LIBRARY APPLICATIONS OF DATA PROCESSING EQUIPMENT AT ABBOTT LABORATORIES

Walter A. Southern

I. Introduction

The Science Information Services (SIS) of Abbott Laboratories is a medium-sized industrial research information service. Its primary function is to serve Abbott's one-thousand man research staff; but being a company-wide service, all North Chicago divisions and, to a lesser extent, Abbott's international offices, avail themselves of SIS services.

The SIS has a staff of five librarians, nine information scientests, and fourteen clerical assistants. It is a collection of 32,000 volumes with 2,000 volumes being added annually; it receives more than 1,000 different journals, and it has an annual circulation of 135,000 items. Since Abbott is a member of a strongly research-oriented industry and because the SIS promotes its services extensively, the SIS is a very active and heavily-used department.

The SIS first made use of data processing equipment in its operations in 1951 when its journal circulation slips and charge records were produced on punched card equipment. Since then various recurring lists such as our list of current journals and our internal dictionaries have been added. In 1959 an automated information retrieval system for the current literature was begun. An IBM 101 Electronic Statistical Machine located in the department is used for our retrieval operations. In addition, the department has an IBM 870 Document Writing System.

Many of our data processing operations are handled for us by our corporate Data Processing Department. Close rapport has always been maintained with this department for professional guidance in developing our own automated procedures and to make use of their

---

Walter A. Southern is Head, Science Information Services, Abbott Laboratories, North Chicago, Illinois.

equipment. Abbott's Data Processing Department has a card IBM 1401 computer, a Burroughs 220 computer, a Burroughs 280 computer, and accessory equipment.

Conveniently accessible to the SIS, in Abbott's scientific divisions, are IBM 402 and 407 accounting machines, an 007 collator, a 514 reproducing punch, a 548 interpreter and a 1620 computer with discpack.

This report reviews SIS' past experiences in the use of data processing equipment for library procedures. In these operations there are three distinct areas where automation is, or will be, used: information retrieval, serials procedures, and acquisition-cataloging procedures.

At the present time, SIS is mid-way in a study of all procedures which lend themselves to automation. This study is complicated by the fact that the company is presently planning for a new large-scale company-wide computer system. The computer needs of the SIS are being considered in this corporate study. Since the final decision as to which computer will be obtained has not been made, the final plans for operations must be delayed until this decision has been made. In the present study, however, the feasibility of developing computer programs on an interim basis which will make use of present computer equipment is also being considered. Such a plan may mean that the procedures will have to be redesigned for the new computer, but hopefully it is expected that it will mean only a minor reprogramming job. Eventually input-output consoles should be in our library on-line with our new computer for information collection and inquiry.

II. Information Retrieval

The SIS has had an automated information retrieval system in operation since 1959. The system now contains 35,000 literature references on Abbott products and other drugs pertinent to Abbott operations. With a five-year coverage of the literature in the system, information searches are now very productive. Approximately 200 searches per year are now being run.

The search system is currently based on the use of an IBM 101. In designing the system of mechanized searching, a method making use of random numbers in preference to direct coding was selected. A random number system was selected for several reasons: (1) Only one punched card per article is needed which keeps the size of the search deck to a minimum. (2) An open-end dictionary is possible which allows us to add new index terms in any subject area as needed. (3) Machine time required to search the punched card deck

is kept to a minimum, and (4) It is possible to correlate all index terms with an individual article as desired. From the Fischer-Yates tables, SIS has had generated and tested for randomness some 10,000 random numbers. The search field for the random number coding makes use of a 10 × 40 field on the punched card (columns 1-40 and rows 0-9). Twelve-digit random numbers are used for the index terms. Thus, random number 063-123-269-304 which represents penicillin V is punched as follows: digit 3 in column 6, digit 3 in column 12, digit 9 in column 26, and digit 4 in column 30 (see Figure 2). Mathematical computations have shown that it is safe to superimpose 50 index terms per article in our 10 × 40 search field. Beyond 50 superimposed index terms, false relationships show up excessively in searching. The maximum number of terms ever allowed is 70. In addition, rows 11-12 of columns 1-40 are used for supplementary direct coding. Terms coded here are either role indicators or frequently used terms, e.g. clinical report, LD_{50}, intravenous drug administration. Other coding for an article consists of the year of the article (columns 73-74), the corresponding abstract number (columns 75-79), and the machine control punches (column 80). (See Figure 3).

