kind of part, letting the amount vary from shipment to shipment, and identifying the parts by placing labels on the shelves.

The University of Illinois chose to work with variable length fields, and must therefore label the bibliographic shelves to identify the parts of the catalog entry. There are, at the present time, 25 different labels in use. Three of these pertain to order information, and the other 22 identify parts of the bibliographic data. Some of these labels will be rarely used, while others will be used every time; it is entirely possible that some of the present bibliographic groups, such as the collation and the notes, may be broken into smaller parts with new labels of their own.

The portion of the card deck containing any particular type of bibliographic data can be of any length, and the cards in each code group are kept in proper order by a two digit card count number punched into columns 16 and 17 of the cards. Only one kind of information is coded into a particular card or group; the next kind of data always begins on a new card with its own code designation and card count sequence. The
end of a bibliographical group is indicated to the computer by a work
mark punched at the end of the data.

The system just described has worked satisfactorily in a test situation
where input data and the timing of update operations could be pre-de-
termined and controlled. It accomplishes the acquisition and cataloging
of material as parts of one continuous, integrated process rather than as
two separate functions. The parallel operation to be started in the fall
will determine whether or not the system is equally satisfactory in a rou-
tine, work-a-day situation, and will attempt to establish cost figures com-
paring the relative economies of the machine system and the manual
system it attempts to replace. If the new system proves to be economically
sound as well as mechanically feasible, we will hope to establish it as a
permanent part of our library operation.

New Edition!

THE INTERNATIONAL
WHO'S WHO
1964-65
28th edition

Basic biographic information about the world's most prominent men and women.
Relatively few of the nearly 120 countries represented have a national "who's
who".

Included in its 1,200 pages are people prominent in government, diplomacy,
finance, science, art, literature, religion, and many other fields. All biographies
are brought up to date yearly. This latest edition contains about a thousand
names not listed before.

$18.20

STECHELT-HAFNER, Inc.
The World's Leading International Booksellers
FOUNDED IN NEW YORK 1872
LONDON/ PARIS/ STUTTGART/ BOGOTA
31 East 10 Street/ New York 3, N. Y.
In discussing the dissemination of information, we can, for convenience sake, consider the problem from two points of view:

1. The degree or span of responsibility that the librarian has in providing special services to his clientele and,

2. The degree or span of bibliographic control that the librarian can afford to exercise on all of the materials received by the library.

The different libraries represent a very wide spectrum of responsibilities to the reader. Some are merely convenient storehouses with the most rudimentary finding tools for their collections, and the reader does all of his own searching. At the other extreme are libraries that try to provide every possible service in supplying information to their clientele. The services not only cover reading material, but sound recordings, films, pictures, and even hand tools of all kinds.

The direct personal service to users has in the past been confined, primarily, to the special libraries which are responsible only to a very restricted clientele and usually in a narrow subject area. The libraries which have much broader responsibilities, both in clientele and subject matter, have not been able to provide many special and individual services nor have they been able, bibliographically, to control much beyond the hard bound book. In most instances, they could not afford to process fully such publications as dissertations, reports, pamphlets, map collections, moving pictures, slides, art collections, and the like.

With the application of the computer to bibliographic processing, it suddenly became possible to provide specialized, direct services to library users and to do library processing very rapidly and on a mass scale. Mechanized processing, however, did not find too much favor with professional librarians. It was only a few rebels and people outside the profession who adopted the new techniques. Also these techniques were confined to special situations and to literary forms that were not normally processed by libraries, such as near-print materials, foreign and domestic journal articles, collections of photographs, statutes, convention papers, the preparation of book indexes, and so on.

There now began a series of tests, a few superficial studies, and a very loud debate as to the effectiveness of mechanized subject analysis. As is usual in such emotional situations, much more heat than light was shed on the situations. Now, however, as the initial temper tantrums subside, as experience is built up, and the more objective viewpoints come to the fore, there is a growing realization that these mechanized techniques,
when applied with discrimination, can be extremely effective and can extend the ability of the librarian to provide services which had been beyond his economic capabilities.

Although I am sure most of you are well acquainted with the computer-prepared dissemination tools, I would like briefly to describe them and the techniques for preparing them. Essentially, what the computer can do is recognize words, sort them, and print them. It does not know the meanings of these words—they are just symbols—but it can distinguish their physical differences. A computer, therefore, can prepare a concordance, can alphabetize, and otherwise arrange lists of words. And by a table-lookup technique, that is by checking a dictionary, an authority list, a thesaurus, a cross-reference list, or whatever you want to call it, the computer can bring synonyms together, provide cross-references, and also eliminate terms that are non-informative, redundant, or otherwise unwanted. With these capabilities, the computer can do what is usually referred to as “word” indexing. It cannot do “concept” indexing. It can do word indexing provided, of course, enough of the words which adequately describe the contents of the item to be processed are available in machine-readable form. We, therefore, have two problems: are there enough words present to describe the contents of the item? are the ideas or concepts in the document to be indexed obvious from the implications of the text (even though not expressed in the available words?) For example, an Atomic Energy Commission document is all about the chlorination of uranium. It is obvious from the text that this is one of several processes being investigated for the separation of uranium from its ore. However, the machine-readable portion of the document: the title, abstract, conclusions, etc., do not mention this. How is it possible, therefore, for the computer to provide the necessary subject tracing for Uranium-Ore beneficiaion or Uranium-Separation processes? Another example: an article describes how a certain drug caused a serious blood deficiency. Nowhere in the text are the terms “toxicity” or “side effects” used. Yet to retrieve this document, somewhere in the subject heading must appear “toxicity”, “side effects”, or something similar.

There are other text-processing problems, such as homonyms (lead and lead), order of terms (there is a difference between “A House of Cards” and “A Card House”), poetic titles or non-informative titles (“First Progress Report of the Radiation Laboratory”), and many more. These problems are usually solved easily by the human but are difficult, i.e. require very elaborate machine programs, or are impossible as yet for the computer.

In other words, straight word indexing cannot in every instance fully and properly index every document. Actually, various surveys, notably of legal literature and of scientific and technical articles, have shown that word indexing is very good. And although the computer does, on occasion, fail to index all relevant concepts, it often picks up useful terms that the human indexer missed, and it certainly is less prone to errors and is much more consistent than humans. In fact, in certain kinds of in-
d3I where the rules can be precisely stated, such as the development of linear notation for organic chemical structures, the computer does a much better job than the best human indexer.

Understanding these capabilities and limitations of the computer, some librarians are beginning to evolve new ideas on how to apply the computer to many of the obvious and routine processing chores, leaving the cataloger or indexer to revise this mechanical indexing and add those special elements which are required. This is sometimes referred to as cataloging by exception. Details of such techniques are the subject of other papers at this conference; I mention this very briefly only in order that we might have the right perspective on computerized dissemination techniques.

Since the computer can recognize words and can sort and print at very great speeds, H. P. Luhn reasoned that the words in the title of an article could be permuted and aligned in alphabetic sequence on a particular column or set out in the margin and printed out. In reality this is a form of catch title indexing and is similar to the “Schlagwort” catalogs which are still to be found in some of the older German university libraries.

Mr. Luhn was very modest in his proposal. He did not present his idea as a substitute for cataloging, but essentially as a technique for very quickly preparing scannable and searchable announcement bulletins. Mr. Luhn called the process Keyword-in-Context or KWIC indexing.

The KWIC programs, in addition to providing a subject approach through permuting the words of the title, also prepare author indexes, both personal and corporate, and, in some applications, other source information. Various types of KWIC indexes are shown in Figures 1-7.

After being tested experimentally, the KWIC program was first applied on a large scale by Chemical Abstracts to produce Chemical Titles. This publication was prepared so that the chemists and other researchers could get information about chemical publications, weeks and even months before they appeared in Chemical Abstracts. In 1963 Freeman and Dyson reported that the average monthly issue of Chemical Titles indexed some 2,800 articles and 5,900 authors and covered about 600 journals. The entire index was prepared by two keypunch operators, an editor, and a clerical assistant with the computer producing an average issue of 125 pages in about 4½ hours.

Since that time KWIC techniques have been applied so widely and so many new applications are being developed constantly, it is impossible to list them all. A few typical examples of KWIC indexes are: library accession lists, section indexes for procedure manuals, indexes for computer programs, special bibliographies, correspondence file indexes, indexes for papers of professional meetings, cumulative indexes for periodicals, indexes of statutes, indexes of technical photographs, lists of standard parts and manufactured products, tool lists, serials lists, and so on.

Details about KWIC indexing are given in IBM General Information Manual No. E20-8091, Keyword-in-Context (KWIC) Indexing, and in many technical articles of which a select few are listed in the bibliography.
accompanying this paper. Machine programs are available. The 7090 program developed for the Bell Laboratories Library can be obtained through SHARE; two 1401, a 704, and a 1620 program are all available from the IBM Program Information Department in White Plains, N.Y. Many individuals, companies, and schools have developed their own special versions of KWIC, and some of these programs will also be supplied to requesters.

The cost of producing a KWIC listing will vary greatly, depending on local conditions, availability of equipment, and salaries. In one instance which may be considered typical, costs were broken down as follows: a monthly index covering 100 journals for a total of 2,000 articles, averaging 4½ punchcards per article generated 9,000 cards. These were processed on an IBM 1401 in a little less than 3 hours, producing a Keyword-in-Context index, an author index, and a bibliographic list. Machine costs were $55 for card punch and verifier and $225 for the computer. Keypunch operator salary was $300 for a total of $580 or an average cost of 29 cents an article. This does not include, of course, the cost of printing multiple copies.

It is this incredibly low cost per item which really makes it possible for librarians to consider applying this processing technique to their holdings which they now cannot afford to catalog. The near-print material, the dissertations, art collections, maps, the various vertical file holdings, and a host of other special materials are all potential candidates for this type of inexpensive bibliographic processing. The only professional effort required is to provide each item with a good title. As a rule, most libraries do this already. Another potential benefit is that such indexes can be produced in multiple copies and thus made available to departmental libraries, branches, and to individuals.

In addition to producing library bulletins, accession lists, and indexes, computer techniques are also being used to provide direct personal service to individuals. With the vast increase in publication and the proliferation of journals, reports, and other separates, it is quite impossible for any individual to keep current in his various fields of interest. Nor is it practical for a librarian to screen all the inputs to his library and direct the pertinent material to the various individuals according to their subject interests. The best he can do is to set out all new acquisitions for a few days so that the library clientele can scan the material or to reproduce the title pages and tables of contents and distribute them.

The library user in browsing through the new acquisitions or checking the tables of contents is looking for clues to see if he should read the article or book. These clues are usually the words used in the title, the abstract, the opening paragraph, or the conclusions. There are, therefore, certain words which stimulate his interest. If he could list these words, it would be a simple matter for a computer to match them with the words which appear in titles, abstracts, opening paragraphs, and conclusions. If there is a sufficient match, then the article or book will probably be of interest to him. In a sense, the computer can browse through the new
literature and pick out the items of potential interest to an individual. This technique has been in use now over four years to provide a current awareness service to many individuals. Generally referred to as Selective Dissemination of Information or SDI, there are a number of variations of the system. As an example, let us look at one programmed for the IBM 1401 and which is available as program 1401-CR-DIX. A general description of SDI is given in IBM General Information Manual Ezo-8092, Selective Dissemination of Information, and in several references in the bibliography.

The two elements which are matched in the computer are the bibliographic record of the citation and the interest register or "profile" of the individual. The user of the system makes up a list of words which reflect his interests; this is his "profile." He may choose the words from a special list, or he may write a paragraph or two describing his interests, these to be indexed by an indexer, or he may just list words that occur to him. Sometimes he is guided by examining interest profiles of others engaged in the same type of work. Figure 8 is an actual interest profile of an individual working in the field of information retrieval.

There are two types of words on the list: exact terms and word roots. The root Librar will match on LIBRARY, LIBRARIES, LIBRARIANSHIP, LIBRARIAN, etc., whereas, the keyword Library will match only on LIBRARY. The reason the same root and full word are included is because library, the full word, is weighted as 3, but the root is weighted only as 1. The "hit level" of this profile is 03. This is a measure of the similarity that must exist between the profile and the document citation before the latter is considered a "hit." The mere occurrence of LIBRARY in a title or abstract will retrieve that document. However, if only LIBRARIAN appears in the citation, then other profile keywords must show up so that the weight sum of matched keywords and roots equals or exceeds 3.

