(6) Wall, E., "Final Report—First Revision of the Thesaurus of ASTIA Descriptors," AD 278168, Armed Services Technical Information Agency, 1962.
(7) Wall, E., "Study of Engineering Terminology and Relationships Among Engineering Terms," AD 432231, Engineers Joint Council, New York, N. Y., August 1, 1963.
(8) Wall, E., "Information Retrieval Thesauri," Engineers Joint Council, New York, N. Y., November 1962.
(9) "Chemical Engineering Thesaurus, "American Institute of Chemical Engineers, 1961.
(10) "DDC Authorized Descriptors," Defense Documentation Center, July 1964.
(11) "Euratom, Thesaurus," European Atomic Energy Community, Euratom, 1964.
(12) "Thesaurus of Descriptive Terms and Code Book," 1st. ed., Department of the Navy, Bureau of Ships, December 1963.
(13) "Thesaurus of Engineering Terms," 1st. ed., Engineers Joint Council, May 1964.
(14) "Thesaurus of Descriptors," tentative edition, U. S. Department of the Interior, Bureau of reclamation, October 1963.
(15) "Thesaurus of ASTIA Descriptors," Armed Services Technical Information Agency, 1960.

Biomedical Information Retrieval: A Computer-Based System for Individual Use

C. N. GILLIS*
Department of Pharmacology, Yale University School of Medicine, New Haven, Connecticut

Received December 8, 1966

A computer-based system has been developed for the retrieval of information from individual collections of abstracts, reprints, and other information relevant to individual research interests. The system and its operation are described.

The problems involved in storage and retrieval of biomedical information are receiving increasing attention from specialists in the field. There exist currently several centralized, computer-based systems that offer, on a service basis, access to the current literature. To use such services—for example, MEDLARS: Medical Literature Analysis and Retrieval System—requests for searches of the stored information must provide key words or phrases relevant to the subject; the occurrence of these in the title of a paper results in the title and journal reference being printed as output. Although difficulties exist in adequately indexing papers for such systems (*1*), they can be invaluable in searching the literature for information on a particular topic.

Much more frequently, however, investigators refer to their own collection of index cards, reprints, and notes, based on papers that are relevant to their particular interests. Over a period of several years, ready and reliable access to specific information in such collections becomes more difficult to achieve. Many people attempt some form of filing of index cards based on broad subject categories. Another widely used system employs cards whose edges are punched at locations corresponding to numbers assigned to topics mentioned in the paper. This system allows reasonably adequate cross referencing of articles, but again a problem arises in searching adequately through an ever-increasing number of such cards. Over the past few years, the author has used a system utilizing the edge-punched cards referred to above. This paper describes a recent adaptation of the system that allows computer-aided searching of the stored information.

Papers read are assigned an appropriate selection of up to 10 primary, 10 secondary, and six tertiary two-digit numbers, each corresponding to one of the topics listed in Table I. The primary topics encompass broad areas of interest, while the secondary and tertiary topics are those dealing with (at least for the author) increasingly specific subject material. The list of topics (Table I) and their corresponding numbers, although quite large is, in practice, easily committed to memory with continued use. As papers are read, notes may be made on index cards which are numbered consecutively and filed in numerical order. The complete information on each paper (termed the reference set) consists of the numbers categorizing the paper, its title, authors, and journal reference. Each reference set is typed on a coding sheet together with the same number as the corresponding index card with notes made at the time the paper was read. If no card was prepared when the paper was read, reference sets

* Work done during the tenure of an Established Investigatorship of the American Heart Association. The cost for development of the computer program was borne by a grant from the Natural Science Foundation (GP-4774) to Yale University. Other expenses were met by Grants H-7249 from the National Heart Institute and O-65-46 from the Life Insurance Medical Research Fund.