SIS information scientists abstract and code the current literature. Strictly controlled chemical and biological dictionaries are used in assigning index terms. For each index term which is in the dictionaries, two punched cards are prepared. The first is used in preparing the printed dictionaries.(See Figure 4). The index term is punched alphabetically in columns 16-80 and the corresponding random number is punched in columns 1-12. Machine controls are punched in columns 13-15. These cards are filed alphabetically and used whenever a new edition of our dictionaries is prepared on the IBM 1403. The second punched card is a tub index term card. The index term's random number is punched as it appears in the search field in columns 1-40. The term's subject serial number is punched in columns 41-45, and the term itself is alphabetically punched in columns 46-72. These tub cards are gang punched on an IBM 519 Document Originating Machine in quantities of either 25 or 50, depending upon the frequency with which the term is used. The index term is interpreted onto the top of the card by an IBM 557 Alphabetic Interpreter. This gives a visual quality control check so the term is properly filed. Green card stock is used for original tub index term cards. These cards are then used as recorder cards when the supply runs low.

The tub cards are filed horizontally in two tubs 10'5" × 2'5" maintained in the SIS. Each tub accommodates 2,400 index terms which are filed behind guide cards which are 1-1/2" longer than the tub cards. The index terms are typed on the guide cards for easy identification when pulling tub cards.

In processing an article for machine indexing, a Documentation Assistant assembles from the tubs those index term cards which have been assigned to the article. A card punched with the date of the article is also pulled as are other punched cards which correspond to the abstract number assigned to that article. Other machine control cards are assembled into the decklet. These decklets are then sent to the Data Processing Department where all the random numbers corresponding to the indexing numbers are accumulated through all the other cards, including the master search card (a top quality, edge coated card) and the duplicate search card. An IBM 519 is used to accumulate this information. The two search decks are then returned to the SIS.

The remaining punched cards for each decklet are kept together by abstract number and put into dead storage. The system was planned so that it can be converted easily to searching on a computer. In the accumulating procedure, the serial numbers corresponding to the index terms are not accumulated. It is the serial numbers which will be transferred to tape, along with abstract numbers. Searching will then be done by serial numbers rather than random numbers.

Studies show that it is faster and less expensive to pull and accumulate tub index term cards than to key punch all the random numbers. One disadvantage is the amount of space needed for the tubs. Should this operation ever be expanded beyond two tubs, an additional disadvantage would appear: the distance a Documentation Clerk must walk to pull cards from more than two tubs.

Requests for machine searches are programmed by information scientists. The formulation of the search question must be carefully done in order that search terms be correctly selected and arranged. A maximum of four searches can be run at one time. Once the search is programmed, the Documentation Clerk runs the search on the IBM 101. The appropriate random numbers, direct punches, and/or date are wired on a panel board. Once the proper abstract numbers have been selected, they are put through the IBM 870, and a printed list of pertinent abstract numbers is produced. The corresponding abstracts are then pulled from the abstract card files and a check made to see that the abstracts are pertinent to the search request. The final answer to the requester is provided by a packet of the pertinent abstracts. All the procedures connected with punched card preparation and the machine searches are well handled by Documentation Assistants.

For machine searching, a permanently wired panel is used which will accept a question containing up to four index terms: ABCD. A typical search would be: The use of ristocetin (A) in combination with penicillin V (B) in the treatment of bacterial endocarditis (C) in children (D). The search panel has been wired to search all possible combinations of the ABCD terms. If this search, for example,

did not provide an answer containing all index terms ABCD, it is quite possible the inquirer would be interested in the articles containing index terms ABC, ACD, or AD. This search procedure is very useful for generic-specific searches for a question such as: Has hydrochlorothiazide been used in the treatment of diabetes mellitus? Here again, the programming must be emphasized. The information scientists must fully understand these coding principles for best results. This request would be programmed: hydrochlorothiazide (A), thiazide derivatives (B), clinical reports (C), and diabetes mellitus (D). If no ABCD cards were produced, it is likely that the requester would be interested in clinical reports on the use of other thiazide derivatives in the treatment of diabetes (BCD), or he might be interested in BD cards which would be experimental reports on the uses of these drugs in experimental animals.