Negative weights may also be assigned to keywords and roots. When these occur, then the total weight of the matched terms is reduced. Examples of negative weights are shown on the profile assigned to three journals. The requester in this instance reads these journals regularly, and he does not want to receive any notices of articles which have appeared in American Documentation, Library Journal, or Special Libraries. The hyphen in Information-Retrieval is present to tell the computer that this word pair must exist, otherwise the match will be terminated at the first space which is immediately after Information.

The preparation of the citation can be done variously. It can be the actual cataloging or indexing done by the library. In fact, this multiple use of library inputs is the most economical. Actually, since the SDI programs have so far in most instances been operated outside of libraries and have included literature not normally processed by libraries (news articles, management reports, patents, patent applications, computer programs, sales brochures, etc.), the inputs have had to be prepared separately.
After a series of tests, it was determined that the citations could be prepared by clerical help—professional indexers or catalogers or other technically trained personnel were not needed. In fact, since these professional personnel were often critical in their selection, they tended to omit useful material and were often slower in preparing inputs than the clerical personnel. The only instructions given the clerical personnel, who were usually secretaries or keypunch operators, were to pick up author, title, journal citation, and, if an abstract were present, to copy it. If an abstract were not present, the clerks were to pick up the first few sentences of the introduction and/or conclusion and anything that looked “real technical,” even if they did not know what it meant. The number of words was limited to a certain maximum, usually 20 lines of 60 characters each. Actually, in the specific applications there were also special instructions such as “be sure to include all references to equipment, to company names and locations, and industry references. Avoid repetitions. If the abstract is too long, remove modifying words, such as adjectives and adverbs, and other unimportant phrases and clauses.” Such instructions, of course, varied somewhat with the subject areas being covered.

When the computer matched the profile keywords with the abstract, it scanned the latter for all terms. That is, it looked at the whole text. Since a text might not have significant keywords, a simple form of indexing was applied. A few so-called “directed terms” were manually assigned by the abstractors if they were not present in the portion of the article selected for SDI. An article on any aspect of librarianship had to have the word LIBRARY present. If it were absent, the abstractor added it. The “directed terms” were broken down by industry and application and in general did not exceed 5 or 6 terms per category for about 30-40 categories.

More important than the actual preparation of the references are their selection. Since many people read their core journals and do not regularly get to see the publications peripheral to their field, it is important that many publication sources, normally considered unimportant, be included. In the profile example shown above, the requester did not want to receive notices extracted from Library Journal, but he did want to be advised of any library or information retrieval article that might appear in non-library journals. Some of the early SDI systems were somewhat less than successful because they concentrated on the “obvious” or “important” journals, and the only information they brought their clientele was already known to the users.

Aside from providing a good and extensive selection of inputs, the other major problem has been the cost of SDI. One rather efficient SDI system serving some 2,000 users and processing between 800 and 900 documents a month sent the average user 110 notices a month at a cost of $7 per month per user. SDI represents a new additional service with new additional costs; does not replace an existing library service and therefore represents no displacement savings. This additional cost must be justified. I need not tell librarians how difficult it is to justify, on a real
and of these 800-900 were entered into the system. The "average" user received notification of 8–9 of these articles or almost 1 per cent of the total which produced some 5 notifications per person each day. About two-thirds of the reply notices indicated that the item was "of interest", while about one-third were marked "no interest." The noise level was about 34 per cent which seemed to be quite tolerable, since it takes only about one minute to read an SDI abstract and punch out the reply card. As the system developed, the profiles were modified, on an average, about 5 times during the first year. This increased the "of interest" replies to 86 per cent and decreased the "no interest" responses to 14 per cent.

Just when the research or even public library will operate a current awareness program such as SDI cannot be foretold. However, once all library inputs do receive some computer processing, and this seems to be
inevitable for all but the smallest libraries, then an SDI service will have
be provided. Once the bibliographic information is machine scannable,
the SDI outputs cost practically nothing, for essentially these are only
the reproduction or printer costs. The dissemination costs will vary with
the situation. In any event, the value of the service and the demands for
it will, I am sure, force the librarian to become a true disseminator of
information and not, as he is so often maligned, just a custodian.

BIBLIOGRAPHY

Brandenberg, W., et al. Selective Dissemination of Information SDI 2 System. York-
Freeman, R., and Dyson, D. M. “Development and Production of Chemical Titles . . .”
Kennedy, R. A. “Library Applications of Permutation Indexing.” Journal of Chemical Docu-
Kraft, D. H. “A Comparison of Keyword-in-Context (KWIC) Indexing of Titles with a
Subject Heading Classification System.” American Documentation, 15:48-52. 1964.
Lewis, R. F. “KWIC . . . is it Quick?” Bulletin of the Medical Library Association,
Luhn, H. P. “A Business Intelligence System.” IBM Journal of Research and Develop-
—- “Keyword-in-Context Index for Technical Literature (KWIC Index).” Yorktown
Heights, N.Y., IBM Advanced Systems Development Division, 1959.
—- Selective Dissemination of New Scientific Information with the Aid of Electronic
Processing Equipment. Yorktown Heights, IBM Advanced Systems Development
Division, 1959.
Montgomery, C., and Swanson, D. “Machine-Like Indexing by People.” American Docu-
Resnick, A. Comparative Effect of Different Education Levels on Indexing in a Selecti-
ve Dissemination System. Yorktown Heights, IBM Advanced Systems Development Di-
vision. Technical memorandum 17-092.
Ruhl, M. J. “Chemical Documents and Their Titles: Human Concept Indexing vs.

• 80 •

Library Resources & Technical Services
KEYWORD INDEX

The page contains a list of keywords for indexing. The keywords are grouped under various categories such as 'Languages', 'Distance', 'Pharmacology', etc. Each keyword is followed by a list of related terms or phrases. The page appears to be part of a journal or academic publication, and the text is formatted in a standard academic style.
FIGURE 2

Library Resources & Technical Services
Figure 4
<table>
<thead>
<tr>
<th>LOC.</th>
<th>DEPT.</th>
<th>NAME</th>
<th>LOCATION</th>
<th>SECURITY</th>
<th>HIT LEVEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>983</td>
<td>597</td>
<td>J. Smith</td>
<td>San Jose</td>
<td>A</td>
<td>03</td>
</tr>
</tbody>
</table>

**S. D. I. PROFILE**

**KEYWORD TYPES**
- MAY WORD
- MAY ROOT
- WEIGHT IS 1 UNLESS SHOWN:

<table>
<thead>
<tr>
<th>TYPE</th>
<th>WT.</th>
<th>PROFILE KEYWORD</th>
<th>WT.</th>
<th>PROFILE KEYWORD</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>2</td>
<td>ABSTRACT</td>
<td>3</td>
<td>LIBRARY</td>
</tr>
<tr>
<td></td>
<td></td>
<td>APERTURE</td>
<td></td>
<td>KIWC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CIRCULATION</td>
<td>%</td>
<td>MICRO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CLASSIFICATION</td>
<td></td>
<td>MICROFILM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CODE</td>
<td></td>
<td>OPTICAL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>COPYRIGHT</td>
<td></td>
<td>READER</td>
</tr>
<tr>
<td>%</td>
<td>2</td>
<td>DESCRIPTOR</td>
<td>2</td>
<td>RETRIEV</td>
</tr>
<tr>
<td>%</td>
<td></td>
<td>DISSEMINAT</td>
<td>%</td>
<td>SCAN</td>
</tr>
<tr>
<td>%</td>
<td></td>
<td>DOCUMENT</td>
<td>%</td>
<td>SIX</td>
</tr>
<tr>
<td>%</td>
<td></td>
<td>IMAGE</td>
<td>%</td>
<td>SEARCH</td>
</tr>
<tr>
<td>%</td>
<td></td>
<td>INDEX</td>
<td>%</td>
<td>STOR</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>INFORMATION-RETRIEVAL</td>
<td></td>
<td>TECHNICAL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LIBRARY</td>
<td>%</td>
<td>UNITERM</td>
</tr>
</tbody>
</table>

**Figure 8**

*Volume 9, Number 1, Winter 1965*
A VARIATION OF THE KWIC INDEXING TECHNIQUE

PHASE (CONTINUATION)
1401 LESS / LEAST-COST ESTIMATING AND SCHEDULING / CARD-SCHEDULING PHASE ONLY / REVISED/
1402 LESS / LEAST-COST ESTIMATING AND SCHEDULING / SCHEDULING PORTION / CARD/
1402 LESS / TAPE LEAST-COST ESTIMATING AND SCHEDULING / SCHEDULING PORTION / TAPE/

PRODUCTION
0650 LINEAR DECISION RULE FOR PRODUCTION AND EMPLOYMENT SCHEDULING

PROFIT
SCHEDULING WITH ARBITRARY PROFIT FUNCTIONS

PROGRAM
CLASS SCHEDULING PROGRAM FOR THE 707A AND 1401
GENERAL INPUT/OUTPUT SCHEDULING PROGRAM

REVISED
1401 LESS / LEAST-COST ESTIMATING AND SCHEDULING / CARD-SCHEDULING PHASE ONLY / REVISED /
1402 LESS / LEAST-COST ESTIMATING AND SCHEDULING / SCHEDULING PORTION / CARD/

RULE
0650 LINEAR DECISION RULE FOR PRODUCTION AND EMPLOYMENT SCHEDULING

SCHEDULING
SCHEDULING WITH ARBITRARY PROFIT FUNCTIONS

CLASS SCHEDULING PROGRAM FOR THE 707A AND 1401
STUDENT SCHEDULING TECHNIQUE USING DEFINED INCREMENTS
0650 MAN SCHEDULING

CRITICAL PATH SCHEDULING / CARD/
GENERAL INPUT/OUTPUT SCHEDULING PROGRAM

MISS LESS MANAGEMENT INFORMATION SCHEDULING / CARD/
MISS LESS MANAGEMENT INFORMATION SCHEDULING / TAPE/
1620 LESS / LEAST-COST ESTIMATING AND SCHEDULING / SCHEDULING PORTION / CARD/
1641 LESS / LEAST-COST ESTIMATING AND SCHEDULING / CARD-SCHEDULING ONLY / TAPE /
1620 LESS / TAPE LEAST-COST ESTIMATING AND SCHEDULING / SCHEDULING PORTION / TAPE /
1620 LESS / LEAST-COST ESTIMATING AND SCHEDULING / SCHEDULING PORTION / CARD/
1620 LESS / LEAST-COST ESTIMATING AND SCHEDULING / SCHEDULING PORTION / TAPE/
0650 MAN SCHEDULING

**Figure 7**

Volume 9, Number 1, Winter 1965 • 87 •
This article describes the reasons for an EDP installation's profitability status and how they are evaluated in a detailed survey of user's top management.

KEYWORDS: Profit, Loss, EDP, Data Processing, Documentation, Information Retrieval, Top Management, Machine Utilization

pp. 28-32


RAND DEE

123 PEOR

123456

SHEEHAN ROBERT

IT'S A NEW KIND OF FORD MOTOR CO.

FORTUNE FEBRUARY 1962


WHAT IS THE SIGNIFICANCE OF THIS DRUMFIRE OF NEWS, WHO ARE THE MEN BEHIND THE COMPANY'S ACHIEVEMENTS, WHAT DO THEY PROMISE FOR FORDS FUTURE, FORTUNE HERE PROBES THE REASONING THAT LED TO EACH MOVE AND ASSESSES ITS LIKELY CONTRIBUTION TO THE MATURE, DIVERSIFIED, AND AGGRESSIVE CORPORATION THAT HENRY FORD II IS SHAPING.

Volume 9, Number 1, Winter 1965

89
The Library of Congress Project

GILBERT W. KING
Itek Corporation
Lexington, Massachusetts

IN APRIL 1961 the Council on Library Resources, Inc., provided a $100,000 grant to an agency of the U. S. Government, the Library of Congress, to examine the feasibility of automation of the large research libraries such as itself. A study group\(^1\) was assembled and investigated the problems existing in the Library of Congress and other libraries, the possible technical solutions likely to be available, and the present state of the art in operating a system from the point of view of programming and organization.

The results of this study were published\(^2\) at the end of 1963. The time taken to complete this study reflects the care with which issues were examined, especially in the light of rapid advances in the pertinent technology in those years. A great many facets were investigated in more detail than might appear in the published report. Some topics were elaborated at the Airley Conference in June 1963,\(^3\) especially in the field of communications in the type of system proposed, equipment for which had been rapidly developing as the report was being written.