Table I. Subjects and Corresponding Numbers

Primary Subjects	No.	Secondary Subjects	No.	Tertiary Subjects	No.
adrenergic	01	age	14	adrenal medulla	04
brain	03	amines	03	aorta	05
cardiovascular	07	anesthetics	16	hypertension	14
cholinergic	04	antagonists	10	nerve stimulation	03
clinical pharmacology	25	anoxia	04	nictitating membrane	02
electrophysiology	05	autoradiography	06	shock	09
electropharmacology	05	binding	01		
enzymes	06	biosynthesis	09		
heart	07	cardiac glycosides	01		
kinetics	18	catabolism	09		
muscle (noncardiac)	08	denervation	05		
nerve	09	depletion	01		
psychopharmacology	10	effects of drugs on	02		
reviews, books, monographs	11	electron microscopy	06		
techniques	12	embryonic	14		
spleen	08	excretion	09		
vascular	21	fluorescence	06		
urine and blood	16	ganglia	07		
molecular pharmacology	18	histochemistry	06		
kidney	22	hormones	08		
		hypoxia	04		
		hypothermia	04		
		inhibitors	10		
		ions	11		
		necrosis	04		
		phosphorylase	13		
		potentiation	10		
		release	01		
		subsensitivity	05		
		subcellular (particles, distribution)	12		
		supersensitivity	05		
		tachyphylaxis	05		
		transport	11		
		uptake	01		

are assigned, consecutively, numbers greater than 50,000. When a sufficient number of coding sheets has been accumulated, the information is transferred to standard 80-column IBM punch cards by key punch operators, according to format given in the coding sheet. All information is then transferred to magnetic tape, using an IBM 1401 computer. Subsequent addition of new reference sets (on punched cards) to the tape is performed during a routine search (see below). It is apparent that the only part played by the investigator in the above sequence of events is to categorize the papers as they are read.

The computer program is written in Fortran, since this symbolic language may be used on computers of different manufacture. (Copies of the computer program listing and coding sheets are available on request.) The computer used currently at the Yale Computer Center, and for which the program was designed, is the IBM 7094/7040 Direct Coupled System. A maximum of four searches through the tape-stored information can be carried out simultaneously, the lists of papers resulting from each search being printed separately during output. Each search consists of a comparison of the category numbers of every reference set with the numbered categories required by that search. Up to nine number categories (three primary, three secondary, and three tertiary) can be included in each search. In practice, it has seldom been found necessary to use the maximum of nine categories allowed. Only those papers whose categorization numbers correspond to those sought in the search will be printed. The program will also add new reference sets to the tape during a routine search; these are also included in the search. The tape used currently by the author contains approximately 1300 reference sets; four simultaneous searches of this information take about 60 seconds.

A sample of the output of a single search, in this case for papers dealing with techniques involving nerve stimulation, amines, and vascular smooth muscle, is given in Figure 1. The availability, in the author's files of reprints of certain articles is indicated, as is the "card number," corresponding to the file card with notes made at the time the paper was read. As mentioned earlier, card numbers greater than 50,000 indicate that no notes were made when the paper was read. Reprints obtained are filed alphabetically under the name of the first author. The first such file is designated "R;" subsequent files are designated R-1, R-2, etc.