In systems using random numbers, false relationships of two types are inherent. The first is that the components of several punched random numbers may form the random number of an index term not in the article. And the more index terms assigned to the article, the more likely false relationships will develop. For this reason, too, it is always more desirable to search on four terms for desired information rather than one term since this cuts down on false relationships. The second type of false relationship is the well-known blind Venetian-Venetian blind example. The terms you are searching for are in the article, but not in the relationship you desire. Both types of false relationships must be sorted out before the answer is supplied to the requester.

A search of the 35,000 articles in the system now requires four minutes for panel wiring and 78 minutes for the IBM 101 run. Additional time is required to pull the pertinent abstracts for the abstract file.

As SIS' store of abstracts grows and as more and more machine searches are being made, it is inevitable that the present system must be transferred to a computer. These plans are presently being made and, as noted earlier, the present system was designed for easy transfer to the computer. A program is being prepared so that the serial numbers (corresponding to index terms) and abstract numbers on the decklets now in dead storage will be transferred to a tape. It has not yet been decided whether an inverted file or serial file will be used.

Computer searching will make it possible to eliminate all false relationships caused by superimposing index terms; it will not, of course, eliminate the Venetian blind type[syntax] of false relationship. Computer searching will also make it possible to run searches faster and more conveniently. In addition, it will be possible to multiplex fifty searches. As a result, SIS expects to run many more searches and not be concerned with the time factor of running

searches as it now is with the IBM 101. A further goal is to prepare a second tape with abstract numbers and complete bibliographic citations for each. It will then be possible to produce printed bibliographies and also to make available a selective dissemination of information (SDI) program which would be a much-used service by SIS clients.

The present use of the IBM 101 gives the convenience of running searches whenever needed. With computer searches, one hour per week will probably be assigned for this purpose. The convenience of searching will be lost, but the benefits to be gained will outweigh the inconvenience.