In addition to all of this material which was published, many specific pieces of the puzzle were examined which will come to light as, hopefully, the project advances.

The principal features of the present plight of large libraries is, of course, the large and growing volume of books and serials, which causes frustration to librarians, with their chronic lack of funds, in control, and to the patron of the library in his endeavors to be informed. The library problem is worth serious national attention because, both traditionally and currently, collections of written material reflect the status and nurture the future of civilizations. The Library of Congress alone represents a great national resource; a mine of information which could be retrieved to advance our culture and save very large sums of money in national projects. This is not to say that the report was deeply concerned with methodology of information retrieval being studied in many places at various levels of sophistication. It was directed at a purely practical means of controlling and finding bibliothecal material by more or less known interfaces with

---

1. The members of the group were G. W. King, Chairman; H. P. Edmundson; M. M. Flood; M. Kochen; R. L. Libby; D. R. Swanson; A. Wylly. The group were greatly assisted by H. J. Dubester; Barbara Markenson; D. F. Loeb; H. T. Spiro; A. D. Kotin.


traditional libraries, although the system concept was designed to permit, without radical equipment alterations, any and all new retrieval methods if and when they appear practical for very large collections.

A solution of the libraries' problem was found and outlined. Many specific details were considered, but not described in the report because in an era of rapidly developing technology, many improvements can be anticipated by the time of substantial funding of the program.

The principal equipment items that are needed for an automated library are mass digital memories, mass graphical memories, a communication network for digital and graphical information, consoles for man-system interaction, and switching equipment to monitor data flow. There is no point in basing a study of this sort solely on existing equipment now on the shelf. Attention should be focussed on what is likely to be available at the appropriate points in time scheduled for a properly funded project. However, although an item would be feasible if a certain amount of development resources were devoted to it, there is some question whether such development will occur. Libraries need equipments devoted to their kinds of data—typically languages, not numbers. The library community is not an attractive market, in its present state of funding, to the equipment manufacturers. This is why it is so important that a large-scale project, such as the automation of the Library of Congress, be adequately funded, for the whole library community to be able to bring itself up to the technological well being attendant to other fields.

The storage of the massive amounts of lexical material in libraries in digital (machine readable) form such as on the card catalogs, not to mention citations, abstracts, tables of contents or full text, requires memories of a size and organization not being met elsewhere in the data-processing industry. Nevertheless, suitable types of memories have been made for special projects, and it seems likely they could be further developed from the libraries' point of view. It is very likely that a digital memory with a capacity of a trillion bits (thirty billion words) will be commercially available before fiscal year 1966.

The storage of large amounts of materials, such as newspapers and serials, as photographic images of their pages is mandatory in an efficient library. Here again there is not available the equipment really exploiting modern technology for the needs of large libraries. For example, reduction of 200 to 400 to one are quite practical for photographs of newspapers, this reduction making phenomenal improvement in space requirement and speed of access.

Methods of communication of the stored digital or graphical information to the requestor is solved, although costs for real time communication over large distances are still high for libraries in their present status in the community.

An automatic library requires terminal sets, or consoles, that is, means for the user to obtain ready access to the contents of the memories. These basically consist of a display such as a television tube, with, however, considerably higher quality, to allow textual material to be read easily.
Good quality and a large font of all the letters and symbols used in literature are essential if substantial use of the library is to be encouraged. The same equipment should be able to display photographs of printed pages and pictures retrieved from the mass graphical file. The console is the site of operation of the user, so it must have certain convenience features. For example, there has to be a simple keyboard by which an untrained user can communicate his requests to the system. It should have a local storage device on which the user can build up a file of the pieces of information he is retrieving, so that he can go back and forth in referring to it. It should have means of giving him low-cost hard copy of selected material he has been shown and temporarily stored.

The console is perhaps the most troublesome piece of equipment visualized for the library system. For, although a console meeting the requirements could be built, its cost is likely to be quite high, unless the market, i.e., number of units, is quite substantial. It is to be hoped that the Library of Congress project will provide such consoles, acceptable to other systems in the library community.

In addition to equipment, an automated library system needs organization and methodology. The novel features of this are that the system itself has to help the user in locating the information he is after. It is not enough to let him ask simple questions, such as (a display of) the catalog card of a book by “Smith.” He would be, in general, flooded by too much material. The program of the system must guide him, with questions and suggestions displayed at his console, in the way a reference librarian discourses with him. Normalized wording of his questions by an automatic thesaurus and even automatic normalization of syntax are distinct possibilities. In this area there is the opportunity of a vast improvement in the interaction of the user with a library.

Related to the development of the automation helping in the use of its contents is the means of loading the memories with the large amount of material generated by a library in its bibliographic control and description. There is the basic difficulty of transferring all the material now printed on cards, reference books, etc., into machine-storable (digital) form, but there is the serious matter of not losing information in so doing—in not losing the typographical and format clues which now aid the user in scanning printed material—aids which must be retained for the display at the console. Here again various studies and prospective solutions were studied as background for the report. These need a concerted effort to become practical and economical, one in which the library community as a whole should become involved.

In considering file conversion, it is important to realize that imposition of standardization as to coding, symbols, or format used is not at all necessary. On the contrary, it is highly desirable, and feasible in the system implied in the report, that essentially any type of organization or indexing can be introduced into the central system. Internally, conversions from one structure to another can be made.

Perhaps the most important feature of the system proposed, in general-
ities in the report, but in detail as background for the report, is its emphasis on a structure which does not impose any constraints on present or future cataloging, classification, or indexing schemes. Thus, all of the bibliographic work which has been done in the past will be saved and in the long run incorporated for future users. In addition, new services, such as citations and bibliographies, generated by the use of the system, will be preserved, and make automated libraries more useful by far than their manual counterparts.
The MEDLARS Project at the National Library of Medicine

CHARLES J. AUSTIN, Chief
Data Processing Division
National Library of Medicine
Bethesda, Maryland

Introduction

MEDLARS (Medical Literature Analysis and Retrieval System) is a computer-based information storage and retrieval system which recently became operational at the National Library of Medicine. It is the first phase of an aggressive program of research and development in documentation aimed at the improvement of biomedical communication and the use of modern technology in library operations.

MEDLARS joins the intellectual talents of professional literature analysts to the tremendous clerical processing capabilities of an electronic computer in a unique man-machine relationship. It is a bibliographic system aimed at control of a large segment of the world's biomedical literature and the rapid dissemination of this data to those engaged in medical research and practice.

The objectives of MEDLARS can best be defined by describing its three main output products. A more detailed explanation of the techniques used in obtaining these outputs is given later.

The first MEDLARS product is the Index Medicus, a monthly subject and author index to some 2,500 biomedical journals published in all parts of the world. The computer improves the index in various ways. The capacity of the machine for storing and manipulating a large volume of data makes expanded coverage of the literature possible, and the timeliness of the index is improved by reduction in the throughput time required to prepare the monthly editions for printing. The second product of MEDLARS is the recurring bibliography—a current-awareness list of citations in specialized medical subject areas. The system will produce up to fifty different recurring bibliographies compiled at regular intervals from data in the computer files. These compilations will be prepared in photo-master form, and will be printed and distributed by organizations working in the specialty fields. Demand bibliographies represent the third and final major product of the system. Rapid searching of the computer's store of data can provide answers to complex bibliographic requests which cannot be effectively handled by reference to a printed index or catalog.

MEDLARS was developed under contract by the General Electric Company's Information System Operation in a three-phase effort. Phase

Library Resources & Technical Services
I, a preliminary study and design phase, lasted from July 1961 to January 1962. This phase included development of a basic set of specifications for equipment, programs, and personnel required to implement the MEIDLARS objectives. Phase II, detailed design, began in January 1962 and included ordering of equipment, writing and testing of computer programs, and development of detailed procedures to be followed by personnel operating the system. Phase III, system testing and implementation, overlapped Phase II and included installation of equipment, file conversion, detailed testing of all parts of the system, and a period of preliminary operation. Phase III will end this summer.

System Loads

A brief description of system loads will help point out the need for mechanization. The current annual volume of approximately 150,000 papers indexed is expected to increase to 250,000 by 1969. In addition, at some future date about 10,000 citations of monographs will be entered into the system annually. This represents an annual input load growing from 62 million characters currently to 100 million in five years. Output loads will be proportionately larger. The system will produce the monthly and annual Index Medicus, up to fifty recurring bibliographies of varying periodicities, and a number of demand requests expected to grow from an initial load of ten per day to ninety per day after five years. The output printing load will grow from about 290 million characters in 1964 to 590 million in 1969.

Data Processing Equipment

In order to meet the speed and volume requirements of the system, the following items of automatic data processing equipment are employed: thirteen punched-paper-tape typewriters for conversion of source data to machine-readable form; a Minneapolis-Honeywell 800 digital computer for editing, sorting, compressing, merging, and formatting data for subsequent printing; an optical printer called GRACE (Graphic Arts Composing Equipment) manufactured by the Photon Corporation and used to convert computer output into high-quality photo-copy; and an automatic film processor for developing the film from GRACE, thus producing a photomaster which can be used directly for burning of printing plates.

With this description of the objectives of MEIDLARS, system loads, and data processing equipment as background, the actual functioning of the system can be described in more detail. Any data processing system normally consists of three parts—preparation of input data, manipulation of this data automatically, and output preparation of the final products of the system.

Input Preparation

Journals received by the Library at the serial record checking area are forwarded to the Index Unit where they are distributed to the index-
ing staff for the translation of foreign titles and subject classification of articles with appropriate descriptors from the Library's controlled list of terms called MeSH (Medical Subject Headings).

To transform this basic data into mechanical form, a punched-tape typewriter (Friden Flexowriter) is used. This machine produces hard copy as well as punched tape when the keys are depressed. The input typist integrates information from a data sheet prepared by the indexer with information from the journal article itself. In addition, codes are added so that the computer can recognize individual elements of the record. Hard copy from the typists is sight-verified by a proofreader who notes necessary corrections on a work sheet. This work sheet is returned to the typist who made the error, and a correction tape is prepared. Both the original input tape and the correction tape are later matched by the computer and necessary substitutions are made. The basic unit record prepared by this process consists of all of the standard elements of a bibliographic citation, plus the subject headings assigned by the indexers.

Search requests for bibliographic information are prepared by a team of search specialists and recorded on a Search Formulation data sheet. Whereas indexers describe articles by selecting the appropriate subject headings, searchers prepare a set of elements which are used in logical combination for retrieving citations already indexed and in the computer files. The request may include as many as 100 search elements, including such items as subject headings, author names, journal titles, language designators, year of publication, and several others. The Search Formulation data sheet is punched into paper tape and proofread in a similar manner to the Indexer data sheet.

**Computer Processing**

Seven programming modules have been designed to fulfill MEDLARS system requirements. A module is a group of computer programs all of which are related in performing one of the system's major functions.

After conversion to punched-paper-tape, indexed citations are entered into the *Input Processing Module*, which edits the data extensively, converts English-language elements such as subject headings into their coded equivalents, and builds the two main computer files on magnetic tape. The Compressed Citation File (CCF) is a highly coded, time-sequential store used for high-speed searching to retrieve demand bibliographies. The Processed Citation File (PCF) contains unit records in expanded print-line format which have been selected and tagged by this module for various recurring bibliographies. Once the two basic data files have been created by the Input Processing Module, the computer subsystem is split into two independent parts—one for retrieval of demand bibliographies and the other for the composition of *Index Medicus* and recurring bibliographies.

The *Demand Search Module* matches search requests entered into the computer by punched-tape against unit records on the CCF. It is important to note that several requests are entered in a large batch (perhaps...
and all searches are performed simultaneously in the computer's central processor. Output from this module consists of a magnetic tape file of retrieved citations plus a printed report of the number of citations retrieved for each request.

The Report Generator Module prints bibliographies of citations retrieved by the Demand Search Module. Sequence and format of these bibliographies are variable and are specified by the search specialist, who first reviews the report telling how many citations were retrieved and uses this information as a basis for determining output format. The specialist also determines whether the final product is to be printed on 3” × 5” cards, 8½” × 11” paper, or photo-type-setter film. The computer takes about five minutes to search one complete reel of magnetic tape containing roughly 25,000 citations. The file currently holds about 150,000 references entered since 1963, and a search of this file for the answers to several requests takes approximately 30 minutes.