With the appropriate numbers, it is possible to program the computer to print stored information on papers dealing with very specific topics. Thus, the output resulting from a search for articles having the following combination of category numbers: (Primary) 01, 07; (Secondary) 03, 08; and (Tertiary) 03, lists all papers read that dealt with these topics: hormones, adrenergic nerve stimulation, and heart (Table I). The inclusion of number 11 in the primary category would result in printed output only of review articles dealing with this topic.

```
PRIMARY NUMBERS= 12, 21 AND  0
SECONDARY NUMBERS=  3,  0 AND  0
TERTIARY NUMBERS=  3,  0 AND  0

REFERENCE FORMAT--(YEAR),VOLUME,PAGE(S)

CARD NUMBER                                                                  REPRINT FILE

     45  RESPONSE CHARACTERISTICS OF PERFUSED MICROVESSELS TO PRESSURE AND VASOAC      R
         TIVE STIMULI
         BALY S.    ANGIOLOGY (1961) 12, 452
    531  FATE OF H3 NE IN SKELETAL MUSCLE BEFORE AND FOLLOWING SYMPATHETIC           R-1
         STIMULATION
         ROSELL S., KOPIN I. J., AXELROD J.    AMER J PHYSIOL (1963) 205, 317
    608  ACTION OF TEA CHLORIDE ON THE SYMPATHETIC GANGLION-PULMONARY ARTERY
         PREPARATION
         BEVEN J. A.      J PET (1963) 140, 193
    630  REGULATION OF CORONARY BLOOD FLOW                                       R-1
         BERNE R. M.      PHYSIOL REV (1964) 44, 1
    664  SOME CHARACTERISTICS OF THE ISOLATED SYMPATHETIC NERVE-PULMONARY ARTERY
         PROPERTIES OF RABBIT
         BEVAN J. A.    J PET (1962) 137, 213
    685  RESPONSES OF ISOLATED PERFUSED ARTERIES TO CATECHOLAMINES AND NERVE      R-2
         STIMULATION
         ROGERS L. A., ATKINSON R. A., LONG J. P.
         AMER J PHYSIOL (1965) 209, 376
    696  THE EFFECT OF SYMPATHETIC NERVE STIMULATION ON VASOCONSTRICTOR RESPONSES R-1
         IN PERFUSED MESENTERIC BLOOD VESSELS OF THE RAT
         MC GREGOR D. D.      J PHYSIOL (1965) 177, 21-30
    698  THE RESPONSE TO TRANSMURAL STIMULATION OF ISOLATED ARTERIAL STRIPS AND   R
         ITS MODIFICATION BY DRUGS
         PATERSON G.       J PHARM LOND., (1965) 17, 341-349
  50005  NEUROHUMORAL MECHANISMS AND VASCULAR SMOOTH MUSCLE - A REVIEW
         CIRC.RES.(1966)18, SUPPL. NO. 1
    974  NORADRENALINE STORAGE IN SKELETAL MUSCLE                                R-1
         SEDVALL.    ACTA PHYSIOL. SCAND.  (1964) 60, 39'50
    977  EFFECT OF VASOCONSTRICTOR TONE ON PRESSURE RESPONSES TO EPINEPHRINE AND R-1
         NOREPINEPHRINE
         KENDRICK.    AMER. J. PHYSIOL.  (1965) 208, 1000'1005
   1082  SEPARATION OF RESPONSES OF ARTERIES AND VEINS TO SYMPATHETIC STIMULATION
         ZIMMERMAN   CIRC. RES.  (1966)  18, 429
  50112  SMALL AND LARGE VEIN RESPONSES TO SYMPATHETIC STIMULATION (NS) AND
         TO NOREPINEPHRINE (NE)
         ABBOUD AND DEUTSCH   FED. PROC.  (1966)  25, 510
```

Figure 1. Computer printout of references on nerve stimulation, amines, and vascular smooth muscle.

There are several advantages of this system, aside from the easy and rapid retrieval of information. Since papers are categorized at the time they are read, a deeper and more detailed coverage can be made of the information of particular interest to the reader. Thus, while all papers describe methods used, it is possible, by appropriate use of the primary category number 12, to list for subsequent output only those dealing with techniques of special interest. For example, the search involving the numbers: (Primary) 03, 12; and (Secondary) 03, 12, would yield all papers read dealing with new or modified techniques for studying the subcellular distribution of biogenic amines in brain (Table I). On the other hand, papers describing results using established techniques would not have a primary category of 12.

In many instances, the titles of papers cannot accurately reflect all subjects dealt with in the paper. However, an advantage of the number classification system is that it becomes possible to categorize an article which has no formal relation to, for example, hypertension [(tertiary) 14], but which nevertheless may be a link in considering a particular research problem involving hypertension.

Another decided advantage of this system is that the basic collection of reprints, cards, etc. is "untouched" by the search. Thus, there is no possibility of losing a reprint or card.

The system described here is clearly only as good as the categorization made when the paper is read. Therefore it is best to develop gradually a system of numbered categories reflecting one's own interests. While developing the system it is of course possible to carry out searches by hand provided the edges of the cards (each of which has all the information comprising one reference set) are punched at locations corresponding to the numbers assigned to chosen topics. When it is clear that the system chosen is satisfactory and that no papers are "lost" through incorrect or incomplete categorization, the information can be transferred to punched cards which are numbered consecutively, in the order in which they are currently filed.

LITERATURE CITED

(1) Kent, A., *J. Am. Med. Assoc.* **196**, 927 (1966).