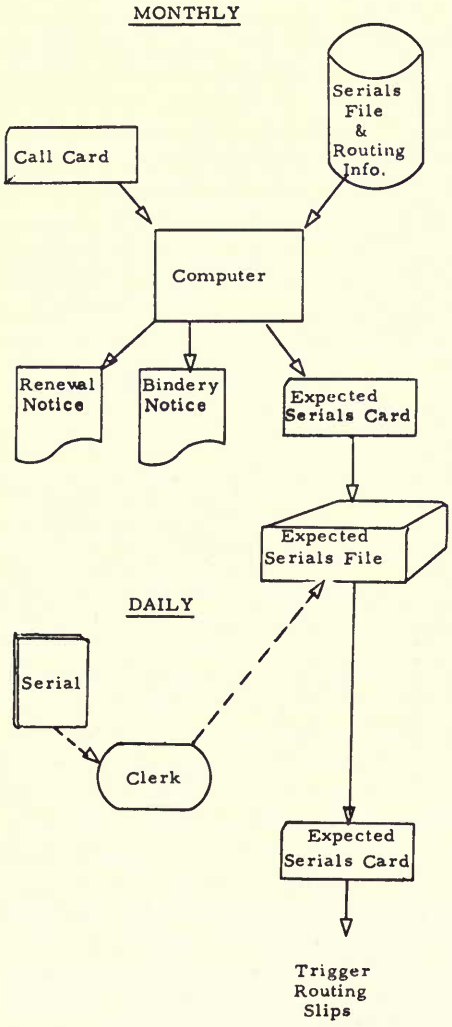
III. Serials

The procedures for the control of current journals present the greatest challenge in SIS operations because of the importance of current journals in the research operations and because of the heavy use of current journals by the scientific staff. It is quite natural then that our first use of automated library procedures was for the preparation of journal circulation lists. This automated procedure for journal circulation lists which was begun in 1951 is still in use today. Circulation slips and duplicate charge cards include the following information: name of journal, copy number, serial number, and employees' names and department numbers [See Figure 1]. Punched cards are used for the input. When the system was started, the IBM 407 accounting machine was used for producing the lists, but today the IBM 1403 printer is used. In addition to the circulation lists and charge cards, lists of journals routinely received by individuals are also produced on an annual basis. These lists are used for annual reviews of current reading, both by the individuals and by their department managers. Other by-products are printed scrolls of all circulation lists which are used for many reference purposes, printed guides for the storage of circulation lists, and statistical data. On an annual basis, SIS receives a report on the number of individuals receiving journals routinely, the number of journals received by each department, and the total number of journals routinely circulated (98,357 journals in 1964). These data are of special interest at this time as SIS is planning additional research facilities away from the present location; they will be useful in planning library facilities for this area.

The next step is for computer control of all journal procedures. This will include: purchase orders for new journals, annual renewal of journals, checking in of current journals, journal claims, overdue notices for journals out of circulation, a union list of journal holdings, plus all procedures covered by the present automated system.

A computer program for all SIS journal procedures has been developed by Mr. Donald H. Kraft, Industry Marketing Representative for IBM in Chicago. The program has been developed for the IBM 1620 computer, but the basic principles are applicable to any computer system. The flow charts for this program follow. The flow charts of a second program, using punched card equipment instead of a computer were prepared for SIS by Mr. L. R. Chapman, Sales Representative of IBM's Evanston Office.

A Generalized Library Serials Check-in & Routing Procedure Using Using Data Processing Equipment



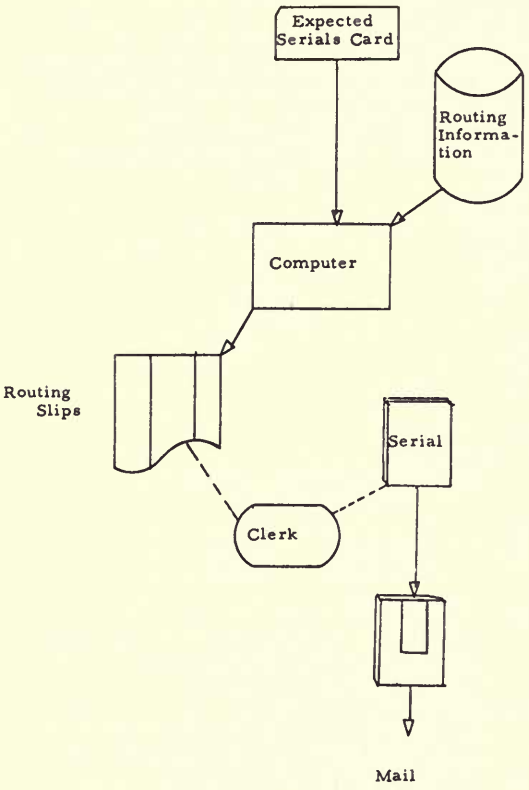
The Call Card calls from disk storage the computer program which generates the monthly Expected Serials File. This file consists of a punched IBM card for each copy of each serial expected the following month. The card is punched with journal name, volume number, month, copy number, destination library, bindery code, class number.

Renewal and bindery notices are written when needed.

Daily Check-in Procedure

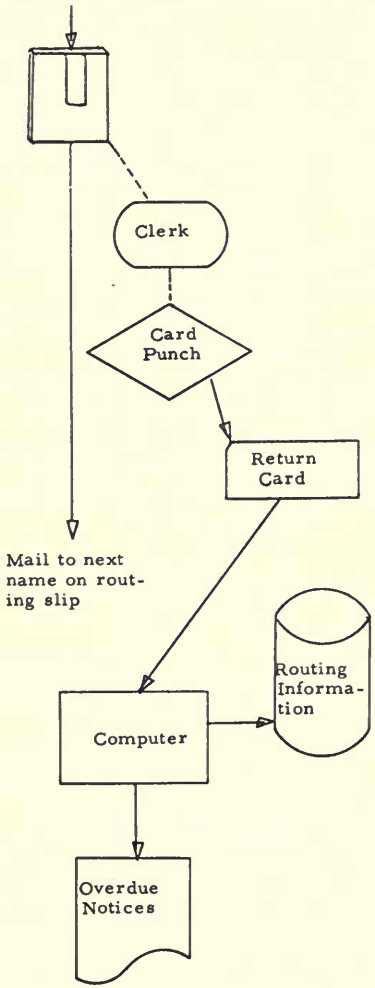
As the serials arrive, a clerk pulls from the Expected Serials File the corresponding IBM cards. These cards will trigger the printing of routing slips.