Composition of recurring bibliographies is the function of the Output Processing Module. Each working day, punched cards are entered into the computer to specify which recurring bibliographies are to be printed that day. The Output Processing Module selects citations from the PCF, using information from these cards. The module then sorts the citations into the proper sequence, formats the data, and writes it onto magnetic tape for later use by the photo-composer.

There are three utility modules in the computer subsystem. The File Maintenance Module is used to update the two main data files (CCF and PCF) with additions, deletions, and changes. The MeSH Generator Module performs a similar function with the master file of medical subject headings. The Statistical Module produces reports on the frequency of use of the system, such as the number of times each subject heading is used, the number of articles indexed by language, etc.

The four main operating modules (Input, Demand Search, Report Generator, Output) are run on the computer once a day, but the utility modules are run at less frequent intervals and only when needed.

**Output Preparation**

The unique aspect of MEDLARS is its special output device, GRACE. This machine prints at the rate of 300 characters per second from a font of 226 characters (including upper and lower case). Data is transferred directly from the computer via magnetic tape and composed onto positive photographic film or paper.

GRACE will be used primarily for compiling Index Medicus and recurring bibliographies. The computer also has an on-line printer which is used for printing demand bibliographies prepared by the retrieval programs. Although this printer is about six times as fast as GRACE, it has a limited type font of 56 characters, upper-case only. A delay in delivery of GRACE one-year beyond the originally scheduled date has forced the Library to prepare the Index Medicus issues this year by the computer's on-line printer. This less-than-desirable format will be replaced by

*Volume 9, Number 1, Winter 1965*  
97
GRACE composition in the very near future, and the *Cumulated Index Medicus* for 1964 will be printed by GRACE.

Magnetic tapes produced by the computer are mounted on the composer which prints complete pages on film in accordance with instructions received from the tape. Upon exposure of 100 feet of film (the equivalent of about 100 printed pages), the film is removed and taken to an automatic film processor for developing. The processed film is later inspected, cut into sheets, packed, and delivered to a printer for platemaking, printing, and binding. GRACE eliminates the expensive and time-consuming job of typesetting required in many publication systems.

*Data Processing at NLM*

Data processing is an integral part of MEDLARS since the very concept of the system is the rapid storage and retrieval of pertinent references from a large information file. For this reason the Data Processing Section was established in the Bibliographic Services Division of the Library in August, 1962. The primary responsibility of this section was to work with the MEDLARS contractor in installation of the system, and prepare to take over upon completion of the contract.

On March 18, 1964, a new Data Processing Division was formed at NLM. This division was charged not only with carrying on the operation of MEDLARS, but also with initiating new systems studies involving the potential use of data processing equipment and techniques in other areas of the Library.

A Systems Analyst joined the Library staff in January of 1964 to begin a systems study of technical processing. It is hoped that this study will lead to improvements in the selection, acquisition, and cataloging of new materials. Other possible future data processing projects include automation of the card catalog itself, investigation of graphic image storage and retrieval and its relationship to MEDLARS, possible mechanization of the serial record, and statistical analysis of the utilization of MEDLARS.

The Library also plans to decentralize the MEDLARS search capabilities by distributing magnetic tapes to university and other medical research centers with adequate library and computer facilities. A pilot study will begin later this year, and will involve re-programming and tape conversion for MEDLARS searching on computer equipment at a large university medical center.

*Impact of MEDLARS on Other Libraries*

MEDLARS will result in much broader dissemination of medical bibliographic data through its new and expanded indexing services. This, in turn, will impose great demands for increasing the size and scope of the collections of local medical libraries to provide the journals cited in a MEDLARS bibliography. The system will provide every local librarian

98

*Library Resources & Technical Services*
with a modernized version of a traditional reference tool for rapid searching of a large store of information, since he will be able to request demand searches from NLM. At the same time it will impose a serious obligation on the librarian to become familiar with the fundamentals of computers in order to understand their capabilities and limitations. Finally, in the future some libraries may find it possible to acquire a magnetic-tape edition of the Index Medicus for use on their own data processing equipment.

Summary

MEDLARS is a computer-based information retrieval system aimed at bibliographic control of the world's medical literature. Medical journals received by the Library are indexed by a staff of professional indexers, punched into paper tape, and permanently stored on magnetic tape. Data from the magnetic tape files are manipulated by the computer in order to produce the monthly and annual Index Medicus, other recurring bibliographies in medical specialty fields, and demand bibliographies in response to subject oriented questions. The Data Processing Division at NLM is heavily involved in the operation of MEDLARS and in systems studies aimed at the improvement of biomedical communications and modernization of Library operations.

REFERENCES

The following documents were used extensively in the preparation of this paper:


CATALOG CODE REVISION SCHEDULE

In LRTS Fall 1964, p. 365, I stated that the catalog code revision schedule calls for a completed manuscript to be presented to the CCS Executive Committee "in the summer of 1966, instead of (as originally planned) 1965."

These dates are not correct.

Actually the schedule calls for a completed manuscript by the summer of 1965, instead of (as originally planned) 1964.—Paul S. Dunkin

Volume 9, Number 1, Winter 1965 • 99 •
The Machine and the Librarian

RAFHL H. PARKER, Librarian
University of Missouri Library
Columbia, Missouri

THERE CAN BE no question who will be the eventual master. When we hear expressions of fear of the machine, what is really meant is that we fear other men's use of it. For this, history provides ample justification. The process of civilization may be characterized as the growing use of mechanical devices and the struggles of mankind to direct that use. Without mechanical assistance man's energies were consumed in providing subsistence; there was no leisure for creative thought. But once a slight margin above subsistence was obtained, it became possible to utilize the resulting leisure either for social good or selfish gain. Even slavery and serfdom can exist only in a society where the means of production exceeds the demands of consumption.

This difference between productive capacity and consumer demand has grown in a very tight spiral. As capacity has increased, so has the standard of living. That part of the difference which is turned back into machines of production serves to widen the spiral. What we know as the Industrial Revolution of two centuries ago ushered in the long period of providing mechanical power to replace human power. Steam was the key in the process which may be characterized by the power driven loom, power saws, and steam boats. We need not elaborate on the significance of these and similar inventions. Steam was followed by the internal combustion engine, then by electricity, and now by nuclear fission and nuclear fusion. The machines which, in the nineteenth century, created unemployment and social upheaval are today taken for granted, and account in large measure for the change in standards of living.

The machines which today are creating unemployment and exerting pressures on the social structure are part of what we might call the Paper Revolution. Its background includes the invention of the linotype, the typewriter, the rotary press, offset lithography, and teletype. More significant, however, are the machines which have descended from a different line. Charles Babbage and his calculating machine of the middle nineteenth century, Hollerith and his concept of punched cards, which in turn was borrowed from the textile industry of the eighteenth century, were fathers of the modern electronic computer. This product of the second half of the twentieth century is more than just a calculating machine. Its capacities for mathematical computation can also be used for the manipulation of non-quantitative data as well, and even for the control of other machines. Thus the machine, which in the eighteenth century replaced man as a source of power and converted labor to the
supervision of machines, is now followed by a machine which even replaces man as a supervisor of a machine.

Of what does automation consist? Mechanization, yes; but two more things are needed: the concept of an integrated system, and what is commonly called "feed-back." There is, of course, automation without computers, but their enormous capacity for feed-back vastly expands the potential for sophisticated systems.

Automation is based on what we might call machine learning, the apparent ability of a machine to learn, a capability such that a decision once made can be remade by the machine without human intervention. Because much human activity is at this intellectual level—masquerading as thoughtful activity—machines have erroneously been endowed with the power of thinking. It is probably unnecessary to repeat that machines do not think, that they simply do rapidly and accurately tasks of routine repetitive nature.

Although there is awe and admiration for machines, there is also snobbery about things which they produce. The Paris gown, handblown glass where an imperfection is a mark of authenticity are examples of this snobbery. The Gutenberg Bible was once a cheap machine-made imitation of a manuscript! It is possible to create cheap machine goods; it is not now economically feasible to use hand labor for this purpose. Paper cups, which are a socially unacceptable substitute for porcelain and fine glass; paper napkins, which are becoming an acceptable substitute for fine linen; the toy balloon and the party whistle can exist only in a system of mass production. But factory production is not necessarily synonymous with poor quality. The electric refrigerator is a marvelous example of beauty, precision, and durability; no handmade product could equal it.

Automated systems do create differences in people, however. The artisan's sense of personal pride in his creation, whether it be a pair of shoes or a page of manuscript, has little place in a system relying on statistical methods of quality control. Until our society has developed a satisfactory substitute for this personal pride of workmanship there will be complaints of employees who do not care. This is an area of challenge for the professional of today and of tomorrow—the professional in industrial relations, the professional educator, and even the professional librarian.

What can machines do in libraries, and what of the librarians themselves? In spite of the predictions of some, computers will perform dull repetitive jobs in libraries in the same way that they are performing dull repetitive jobs in business and in scientific research. When mixed with ingenuity of human beings, the results may well be marvelous.

The books of which libraries are composed are rather mobile. Who hasn't heard allusions to them walking off or to the ease with which they may be borrowed, secreted, or purloined by the willful? Records, on the other hand, are viscous. The bibliographic organization of a library is far more durable than its books or the building which houses them. Li-
brarians may not have analyzed the problem in this light, but this viscos-
ity is primarily responsible for their being called narrow-minded, unre-
sponsive to the needs of the user, and even obstructive to culture and enlightenment. How often the librarian is unable to fulfill the wishes of the user because of the costs of recording the change in location of the book or of maintaining adequate control of its availability!

The newer technology which is now emerging will free the record; it will be possible to reorganize collections for short term use as well as for long term repose. The distinction between the circulation file, the catalog, bibliographies, and indexes will largely disappear.

But there is a danger that the machine will become temporarily the nominal master. Machines have limitations in their capabilities, and the people who operate them tend to let the machine dictate what and how things are done. When this happens, it is a defeat for human ingenuity. Of course we must work within the confines of existing technology, but it need not be a strait jacket.

Printing in more than one color was economically not feasible in the sixteenth century. Type designers overcame the limitations by development of variations in style and size of type. The expert photographer in black and white can use techniques to compensate for the loss of color. The illusion of depth can be created in a flat representation. We must not cease to strive for improvement in the technology with which we must live. Just because most data processing machines have type fonts limited to 39 or 47 or 63 characters is no reason why bibliographies or catalogs in libraries must be geared to them. If the use of machines means lowering of quality, assuming that our standards of quality have real validity, then we must avoid using the machine.

The fact that it is not possible to program a computer to file by the ALA rules is not a valid reason to settle for records arranged in a sequence which the computer is capable of creating. Although it may not be possible for a computer to figure out how to put an entry into the desired sequence, it can be taught to put it there after it has been shown the first time. And who will decide where it should be put in the first place? The Librarian, of course; but he need make the decision only once.

Let us repeat, the machine cannot do intellectual tasks, and presumably the professional work of the librarian is intellectual. But the machine can do repetitive ones and do them far more effectively than can the librarian. The effectiveness of the professional can be multiplied by storing the results of intellectual activity and retrieving them by machine when they are needed. This is what we do, without a machine, every time we consult an index or even the text of a book. But the machine can extend the fruitfulness of the original creator by searching and comparing and printing out only those items which match—or do not match—the criteria established.

In short, automation of records in libraries will free librarians, whether they wish it or not, to become truly professional. Their jobs may
be far different, but they will still be needed. The library of the future will also need technicians to operate the machines, and for some time to come, it will need people to shelve books. The catalog typist, the file clerk, and the librarian who is only a paper shiffer will largely disappear.

These prospects may create some sense of fear, but they will have little impact on the change which is taking place in our society. For one thing, libraries are too insignificant a part of it; furthermore, most libraries have too much difficulty in obtaining people to do these tasks now for anyone to regret the passing of the jobs. In the long run, the capacity of the library for rendering service and the demands for that service will be so vastly increased that the number of people still working in libraries will continue to grow.

THE DECIMAL CLASSIFICATION EDITORIAL POLICY COMMITTEE: ANNUAL REPORT 1963/64

The chief business of the Committee at its meeting in October 1963 was the approval of final details of proposed changes in Edition 17. Perhaps the most important item was the decision to adopt a new method of indicating geographic arrangement through the presentation of the full list of geographic headings in standard subdivision .09 rather than referring from this point to the history schedules in the 900's. We believe that this procedure will make the application of the numbers easier and also that it will permit the logical separation of geographic concepts from the present limitations of the historical approach. Among other actions, the Committee also reversed a previous decision to print both new and obsolescent schedules for Psychology and decided that only the preferred schedule should be presented.