Daily Preparation of Routing Slips



The Expected Serials Cards are fed into the computer. The disk file contains the routing lists, which are stored magnetically on its disk surfaces. The computer then prints the paper routing slips corresponding to the Expected Serials Cards. The routing slips contain serial name, number, issue, copy, and date, as well as a numbered list of the recipients.

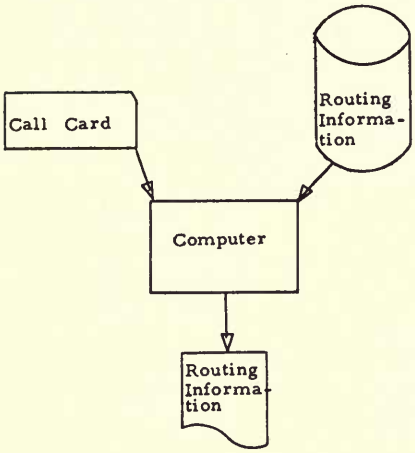
The serials clerk staples the routing slips to the appropriate serials.



After a reader has finished the serial, he returns it to the library. The serials clerk key punches an IBM return card with the journal identification and the line number of the last name scratched off the list. The return card can contain information for more than one transaction.

The return card is read into the computer, which searches the disk file for the appropriate routing record. The computer then posts today's date beside the next name on the disk record, indicating the date the serial was sent to him. Overdue notices are printed for those serials not returned within the loan period. The return cards are destroyed.

Weekly Printout of Current Routings

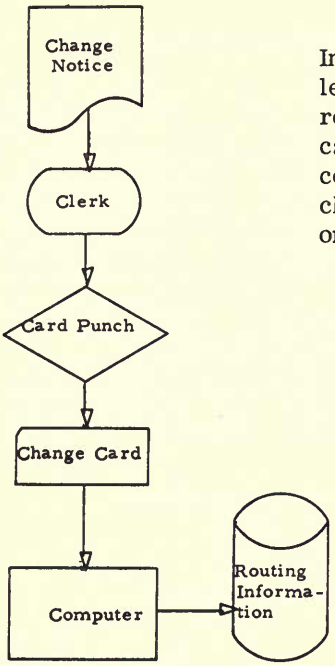


Each week, a program is initiated by a call card to print "CURRENT ROUTING INFORMATION." The computer prints the journal name, identification number, date, volume, number, copy as well as the names of the next two persons on the routing list. This list is consulted by the serials librarian to answer queries relating to the whereabouts of particular journals.

A sample is shown.

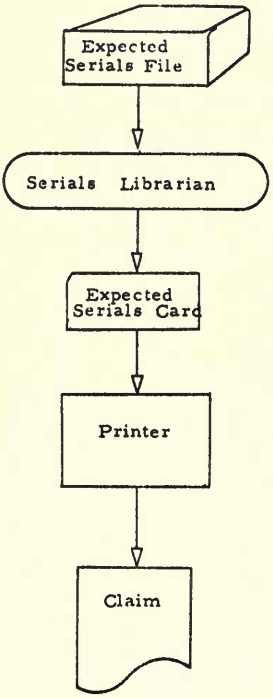
```
        CURRENT ROUTING INFORMATION
              WEEK OF 12/18/64

THE LIBRARY JOURNAL, 7825, 10/64, VOL. 34, NO. 4, COPY 1
          JOHN JONES      DEPT. 72
          BILL SMITH      DEPT. 90
```



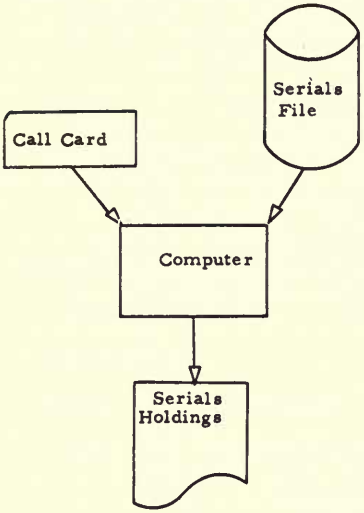
Information relating to additions, deletions and address changes to the routing lists are punched into IBM cards. These cards are fed into the computer which makes the necessary changes to the lists stored magnetically on the disk surfaces.

Monthly Claims Reporting



Cards remaining in the Expected Serials File at the end of the month are examined by the serials librarian. Those cards representing claims are read into a printer or computer, and claims reports are written.

Serials Holdings List



Periodically, the computer can print a Serials Holdings List from information stored magnetically on the disk file. This list can be prepared in multiple carbon copies or directly on reproduction mats by the computer.

A Keyword-in-Context (KWIC) Index can also be prepared by the computer, thereby indexing the holdings list by each significant title word

IV. Acquisition-Cataloging

Up to the present time none of the acquisition-cataloging procedures have been automated. But in our study we believe that a total systems approach will be established for them. Whether or not the procedures will be developed on an interim basis for the Burroughs 220-280 system has not been decided.

The following procedures will be included in the automation of the acquisition-cataloging operations:

1. Purchase orders.
2. Outstanding order file.
3. Claims.
4. Book catalogs (and possible catalog cards until the complete collection is automated).
5. Book charge cards, book pocket labels, and book spine markings.
6. Book overdue notices.
7. New book announcement lists.
8. Printed lists of books on specific subjects.
9. Various statistical reports.

The automated procedures being considered for the acquisition-cataloging procedures follow closely those developed by Mr. R. E. Durkin for the IBM libraries at Kingston and Poughkeepsie. Flow

charts, diagrams, and examples of their operations follow. This progran, called the Program Library Tape, is in the process of being documented by IBM, and the descriptive brochure should shortly be available.

Program Library Tape

A. Basic Equipment needed:

1. 1401 computer with 4,000 positions of core storage
2. 4 tape drives
3. 1402 punched card reader
4. 1403 printer.

B. Hourly Costs:

The estimated hourly charges are in the area of $45 - $50.

C. Program Features:

1. AUDIT - Checks, and in some cases, corrects all key-punched input; prints a proof-reading copy of the input; and transfers information from punched card records to magnetic tape.
2. PRINT - A printing program to prepare reports of listings in any of variable page formats.
3. INDEX - A permuted title indexing program.
4. TWOUP - A second printing program which prints data on two column page format.
5. THREE (&) FIVEP - In combination, they print 3 × 5 cards.
6. FOURD - Allows further refinements in producing 3 × 5 cards.
7. SUPVS - The basic function of SUPVS is to control the loading and unloading of the other programs.

V. Centralized Information Retrieval Systems

It has been evident for many years that no company can by itself adequately index, abstract, and disseminate all current and past literature pertinent to its operations. This is especially true in the pharmaceutical industry with its extensive interdisciplinary research. As a result, many new centralized retrieval systems are now in operation. And we must not overlook the many proposals for international cooperation in medical documentation services, some of

which will certainly materialize. On an industry basis, the American Petroleum Institute and the American Society for Metals have already established computerized information retrieval systems. Other established automated information retrieval systems are now available in the chemical-pharmaceutical areas.

Because of these developments, the best laid plans of a pharmaceutical information service for developing its own retrieval stystem are going to be influenced—perhaps even drastically reduced—as these new centralized services develop. More likely, however, it will be that the individual company will tailor-make these information services to meet its own information needs.

In the Abbott SIS the following automated centralized information retrieval systems will be influencing the development of its own retrieval systems:

1. Steroid Index of Patents

The U.S. Patent Office for several years now has been issuing a punched card index of all new steroid compounds reported in the literature and in U.S. patents. An electronic sorter is needed to use this service. No plans have been announced to transfer this operation to computer usage. But it is not inconceivable that this will be done since the file is now becoming unmanageable.