The Foreign Survey of the Use of the Decimal Classification got under way in the Spring, under the general supervision of Edwin B. Colburn as Chairman of the Steering Committee. Harriet MacPherson was appointed as Director of the Survey but was forced to withdraw because of her health. Sarah K. Vann succeeded her, and secured the assistance of Pauline Seely as a surveyor. Both Miss Vann and Miss Seely have been abroad during the Spring. Miss Vann is planning a further trip during the Fall. A preliminary report has been made of suggestions so far received, but a final report awaits the completion of Miss Vann's travels.

At the October meeting Wyllis E. Wright was elected Chairman and Carlyle J. Frarey Vice-Chairman for the term 1963/66, and Deo B. Colburn Secretary with an indefinite appointment.

The members of the Committee during the past year have been Carlyle J. Frarey, Esther J. Piercey and Pauline A. Seely, representing the American Library Association; Edwin B. Colburn, Marietta Daniels Shepard and Wyllis E. Wright, representing the Lake Placid Club Education Foundation; Godfrey Dewey, Virginia Drewry and Joseph W. Rogers as the continuing members representing the Forest Press, Inc., American Library Association, and the Library of Congress respectively.—Wyllis E. Wright, Chairman

Volume 9, Number 1, Winter 1965

103
Statement on Types of Classification Available to New Academic Libraries*

Introduction

NEW ACADEMIC LIBRARIES are faced with a rather important decision at their very beginning. What classification system will prove most adaptable and most durable over a long period of time? The Classification Committee, at the request of the Cataloging Policy and Research Committee, has studied this problem carefully. The statement embodying our conclusions does not consist of arbitrary recommendations for one scheme or another, but sets forth the characteristics of the major classifications as they apply to different situations. The question-and-answer method has been used for convenience in helping a library define its own situation. The Committee was extremely fortunate in having as a member, Miss Gertrude L. Oellrich of Alanar Book Processing Center, who is actually engaged in classifying with several schemes simultaneously and who, therefore, is in a good position to compare them.

Definition

The purpose of classification is defined in this statement as a systematic, subject-oriented arrangement for shelving, a location device for open or closed shelf collections of books, not for the classification of knowledge.

Statement

The field of study was narrowed to a survey of the comparative merits of the Dewey Decimal (DC) and Library of Congress (LC) Classifications. Both these systems are growing, are being kept up-to-date with quarterly revisions, and Dewey, at least, now has a users' guide to go with it. One or the other of these two classifications is now used in the majority of academic libraries.

1. Is it important to consider other classification systems in addition to the Dewey Decimal and Library of Congress schemes?

There are several other classification schemes available, most being used in libraries somewhere in the world.

BLISS—This is used rather extensively in Australia. It is a good logical system, but is not being kept up-to-date for ease of usage. The manuscript version does not agree with the published version. Letter notation.

RIDER—This is a new scheme quite similar to the Dewey Decimal Classification, but with a letter notation as in Bliss. There is no way of updating it except as individual libraries undertake the job.

READER INTEREST—This system is more suitable for a public library which must cater to constantly changing interests.

UNIVERSAL DECIMAL—This is a European adaptation of the Dewey Decimal Classification. It is greatly expanded in the science and technology sections to serve the purposes of scientific documentation. Except in occasional areas, the rest has scarcely more depth, and, in some cases, less modernity than Dewey. It is too lopsided for a general library, but would be suitable for special collections in scientific and technical subjects. A scientific-technical edition in English, with a good guide written by Jack Mills, is available. This classification is updated periodically. If a new and centralized secretariat is established and the major revisions now under consideration are adopted, it may be of greater significance than it is now. A classification system to watch.

COLON—This system is used at various establishments in England and in India. The current (6th) edition schedules are rather limited in scope. It is very difficult to use because it necessitates an attitude of mind that is totally different from that employed in any other classification process. At present there is no good guide to its use. The explanatory portions in the 6th edition are extremely difficult reading and less than clear.

FACETED CLASSIFICATION—The schemes of this type developed so far are for specialized subjects. Until a general system is developed, this type of classification is not suitable for a generalized library.

2. What characteristics influence the choice of a classification?

COMPREHENSIVENESS—LC is much broader and more comprehensive than DC. It permits finer (closer) classification. The “P” schedules, in particular, are tremendous in size and, while hard to learn to use, have much “elbow room.”

FLEXIBILITY—LC has the advantage of not being logical in exposition, as a rule, and while it is practically impossible to memorize, it is easy to expand without upsetting existing classified books. The advantage of a non-logical classification is apparent in dealing with rapidly advancing subjects, as the sciences, where a major change in thought can throw out a whole branch in a previous arrangement of knowledge. LC can interpolate where DC must compromise.

Dewey has to be expanded through further breakdown, sub-classification or re-naming and reassigning classes. LC can be expanded by interpolation because the whole system does not have to be logical but can, to a considerable degree, grow like Topsy without regard to its environment. It has been possible to abridge Dewey, but not LC.

LC permits variation in the treatment of specialized topics. Sayers states that LC was the “first to recognize the necessity for variations.
of treatment as between the different classes, and it is this feature of
the scheme which has found so much favor in academic libraries.

COMPLEXITY—The mixed notation of LC is more complex than the
pure notation of DC. However, Gulledge® stated that the LC num-
bers are on the average shorter than DC.

INDEXING—LC has no combined index and this is considered a fault
by users. The relative index of DC has been praised, although the in-
dex for the 16th edition is inferior to that of the 14th edition.

BROWSABILITY—DC has the advantage of providing browsability. In
open stack libraries, this is important. It is practically impossible to
browse with LC although people try it all the time.

NOTATION—Dewey’s notation is positional, each position represents
a classification level. LC notation is ordinal. Each class has a number
of its own not necessarily related to preceding or following classes.

CLASSES—LC has three times as many classes as DC. Neither classifica-
tion fits the college curriculum.

INTERPRETATION OF USE—DC seems to be superior in this because
there is a Users’ Manual for DC.

SYSTEMS OF SUBDIVISIONS—The system seems to be better in LC,
but the tables are difficult to use. Students have some difficulty learn-
ing to build numbers in DC, but once learned, the application is uni-
form throughout the system. However, if one uses LC cards there
may be fewer instances when numbers have to be built.

REVISIONS—Both LC and DC are now being kept current with quar-
terly corrections.

Problem areas noted are fiction, translation, literature subdivi-
sions, political subdivisions, and study and teaching. Both schemes
have been criticized.

3. Is the choice of LC or DC a function of the size of the collection?

It seems to be an accepted fact in the literature on classification that
the LC scheme, because it lacks general numbers for many areas, does
not serve the small library needing broad classification.

The ceiling for the 15th edition of Dewey was for libraries of 200,
000, though this ceiling was lifted for the 16th edition to include li-
braries of whatever size. However, in a survey conducted by Thelma
Eaton among college and university libraries, only in libraries of less
than 200,000 volumes was the value of DC stressed.

The LC classification is used by 300 university, special, and govern-
mental libraries in the United States and abroad. The scheme does not
lend itself easily to abridgment for use in libraries with small collec-
tions, and serves best in libraries with large collections or special libraries
which require minute subdivision of limited subjects.

The Committee recommends Dewey for libraries with general collec-
tions up to 200,000 volumes in size, and the Library of Congress system
for those expected to be larger and for those small libraries with spe-
cialized collections in the social sciences and humanities.
4. What local or existing factors must be considered in making a choice of a classification?

Various questions were considered:
if the library is in a state system, what is the rest of the state doing?
should such factors as the use of fixed location (shelving by size), open
or closed stacks, availability of a centralized catalog or even adoption
of some special subject heading system be considered?
Apparently none of these factors had much to do with the choice of a
classification system, since nothing could be found in the literature of
librarianship relating to them.

5. Is a divisional library vs. a central library a reason for preferring one
classification or the other?

In their description of the process of adapting the Dewey Classification
for use in a college library, Ashton and Hansen note that "no system
has been devised with the divisional plan as its basis." On the other hand,
they do not recommend attempting to develop one. They modified the
DC system to suit their needs with only 10% of the collection requiring
reclassification. But their conclusion was that "Dewey, as it now stands,
confuses the divisional library issue." The LC system, having more
classes, could be more easily adjusted to the divisional concept.

6. Which classification, LC or DC, is more satisfactory for centralized cata-
logging?

A. Library of Congress characteristics:
(1) LC cards give class numbers plus LC-style Cutter numbers in the
following proportion (barring law and lesser known languages): 85%.
(2) There are fewer changes in LC class numbers than in DC. (see, for
for example, literature periods between the 14th and 15th editions
of Dewey).
(3) LC will often serve specialized departments better than DC, and
since a single system is easier for the whole college, the LC system
becomes the chosen one where centralized cataloging is done.

B. Dewey Classification characteristics:
(1) DC numbers appear on LC cards for about 35% of titles. This
means that about 65% of cards purchased from LC have no DC
number.
(2) DC changes cause confusion when reprinted LC cards are used
and require constant professional attention.
(3) DC is too permissive. This is a boon to custom cataloging or to
local cataloging preferences, but a Pandora's Box in centralized
cataloging.
e.g. geography and history combination.
rearrangement of class to bring related classes together, such
as philology and literature, 400 and 800.
biography in 920 or subject number.
bibliography in 016 or subject number.

extension of class numbers or building numbers beyond what
is given on the LC card.
shortening of numbers when class number ends in .01, .08,
etc. Requires common sense—one cannot stop at zero (821.0)
if a general rule for no more than four numbers has been
made.

7. Should the classification numbers on the LC cards be accepted in preference to making local changes?

The consensus in the literature is that catalogers should accept the
Library of Congress classification choices in preference to making local
changes. The reasons for this are:

(1) This practice makes the cataloging function easier and more econo-

(2) Few, if any, libraries use all the numbers assigned by LC, but the
majority indicate that they accept 90-99% of the numbers on cards
and proof-sheets.

(3) Dawson actually examined cards in a selected number of libraries
and found only 84.45% of the numbers were accepted.

Some of the reasons given for NOT accepting the numbers given on LC
cards are:

(1) Changes in schedules since older cards were printed give obsolete
numbers.

(2) Absence of numbers.

(3) Failure to accept wholly or apply consistently the various revisions
of the classification systems.

8. What are the relative costs of using LC and DC?

The comments below are based on daily observance of LC cards in a
catalog department, a limited study of 500 LC cards as received, and a
study of 500 LC entries in the National Union Catalog, excluding law
and lesser known foreign languages.

Classification nos. on cards (excluding law & lesser known languages)

<table>
<thead>
<tr>
<th>Daily cards</th>
<th>500 cards</th>
<th>Nat. Union Cat. LC card entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO LC class number</td>
<td>10%</td>
<td>16%</td>
</tr>
<tr>
<td>NO DC class number</td>
<td>50%</td>
<td>63%</td>
</tr>
</tbody>
</table>

To supply class numbers where these are lacking on LC cards requires
from 1 to 10 minutes per title by a person familiar with the LC and DC
systems. Obviously, the DC is more costly in this respect, and only the
advantages derived from its use can counterbalance this cost. It is said
that fewer DC numbers will appear on LC cards in the future.
To this basic cost, which requires the employment of a professional cataloger for DC, must be added:

1. variations due to edition changes on reprinted cards 10%
2. extension of the DC number when libraries prefer close classification. Longer numbers are now being given on more recent cards 5%
3. assignment of Cutter numbers for every title 5%
4. assignment of fiction number (i.e. 813) if libraries do not use F or have no number for fiction 100%

The percentages here occurred in the 500 cards. A closer study might be useful.

Considering the LC classification, to the basic cost of supplying 15% of the numbers must be added the following:

1. supply numbers for all subjects treated from the legal aspect 10% of the 500 cards
2. supply literature numbers for PZ 100%
3. supply numbers for the papers, proceedings, etc., of universities, societies, etc., where LC has assigned an A class number (for the society) instead of a number for the subject in the paper e.g. Riabov's Rules of Motion of Artificial Celestial Bodies, LC cards gives: 629.1388 (a case of former DC number)

Some colleges want this material with the subject, so an LC number must be supplied for it.
In some cases only an LC A class number is assigned and no Dewey number, necessitating a new class number with either classification.

It is obvious that the LC system is less expensive even though some LC numbers are lacking and the problems of law, fiction, and series remain.