2. Information for Industry Uniterm Index to U.S. Chemical Patents

This service has been in operation for ten years. The basic service is a manual, book-type, uniterm index of all United States chemical patents. A computer tape index became available two years ago; thus both manual and computer retrieval methods are available. An improved computer search system which would incorporate all minor terms (terms used less than ten times in any one year) is being planned. If this should become available, it would greatly enhance the value of this service for pharmaceutical firms.

3. RINGDOC

This indexing-abstracting service of the pharmaceutical literature was started in July 1964 by Derwent Publications in London. It is perhaps the most advanced centralized information retrieval system now available. It provides 40,000 abstracts each year along with manual, punched card, and computer tape indexes. The present tape index contains only alphabetical index terms and bibliographic citations. But in October 1965, it is planned to add a detailed chemical coding system.

4. Chemical Titles

In January 1965, the American Chemical Society made available computer tape indexes for all entries in their publication, CHEMICAL

TITLES. The tape format and search programs are designed for an IBM 1401/1410 computer system. The tapes will enable users to make their own searches and enable us to provide a selective dissemination service. (This new service also provides for the alternative of having searches made in the offices of Chemical Abstracts in Columbus for those companies not having computer services.)

5. FARMDOC

For the past three years, an abstract-index service of pharmaceutical patents from the major countries of the world has been available from Derwent Publications in London. Both manual and punched card indexes are supplied. As this store of patents grows, it is likely that the punched card index will also be supplied on computer tape. This service has become so successful that the supplier is planning to extend his patent coverage into the areas of food and agriculture.

These five examples of centralized information retrieval services have been noted to illustrate how national information suppliers will be influencing the operations of company information services in the future. More and more, the company information service must be planning for reprogramming and "packaging" computer tape indexes for their own use, making use of computer service centers, and justifying the costs of national services in relation to company needs.

ILLINOIS UNIV.ORG.AB. I 6 4400 PO

Return within two days.

ZAUGG	
*FREIFELDER	
*CLOSE	
JOHNSON R P.	
SCHOCK	
COLE J W	
*COTA	
PIECZ	
SMART	
SHIPTON	

ILLINOIS UNIV.ORG.AB. 1 6 4400 PO

Return within two days.

ZAUGG	482	
*FREIFELDER	482	
*CLOSE	482	
JOHNSON R P	482	
SCHOCK	480	
COLE J W	482	
*COTA	453	
PIECZ	453	
SMART	453	
SHIPTON	453	

ABBO
*Pass
Do not

ABBOTT LIBRARY
*Pass Journal to next person.
Do not clip or mutilate Journal.

Figure 1
Computer Produced Journal Circulation Lists
and Charge Records

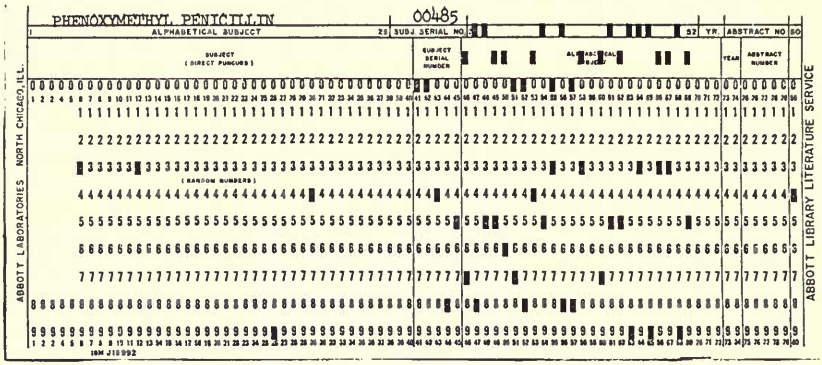


Figure 2
Index Term Card for Phenoxymethyl Penicillin

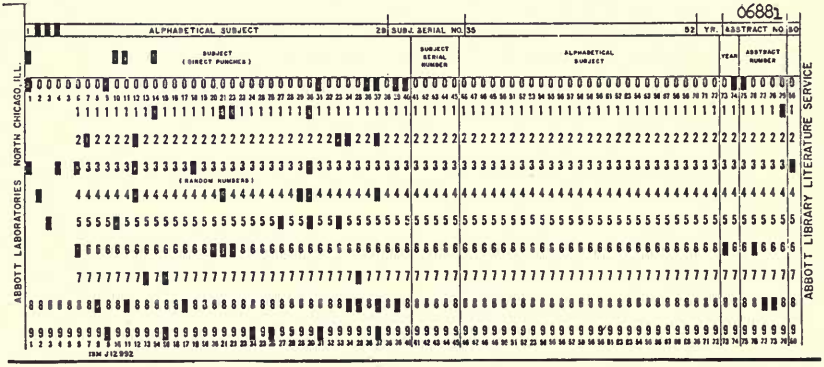


Figure 3
Search Card. The Term "Phenoxymethyl Penicillin"
is Superimposed on Other Index Terms

```
                                                                    PAGE      6

RANDOM NUMBER        DICTIONARY TERM                                    SEQ NO
157 194 271 330 C      GIC, ACETYLCHOLINE-LIKE, AUTONOMIC AGENT, DIRECT-  A 11694
157 194 271 330 D      ACTING /3/ II $                                    A 11694

062 135 214 230      *ACETYLCHOLINESTERASE INHIBITOR /3/                  A 12024
062 135 214 230 B      $ ACETYLCHOLINESTERASE INHIB., ENZYME INHIB., NICOTINE-  A 12024
062 135 214 230 C      LIKE, CHOLINERGIC, MUSCARINE-LIKE, AUTONOMIC       A 12024
062 135 214 230 D      AGENT $$ ENZYME INHIB., ACETYLCHOLINESTERASE INHIB. $  A 12024
062 135 214 230 E      $ SEE SPECIFIC AUTONOMIC ACTIONS II $$ GANGLIONIC  A 12024
062 135 214 230 F      BLOCKING, AUTONOMIC AGENT $$ SPASMOGENIC /MUSCLE SKEL./  A 12024
062 135 214 230 G      II $                                               A 12024

                       ACHALASIA    SEE    EXOPHAGUS DIS. /539/ /6/       A 12352

                       ACHLORHYDRIA    SEE    STOMACH FUNCTION DIS. /544.0/ /6/  A 12678

053 067 171 177      *ACHROMOBACTER /9/                                   A 13001
053 067 171 177 B      BACTERIA /9/                                       A 13001

109 176 190 354      *ACIDIFIER /3/                                       A 13322

012 042 058 165      *ACIDITY /4/                                         A 13640
012 042 058 165 B      PH, PROPERTIES, CHEM., PROPERTIES, PHYS., /4/ /SEE  A 13640
012 042 058 165 C      ALKALINITY, II, NEUTRALITY II/                     A 13640

                       ACIDOPHILE    SEE    EOSINOPHILE /5/               A 13956

292 311 339 349      *ACIDOSIS /788.6/ /6/                                A 14270
```

Figure 4
Biological Dictionary. A Computer Produced Dictionary Used in Connection with Abbott Abstracts

GENERAL REFERENCES

Griffin, H. L. EDP Procedures in Technical Library Operations. IDO-16881. Phillips Petroleum Co., Atomic Energy Div., Idaho Operations Office, AEC, 1963.

IBM DSD Technical Information Center. Selected Papers on an Integrated System for Disseminating, Storing, and Retrieving Information. TR 00. 1103. Poughkeepsie, New York, IBM Development Laboratory, Data Systems Division, 1964.

IBM Data Processing Division. Index Preparation and Library Processing at Monsanto Chemical Company's Research Center. White Plains, New York, IBM Data Processing Division, 1964.

Nicolaus, John J. The Automated Approach to Technical Information Retrieval; Library Applications. Washington, D.C.,

Department of the Navy, Bureau of Ships, (For sale by the Superintendent of Documents, U.S.GPO) 1964.

Southern, Walter A. "Mechanized Processing and Retrieval of Bio-Medical Information," Methods of Information in Medicine, 1:16-22, January 1962.