9. Which classification would be easier to use for a mechanized system?

Several factors have a bearing on which classification would be most easily mechanized.

STORAGE—Since the classification notation must be converted to Double Digit form, the storage space for each class number would be twice its length.

Storage units necessary would depend on the type of machine.
For example:
- Decimal machine (IBM 7070 series) 10 digits to the field, allowing 5 letters or numbers per field.
- Binary machine (IBM 7090 series) 18 digits to the field, allowing 9 letters or numbers per field.

Few LC numbers run higher than 10 digits (20 in the Double Digit form), which would take 2 fields in either machine.

For Dewey, if a maximum size were not predetermined, one would have to use the longest number in the library as base for determining how many fields were needed.
INPUT-OUTPUT PROCEDURES in compiler language—WRITE TAPE or PRINT routines would be very simple for Dewey, but complex for LC or UDC. The three statements below are written in FÖRTRAN for a number in each system:

- FORMAT (3H TK, F4.0, 1HM, 12) LC (TK872.45 Masers)
- FORMAT (F9.6) DC (629.13453 Rockets)
- FORMAT (F4.1,2H+, X,F4.1) UDC (655.1 + 688.1 Printing and Binding)

Obviously Dewey is easier for the machine to handle on output since there is less work to be performed internally. The same is true of input statements, again using FÖRTRAN as an example:

READ 5, K

- 5 FORMAT (2A1, F4.0, A1, I2) LC (TK872.45)
- 5 FORMAT (F9.6) DC (629.13453)
- 5 FORMAT (F4.1, A1, F4.1) UDC (655.1 + 688.1)

Some of the connectors in UDC could not be used in their present form because they already have a meaning in FÖRTRAN. (e.g. /= start a separate line) Again, Dewey is easiest to handle as far as the machine is concerned.

USE OF DC for storage and compiler—If DC were arbitrarily limited to 15 places after the decimal, it would require 2 fields for storage in a binary machine of the 7090 type, or 3 in a decimal machine of the 7070 series. It would take the format (F18.15) in FÖRTRAN for both input and output.

The largest LC number would require the same field space in the binary machine, but less in the decimal. It would take more machine processing in FÖRTRAN compiler language.

AN ALL LETTER NOTATION, with a decimal for subdivision, would be even better than DC from a classification point of view, and no worse from the machine point of view. Such notation does not now exist.

IF A MECHANIZED SHELVING AND FETCHING SYSTEM is developed to replace stack men, classification as a shelf location code could end. Books could be filed by accession number or some other numerical system, or by size, etc. In such a case the dictionary catalog might be replaced or supplemented by a classified catalog. Classification itself could be more completely developed as an organizational system if it did not have to serve as a shelf location device. Thus, relationships among concepts and structure could be described. Multiple generic relationship classifications could be made by computer.¹⁰

REFERENCES


• 110 •

Library Resources & Technical Services

NOMINEES. 1964/65

Resources and Technical Services Division

For Vice-president (President-elect):
Wesley C. Simonton, Library School, University of Minnesota, Minneapolis, Minn.
Wyllis E. Wright, Williams College Library, Williamstown, Mass.

For Director-at-large:—three-year term:
Dale M. Bentz, State University of Iowa Libraries, Iowa City, Ia.
Mrs. Avis G. Zebker, Brooklyn Public Library, Brooklyn, N. Y.

Acquisitions Section

For Vice-chairman (Chairman-elect):
Carl Jackson, University of Colorado Libraries, Boulder, Colo.

For Secretary:—three-year term:
Marietta Chicorel, University of Washington Library, Seattle, Wash.

For Member-at-large:—three-year term:
Robert D. Stevens, Director of Research Collections, East-West Center, Honolulu, Hawaii

Volume 9, Number 1, Winter 1965
Cataloging and Classification Section

For Vice-chairman (Chairman-elect):
Alex Ladenson, Chicago Public Library, Chicago, Ill.
Marian Sanner, Enoch Pratt Free Library, Baltimore, Maryland

For Member-at-large:—three-year term:
Mrs. Ruth F. Strout, Graduate Library School, University of Chicago, Chicago, Ill.

For Secretary:—three-year term:
Richard O. Pautzsch, Brooklyn Public Library, Brooklyn, N. Y.
Hilda Steinweg, Ohio University Library, Athens, Ohio

Copying Methods Section

For Vice-chairman (Chairman-elect):
Joseph Popecki, Catholic University of America, Washington, D. C.
Stephen Salmon, Washington University Libraries, St. Louis, Missouri

For Secretary:
Jonathan R. Ashton, School of Library Science, Queens College of the City University of New York, Flushing, N. Y.
Albert J. Diaz, Microcard Editions, Inc., Washington, D. C.

Serials Section

For Vice-chairman (Chairman-elect):
Thomas D. Gillies, Linda Hall Library, Kansas City, Mo.

Member-at-large:—three-year term:
Mrs. Jacqueline W. Felter, The Medical Library Center of New York, New York, N. Y.
Second International Study Conference on Classification Research: Conclusions and Recommendations

1. Present Situation

The first International Study Conference on Classification for Information Retrieval was held at Dorking under the auspices of FID in 1957. It was the continuation of the efforts undertaken by the FID over a number of years, in its CC and CA committees and at its annual conferences, particularly at the Rome 1951 and Brussels 1955 conferences.

In the seven years since Dorking, much progress has been made, both in the design of classification systems and in the application of machines to information retrieval. There were five nations represented at Dorking. Individuals from 16 countries and two international bodies attended the present conference. The scope of the second conference is much broader than the first. This reflects the growing interest and increased volume of research being carried on in this field.

It is no longer necessary to insist on the role of classification in information storage and retrieval languages. The earlier doubts on the feasibility of machine retrieval have largely disappeared. Moreover, it has been widely recognized that paradigmatical organization is an essential feature of any effective machine system. Many theoretical issues have been clarified, and progress in engineering capabilities for the processing of large information files has been significant.

Important contributions to classification theory have been made by various disciplines, such as structural linguistics, semantics, mathematics, logic, and epistemology. Experimental testing of existing classification systems has been pursued on an increasing scale. The individuals at the present conference reflect this multi-disciplinary approach to the problems in classification. The purpose of the present conference has been to consider the situation and point the way toward productive future work.

2. New Directions

2.1. Theoretical research

The existing body of theory is in need of further elaboration on various lines, for both general and special classification systems, as well as

* Held at Elsinore, Denmark, Sept. 14-18, 1964; approved Sept. 18, 1964. This paper was issued at the close of the Conference.

1 By "classification" is meant any method creating relations, generic or other, between individual semantic units, regardless of the degree in hierarchy contained in the systems and of whether those systems would be applied in connection with traditional or more or less mechanized methods of document searching.

2 See footnote to 2.1 (a) below.

Volume 9, Number 1, Winter 1965
for such established applications as shelf arrangement, card catalogues, indexes, bibliographies on the one hand and machine systems of various degrees of mechanization and automation on the other hand.

This theoretical approach embraces among other things:

- the study of the mutual interrelationships between thought and language, i.e. the connection between concepts, relations between concepts, and their expression in the natural language;
- the linguistic study of terminology in scientific and technical fields;
- the construction of controlled vocabularies, thesauri (with or without hierarchical relations expressed) as well as classificatory pre-coordinated structures;
- the study of various methods for embodying "analytic" relations given by context (so-called syntactical structures);
- the analysis and evaluation of the functional relationship between the various components of systems (including classification, codes, and equipment);
- the study of the behavioural processes, e.g. the inductive processes, (both at the individual and group levels), which largely determine the choice of semantic categories.

In this connection, a number of specific questions relating to classification theory should be investigated further, such as:

(a) the possible separation of paradigmatic and syntagmatic relations;
(b) the use of universally applicable categories or categories applicable to several fields;
(c) the domains of application and conditions for the use of integrative levels;
(d) the formal (mathematical and logical) foundations of classification;
(e) the relevance of a classification system to the subject being classified, taking into account related semantic questions from the socio-psychological point of view;
(f) data classification (look-up systems) as contrasted with document classification;
(g) the optimal stage of precision in classification language when expressing complexity.

At the frontier of theoretical research and practical application we should investigate:

(h) symbolization (notation) problems;
(i) relationship between "general" encyclopedic classification schemes, and "specialized" classification schemes.

2.2. Applications

Theoretical studies mentioned above should be applied to:

1 These terms being respectively equivalent to: lexical and syntactic structures, vertical and horizontal axes, etc.

• 114 •

Library Resources & Technical Services
(a) the improvement of existing classification, including work on methods for construction of thesauri and related tools;
(b) the achievement of better design in new classifications;
(c) the exploration and implementation of compatibility among classification systems and thesauri, including standardized vocabularies;
(d) the convertibility of the records of material indexed in one system into another;
(e) the study of the interaction between the classification systems and computer technology in the process of system analysis and programming; and the effects of the cooperation between the classificationist and the systems engineer.

2.3. Evaluation of Classification Systems

The objective of work in this area should be to obtain generally recognized and standardized techniques for evaluation as well as the measurement of the dimensions of a classification system. It is necessary to devise:

(a) more adequate experimental and operational test designs,
(b) better evaluation techniques,
(c) mathematical models for the more precise and reliable description of systems,
(d) better methods for the comparison and evaluation of classification systems.

Tests and evaluation of existing systems in a variety of disciplines should be encouraged. An international cooperative effort should be made on collections of sufficiently large size to test the utility of classification systems. There is a need for further work along these lines:

(e) tests which include the users of classification systems, the classifiers and indexers, and to uncover divergencies between index description and the author's own analysis of his paper.
(f) studies of the reliability and consistency of results of classification performed by different classifiers within one classification system and/or various systems;
(g) more precise and reliable methods of measuring documentary relevance to search queries.

2.4. Automated Classification

The problems related to the construction and possible application of automated classification could have appeared under all the preceding headings, but the high level of interest in this area at this time called for special and separate treatment.

Automated classification includes (1) the mathematical derivation of classification schedules (the work of the classificationist); and (2) the automated assignment of documents to categories (the act of class-
ing which is the work of the classifier), regardless of whether the categories were automatically derived or were chosen from a classification scheme previously devised.

The automated assignment of documents using a prepared classification schedule is already operationally feasible. The other kind of automated classification has been subject to some experimentation. Further experimentation on an adequate statistical basis is desirable in order to determine the scope and limitations of these procedures in comparison with other classification methods. Statistically reliable studies are needed to determine how the automated classification compares with the vocabulary and the document distribution applied by classifiers and users.

3. Ways and Means

3.0 Steps have been taken to systematize terminology on a national level, in order to standardize and simplify the vocabularies used for describing classification systems. It is recognized that standardization of terms requires prior coordination of concepts by the different specialists concerned. It is urged that this effort be organized by an international body such as FID/CR or ISO.

3.1 There is already some cooperation among research teams on an informal basis. This should be extended and placed on a more organized footing, particularly through international exchange of research personnel, exchange of data and computer programs, fellowships, etc.

3.2 A register of classification research projects in progress should be maintained by some suitable organization. A frequent and prompt publication containing information on new activities is needed. Clearing-house arrangements should be made for the collection and provision of information on classification systems and thesauri for special fields, and materials available in machine code for the use in classification research. Critical reviews of research progress should be encouraged.

3.3 Shortage of personnel for classification research is a serious obstacle to progress. Measures to improve this situation by training are being made and should be encouraged in institutions of higher learning. Attempts should also be made to interest research workers from certain other disciplines in classification research problems.

3.4 In accordance with the FID Bureau recommendation of June, 1964, the formation of national groups for the study of classification and retrieval languages should be encouraged where such groups do not exist. The FID/CR Committee is to be considered as the link between these various groups.

3.5 More financial support will be required for classification research and its supporting organizations. The responsibility for financing research projects is primarily in the hands of national organizations for scientific research, foundations, and international governmental organizations, like Unesco, and non-governmental organizations like FID itself. Eventual inclusion of this list of priorities for
fundamental research on classification problems in the FID Long Term Programme could stimulate additional support and encourage further cooperation among the various organizations concerned.

3.6 Further symposia or study groups should be organized at reasonably short intervals. They should be planned in such a way as to facilitate detailed study and critical review of work in progress.

**Variant Pricing of Serial Publications**

**PETER GELLATLY, Serials Librarian**

*University of Washington Library, Seattle*

**THE PROBLEM** of the variant pricing of serial publications is a complex one. Fixed pricing is by no means the rule and seldom occurs, except in the case of such large, mass-distributed publications as *Life, Time, the Saturday Evening Post*, and so on. But even here one cannot plump too solidly in his conclusions. The rates are seldom as rigidly fixed as might appear at first glance. They are, in fact, almost always lower in the first year of subscription than at any other time, simply because it is harder for a publisher to acquire a new subscriber than to retain an old one. In the following years also, the rates can vary, and often do. Serial prices are anything but stable, and no easy assumptions can be made about them.

The way in which a subscription is obtained matters. For the private individual, buying a subscription at his door or over the telephone is the least economical way to do it (few subscriptions are available in this way, despite the omnipresence of the salesmen); and, for that matter, publications bought from salesmen cost about as much as they do on the newsstands, the large and resounding offers of the salesmen notwithstanding. It is a much better idea for the private individual to go directly to the publisher for his subscriptions. Libraries, of course, usually make their purchases through jobbers for the convenience this arrangement affords, and while occasionally they receive from the jobbers the favorable long-term subscription rates that are available to subscribers in general, this is not always the case. They have to be content with what they get; and whether or not a reduction is available depends upon the jobbers themselves. The fact that many libraries have budgetary prohibi-

* Paper prepared on request of the RTSD Acquisitions Policy and Research Committee.
tions against advance expenditures also influences the rates they pay. Most serials are paid for in advance; and unless it is possible to go two or three years into the future, there is little hope of securing the multiple-year advantage. However, in cases in which orders are sent out in the "until-forbidden" way, jobbers invariably make long-term purchases, and even occasionally pass on some of the savings they obtain in this way to the libraries concerned. In such cases, a pro-rated adjustment is required when premature cancellation occurs; but even so, the arrangement is usually of benefit to the library.

The rates vary for another reason. Publishers, or at least their circulation managers, are increasingly anxious to sell their publications, and one device they use for doing this is to offer bonuses of various sorts to their subscribers. Bonuses are typically offered to new subscribers as an enticement to subscribe, but sometimes also to old subscribers at renewal time, on the theory evidently that some compensation has to be offered them in view of the increased rate charged after the first year, and certainly in view of the lower rate granted new subscribers. The bonuses for new subscribers are, of course, in the way of price concessions. This is rarely the case where renewals are concerned, although multiple-year rates are offered which become progressively more favorable as the subscription period lengthens. What is more usual is for a publisher to offer a few extra issues for a prompt renewal (with cash enclosed as a means of reducing bookkeeping costs to the publisher), and less frequently, a small book, often but not always made up of writings gleaned from past issues of the publication in question. The Harvard Business Review and Changing Times, among others, issue books that are used in this way.

In the matter of renewing subscriptions, it is of interest to note that bargaining exists—not in the common sense of haggling, but in a real sense nevertheless. This is what happens. The longer a subscriber waits before he makes his renewal, the lower the price becomes. With each renewal notice, the publisher makes a concession until at last, in a spurge of what must be tie-making generosity, he offers a rate that very nearly approximates the first-year rate. Not all publishers are amenable to the threat of a discontinued subscription, as not all subscribers are to the blandishments furnished by the publishers for their continued loyalty; but certainly bargaining is possible to a point in dealing with the publishers of the large, commercial periodicals. While such experimenting in the open market is fine for the individual, no library would think of indulging in it—not presumably because it lacks dignity, but because it subjects the library to the danger of having its subscriptions interrupted or even cut off. Most publishers are almost paternally indulgent these days, and one must wait a long time before this extremity is reached; but reached it finally is. Losing a subscription through default is annoying for anyone, but a minor disaster for a library.

Price variants assume many forms. One with which we are all familiar, sometimes happily and sometimes not, is the service-basis pricing used, among others, by the H. W. Wilson Company, according to which a li-
ibrary pays for Wilson publications at rates depending upon its income and the size of its collection. This idea is socially useful in that it allows every library the possibility of purchasing Wilson publications (many, of course, quite indispensable), regardless of its size and solvency, and is certainly commercially astute in that it ensures a large sale for these publications. Its disadvantages have been pointed out often enough, but the only one that seems to matter much nowadays is the excessive amount of paperwork that subscribers find themselves involved in from time to time in order, among other things, to keep the rates up to date.

To say that librarians are addicted to the pay-what-you-can-afford principle is perhaps saying too much, but this principle appears fairly frequently in the pricing of library publications. As the cost of professional memberships varies with the applicant’s salary, so the price of some library publications varies. The latest example of this to come to notice concerns the new Canadian Library Association publication, *Canadian Library Horizons*, which is available to individual subscribers at $9 a year, to libraries with incomes of less than $100,001 at $10, and to libraries with greater incomes at $25. There are many variations on this particular theme. In the case of the new *Economics Library Selections List*, published by the University of Pittsburgh, the rate for students is $1 a year, for professors $2, for university and public libraries and for unaffiliated individuals $10, and for all others $15. This elaborate price schedule, while it was not produced by librarians, is typical in that it illustrates the way in which price schedules are arranged to secure an advantage for special groups.

Instances occur in which libraries are charged less for their subscriptions than are individuals; and certainly the opposite case is not unknown. Libraries are considered fair game by some publishers, who feel no doubt that selling a subscription to a library will decrease their private sales. This assumption is hard to disprove, but equally hard to prove. It is not at all certain that a person who uses a publication in his local library would continue using it if he were required to take out a subscription himself; but of course he might. Still the anti-library (or better, perhaps, the pro-individual-subscriber) bias persists. For all this, many publishers now seem to be coming to the point of making a special effort to woo libraries by giving them favorable rates rather than using discriminatory rates against them. This is particularly so in the case of the publishers of learned journals who are aware that a large part of their subscription revenue comes from libraries. Firms that issue mass-directed publications, on the other hand, are less concerned with the income obtained from their library subscribers and so are less inclined to give them special consideration.

One notable example in which an effort has been made by a publisher to attract library subscribers is seen in the dealings of the Pergamon Press. Subscriptions are supplied by this firm at what it calls its “A” and “B” rates. The “A” rates, which are considerably lower than the “B”, apply to libraries and to various other sorts of institutions, while the “B” rates
are assayed against individual subscribers (including, one assumes, in-
dustrial firms and manufacturers.) Individual subscribers are further
discriminated against in that they are required by Pergamon to certify
(whatever that means) that they will use their subscriptions for their
own personal research and neither lend nor sell them. Pergamon has not
always been so magnanimous in its treatment of libraries (nor, for that
matter, so indifferent to the needs of the individual), but evidently the
realization has been borne in upon it that as libraries are its best cus-
tomers, some concessions ought to be made to them.

Libraries are often the recipients of less generous treatment (although
it must be pointed out that Pergamon’s prices are high, even in the “A”
category), and charged considerably more for their subscriptions than is
the individual subscriber. The feeling here, if it is not hostile to libraries,
is scarcely sympathetic. Apparently it is thought that libraries will pay
any amount for a good publication, whereas individuals are limited in
the amount they can afford to pay and should be shown some preference
in the rates they are charged. A case in point is Iron Age, a publication
sold to individuals at $2, but for which libraries must pay $25. Other pub-
lications which are more expensive to libraries than to individual sub-
scribers: Arithmetic Teacher, Automotive Industries, Journal of Hetero-
cyclic Chemistry, Mathematics Teacher, Progressive Architecture.

Preferential treatment is accorded in still other ways. Members of
learned and professional organizations, for instance, often receive the
publications of their organization as a membership privilege. In this re-
gard, one need only mention the ALA and the generosity with which it
sends out its publications. Such an arrangement, while still very common,
is becoming less so as the cost of publication increases. (It appears that
the cost of publication is mounting at a considerably greater rate than is
the general cost of living, but this is beside the point.) A more usual ar-
rangement is for an organization to make its publications available to its
members at a reduced rate, generally from thirty to fifty per cent of the
cost of these publications to the ordinary subscriber. Examples abound,
but a few will serve to give an idea of how concessions are made in favor
of member-subscribers. Speculum is sold to members of the Mediaeval
Academy of America at a reduction of twenty per cent. Scandinavian
learned societies offer their publications to members at a standard and
invariable thirty per cent off list-price. Finally, the American Institute of
Aeronautics and Astronautics makes its many publications available to
its members at a fifty per cent reduction.

As for library memberships, there are a number of things to be said.
First, a library may become a member of a learned society or other organ-
ization willy-nilly simply by subscribing to one of its publications. This
is a frequent enough happening; and any library with a reasonable col-
lection of serials finds itself also in possession of a number of inadvertent
memberships, most of which bring it nothing but the desired publication.
There are others, however, that bring all sorts of secondary publications,
some perhaps wanted and some perhaps not. Second, it is occasionally

- 120 -

Library Resources & Technical Services
necessary for a library to take out a membership before it is put on a society's mailing list. This is usually no hardship, although there may be a charge both for the membership and for the publications. In some such cases, the favorable subscription rate for members more than offsets the cost of the membership, but this is not always so. It might well be that a large technical library would find paying the $500-a-year membership fee of the American Institute of Aeronautics and Astronautics an advantage, since the number of publications put out by this organization each year and made available to members at half-price is enormous. On the other hand, many libraries would find so expensive a membership economically unfeasible. The advantages and disadvantages of the membership arrangement have to be considered in each case separately. Third, memberships are not always available to libraries but sometimes to individuals only. This creates no difficulty for the library if the needed publications can be obtained without the necessity of membership; and certainly no library can complain if it gets the publications it is after without excessive cost and bother. It is true, moreover, that society publications are almost always made readily available to libraries, either as gifts or purchases. Fourth, in cases in which personal membership is mandatory, there is sometimes difficulty in finding a librarian with appropriate qualifications, but generally in such cases the society is willing enough to ease its regulations in order to oblige the library and ensure, incidentally, that its publications are given a place in the library's collection.

How does a library find its way about in a situation in which an increasing number of its serial publications have more than one price? The question remains unanswered. There is nothing predictable, or even entirely rational, in the pricing of serial publications; and what is paid for them depends upon a number of considerations, not the least of which is the publisher's state of mind at the time the purchase is made. One thing more: despite the almost frenzied efforts of publishers to sell their publications, real bargains are as rare now as they have always been. The hard sell is here to stay, and it should be recognized for what it is.

EDITOR RECOMMENDS:

"Push Button Bibliography Today and Tomorrow," Bulletin of Bibliography, 24:73-78, 86-88. May-August 1964. Kenneth Shaffer (Director, Graduate School of Library Science, Simmons College, Boston), Ralph Parker (Librarian, University of Missouri, Columbia), and Ludwig Sickmann (teacher at the Bibliothekar-Lehrinstitut, Cologne) discuss the realistic and false hopes of automation in bibliographic retrieval.
The Advantages and Disadvantages of a Classified Periodicals Collection

JOSEPH C. BORDEN, Associate Librarian
University of Arkansas Libraries, Fayetteville

ON THE QUESTION, "What are the advantages and disadvantages of a classified periodicals collection compared with an unclassified collection," it will be desirable at the outset precisely to define or describe a classified periodicals collection and an unclassified periodicals collection. The periodicals belonging to a library need to be shelved in such a way that they can easily be found when needed. This is true whether the periodicals are bound or unbound. If the periodicals are not classified in the library by some scheme of library classification such as the Dewey decimal classification or the Library of Congress classification, then they are usually kept in alphabetical order by title. In either case, all issues of the same title are kept together, supposedly in chronological order, excepting possibly the current issues.

Classification of the periodicals may be defined as the placing of numbers or letters (known as call numbers, class numbers, or class marks) on the title to achieve its being in the library with other publications on the same subject and near to other publications on related subjects and in some logical numerical or alphabetical relationship to other publications in the same or related subjects. Classified periodicals in a library collection are usually shelved among or alongside the classified books in the same or related classifications. An unclassified collection of periodicals may be defined as the arrangement of periodicals in a library by some other method than by classification; and at the present time the only other practical method in use is the alphabetical method of arrangement.

Large general collections of periodicals, such as in university libraries and large reference libraries, usually hold to the classified arrangement, whether the stacks are open or closed; while some smaller collections, such as in special libraries, college libraries, and public libraries, hold to the unclassified or alphabetical arrangement of their periodicals. For example, Gloria Whetstone, in a recent survey of serial practices in 16 selected large university libraries, found that 12 of the 16 university libraries classified all their periodicals, two more classified some of their periodicals, and the other two, namely Wayne State and Rutgers, did not classify any. On the other hand, in special libraries, where most of the periodicals are in a special and restricted subject field, the classifying generally would not serve a useful purpose.

* Article based on a paper delivered before the Arkansas Resources and Technical Services Group Meeting, Little Rock, October 28, 1963.

Library Resources & Technical Services
Since the individual library will be already committed to one method or the other, it may be asked who is interested in the comparative merits of the two means of shelving periodicals. Aside from the theoretical interest of the subject, it is a matter which some librarians have to face in actuality from time to time: when a new library is established, decision must be made about classifying or not classifying the periodicals; when a going library experiences difficulties with its periodicals, it may wish to consider changing to other means of shelf arrangement; when a library is to occupy a new building, it may want to review its way of shelving periodicals; when a faculty member or student in an academic library may ask why classify or why not classify, it is useful to know the answers. For all of these reasons it is worthwhile giving some thought to the question.

Advantages of Alphabetical Arrangement

Processing is less costly for the unclassified or alphabetical arrangement of periodicals. This refers not only to the savings made by not having to classify each title in the first place, but also may carry through to such details as the savings from not needing to mark the call number on each bound volume. From the use standpoint, the finding of the title is more direct, since it is not necessary to find the call number first. In an open shelf arrangement it follows from this that the user of the library can find what he needs more quickly, or at least more directly.

Keeping the periodicals in one part of the building by themselves instead of scattering them with the books makes it easier for persons interested only in periodical references since all of the periodicals are together. This also simplifies matters for the technical staff working with periodicals. For the same reason, if the library enters a period of expanded acquisitions in the field of back files of periodicals, when more shelf space for a particular title or a particular section of titles is needed than originally planned, the necessary shifting is less cumbersome than if the books also should need to be displaced and moved.

Checking holdings for ordering from offers of periodicals from book-sellers is also simplified, as a single alphabetical shelf-list should be easier to consult than a classified shelf-list.

Advantages of A Classified Arrangement

In a classified arrangement, changes in a periodical title permit shelving the whole file of the periodical together despite the title change. Similarly, the various bulletins, transactions, and proceedings of the same organization can be easily kept alongside one another. Users interested in a particular subject or group of related subjects, not only all the periodicals in the subject, but also the books in that subject, will find them relatively close to one another.

The classification of the periodicals usually scatters periodicals of similar titles and thus tends to prevent confusion in shelving and finding; e.g., such titles as begin with the words Journal of the . . . In addition,

Volume 9, Number 1, Winter 1965
difficulties in alphabetizing caused by such small parts of speech in titles as prepositions and articles can be eliminated by classifying. If, later, the collection is to be broken up into departments, a classified arrangement will make the change easier than would a non-classified arrangement. Non-indexed periodicals in special fields are virtually "lost" to the scholar unless they are classified by subject.

Looking over the respective advantages of the two types of arrangement, one realizes that the chief advantage to the library user of the unclassified shelving of periodicals is the direct approach made possible by the simple alphabetical arrangement and that the chief advantage to the library user of the classified arrangement of periodicals is the grouping together of periodicals dealing with the same or similar subjects. The respective disadvantages of the two types of arrangement are the obverses of their advantages. It is perhaps worthwhile to take the time to list the disadvantages, also, with comments.

Disadvantages of Alphabetical Arrangement

When a periodical changes its title, the volumes under the new title will be separated from the volumes of the old title if the alphabetical arrangement is adhered to. This disadvantage is not difficult to overcome partially, by the use of dummies on the shelves, but there is a certain inconvenience to it.

Titles beginning bulletin, proceedings, transactions, etc., of the same organization will be shelved with the bulletins, proceedings, transactions, etc., of other organizations if shelving is strictly by title. Such shelving separates, for example, the proceedings and transactions of the same organization, which for the user of the library will sometimes be inconvenient. This difficulty can be overcome in many cases by shelving such files under the name of the corporate body issuing them. Parenthetically, it should be added that some library users, such as research scholars, especially scientists who are journal-oriented, would prefer to have proceedings and transactions kept under their titles rather than moved to entry under name of the organization, this for the reason that scientific indexes and abstracting services enter under "Bull.", "Proc.", "Trans.", etc.)

The periodicals will not be grouped by subjects, which makes them less convenient for the use of scholars interested only in particular subjects. In many cases this will not be a real difficulty. For example, in libraries with closed stacks, most persons needing the periodicals will not care whether the arrangement is by subject or not; and in most cases the users of periodicals in libraries are found to want a particular citation or several particular citations in a given title or in given titles and are consequently interested only in finding these citations and not in browsing among many titles on the same subject.

Titles beginning Journal of ..., and certain other titles which begin similarly, will be together in one section of the periodical collection. This
is not a serious difficulty, except perhaps to the casual user, and can be re-
duced by careful attention to the alphabetical rules under which un-
classified periodicals are shelved. The same comment may be made on
the disadvantage of long titles containing prepositions and articles.

If the collection is to be later transformed into a departmentalized
collection by broad areas, such as humanities, social sciences, and phys-
icales sciences, the transformation will be more difficult where the titles
have not been previously classified. However, this is merely an incon-
venience to the staff at the time the change is made and should not be a
strong argument against leaving the periodicals unclassified.

The availability and contents of a periodical not indexed in one of the
indexing services tend to be more or less overlooked in an unclassified
collection. 2

Disadvantages of Classified Arrangement

The processing of periodicals in a classified arrangement will be more
costly. This is not really a disadvantage if it works to the well-being of the
library. For example, it might be said that binding periodicals is more
costly than keeping them unbound, but the advantages of binding for
reference purposes are so apparent that the cost is absorbed as a matter
of course. The same thing happens with the cost of classifying in libraries
which classify. It is considered a necessary expense, fully justified by the
advantages the library believes it receives from the classification of its
periodicals.

The user of the library, or the staff member, must find the call num-
ber before he will be able to obtain the desired periodical. Thus it will
take longer for him to get to the citation than in a library where the
shelving is strictly by title. This is generally true of periodicals with sim-
ple titles, but will not necessarily be true of more complicated titles, where
the alphabetical arrangement, theoretically simple, may prove to be more
difficult for the locating of given titles than the classified arrangement
with its call numbers. It should also be remembered that in an open-shell
library with a classified periodicals collection the library staff members
and the constant users will know the classification scheme sufficiently well
after a time to be able to go directly to the shelves without the need to
consult the catalog or other listings first.

Persons concerned only with periodical references will find the peri-
odicals scattered among the books of the library when a classified arrange-
ment is utilized. The disadvantage here is really a matter of inconven-
ience. Even when the periodicals are arranged alphabetically, there is a
certain amount of inconvenience in spacing from one part of the alpha-
bet to another. And where the periodicals are classified, there is a clear
gain of convenience for persons working largely in one field of knowledge.

The disadvantage to a classified collection when shifting time comes
is that books as well as periodicals will need to be shifted. This disadvan-
tage, fortunately, is experienced relatively infrequently.

The checking of holdings of periodicals for ordering purposes is

Volume 9, Number 1, Winter 1965
more cumbersome when the periodicals are classified than when they are shelved alphabetically. This disadvantage is absorbed entirely by the staff and does not trouble the user of the library. It is in the same category of disadvantages as the cost aspect of classification.

It is the specific purposes and objectives of the individual library which should determine whether or not the library should classify its periodicals or classify some and not classify others. For instance, who will use the library—the general public, a general faculty and student body, selected advanced scholars, or subject specialists? How is the library used—open shelves so that the public is getting its periodicals from the shelves directly, or closed shelves so that staff only has access to the shelves? What is the arrangement of the building—is it practical for the periodicals to be in among the books, or do shelf considerations make the separation of the periodicals advisable?

The average librarian today does not encounter the problem of whether or not to classify periodicals. The decision will have been made, probably years before the current librarian came on the job. Yet the fact that there are two general ways to arrange periodicals and the fact that both ways have their adherents should lead the librarian at least to think whether his periodicals will be better utilized one way or the other. Some libraries have made the decision to change from one system to the other and have found the change worthwhile, either in reducing staff time in servicing periodicals or in improving their availability. It behooves the librarian to analyze the purposes and objectives of his collection and then to decide whether better to hold with what he has or to adopt and change over to the other system.

REFERENCES
DO SCIENTISTS NEED A COMPREHENSIVE INDEX TO ALL OF SCIENCE?

When we asked scientists this question recently, most of them replied yes—a few said no. If you are among the few who are perfectly satisfied with the "selective" approach of conventional indexing systems, stop right here. However, if you, along with the majority, feel that improved methods and comprehensiveness are needed in information retrieval and dissemination, read on.

Every research scientist has, on one occasion or another, been stumped by the complex problems of searching the scientific literature. Today's search requirements are hobbled by the limitations of traditionally organized indexing systems.

At the Institute for Scientific Information, we have tested and developed a new dimension in indexing. To this we have added comprehensive coverage, resulting in a unique tool for information discovery. We call it Science Citation Index.

The Science Citation Index instantly identifies the most recent publications referring to a particular work since its publication. The Science Citation Index provides the scientist with an improved, novel starting point: the specific paper or book of a specific author...and from there the searcher is brought forward in time to current papers relating to the earlier work.

Revolutionary? Invaluable? We think so and we think you will too after you've had an opportunity to review descriptive material on this new technique. Write us now and we'll send details by return mail—without obligation, of course.

Dept. 23-2

INSTITUTE FOR SCIENTIFIC INFORMATION
325 Chestnut Street Philadelphia Pa. 19106
The H. W. Wilson Company takes pleasure in announcing the first issue of Biological & Agricultural Index, the new successor to Agricultural Index. A detailed subject index to 146 periodicals published in the United States, Canada, and the British Commonwealth, Biological & Agricultural Index begins publication this month and is expected to be of great value to libraries that need ready reference to magazines in the fields of:

- agricultural chemicals
- agricultural economics
- agricultural engineering
- agricultural research
- agriculture
- animal husbandry
- biology
- botany
- dairying & dairy products
- ecology
- entomology
- feeds
- forestry & conservatism
- genetics
- horticulture
- microbiology
- mycology
- nutrition
- physiology
- plant science
- poultry
- soil science
- veterinary medicine
- zoology

Periodicals were selected for indexing in Biological & Agricultural Index by the subscribers to Agricultural Index and include 78 publications oriented toward biology and 68 oriented toward agriculture. The form of indexing is similar to that used in most of the other Wilson indexes, with subject headings based on those used in the dictionary catalog of the Library of Congress, and numerous subheadings and cross references to facilitate quick reference. Entries include the title of the article, author, periodical, volume, inclusive paging, and date; bibliographies, illustrations, tables, graphs, and diagrams are also noted.

Biological & Agricultural Index will be published monthly except in September, with bound annual cumulations, and is available by annual subscription on the H. W. Wilson service basis. Each subscriber will be charged only for the indexing of those periodicals received by his library. For a quotation of your service basis rate, write today.

THE H. W. WILSON COMPANY
950 UNIVERSITY AVENUE, BRONX, NEW YORK 10452
Presenting a Master Key to the greatest treasure house of applied scientific knowledge in the world.

Important development in information retrieval as beneficial to the reference librarian the Cumulative Book Index. The International Index of Patents is the shortest route to of the information, engineering and scientific technology to be found in patent literature. Developed by the Interdex Corporation with the operation of the U.S. Patent Office, and distributed through Bro-Dart Books, Inc., these sections are fully cross-referenced to enable you answer any question. Material is indexed by Date . . . by Patent Number . by Subject . . . by Class and Subclass . . . by standard Industrial Classification. There is even a list of those patents, in each field of interest, issued before the inception of the numbering system in 1836.

Subscriptions are being accepted now for the first six volumes, covering every United States Chemical Patent issued from 1790 through 1960. Three additional units are in process, covering Foreign Chemical Patents and U.S. and Foreign Electrical Patents, as part of a program to publish the complete body of patents covering all fields of endeavor.

For further information on the International Index of Patents please write to:

BRO-DART BOOKS, Inc.
Dept. 1401A • 1609 Memorial Ave. • Williamsport, Pennsylvania
For Libraries That Want Quality Bookbinding

GLICK BOOKBINDING CORP.

Specialists in the Binding and Rebinding of Books and Periodicals

Serving Institutional, Public And Research Libraries Since 1905

We Have Moved—Our New Address Is
32-15 37th Avenue
Long Island City 1, New York
Stillwell 4-5300

In Nassau and Suffolk
IVanhoe 3-9534

In New Jersey
Mitchell 2-5374