

In-Depth Indexing of Monograph Literature for an On-Line Retrieval System:

A Pilot Project*

BY JANET EGELAND, *Director of Educational Services*

*SUNY Biomedical Communication Network
Syracuse, New York*

ABSTRACT

One of the unique features of the State University of New York (SUNY) Biomedical Communication Network is the availability of depth-indexed monograph information for on-line retrieval. Approximately 8,000 titles from the Upstate Medical Center Library collection in Syracuse, New York, were indexed in depth, chapter-by-chapter, as to subject content using the control vocabulary *Medical Subject Headings*. Detailed indexing of monographic conference proceedings was considered of particular importance to the project. This attempt to make monograph literature more readily accessible to medical library patrons is interesting from both a philosophical and a technical point of view. This paper will discuss both of these aspects, giving an overall description of the nature of this innovative project.

BACKGROUND

THE fact that medical professionals rely more heavily upon journals than books to satisfy their information needs is well known among medical librarians. The actual content of medical libraries attests to this fact, with journals making up nearly two-thirds of the collection in many medical libraries.

The reasons for this emphasis on journal information are obvious and need little discussion. Journals offer the most current information and, most importantly, this information is made easily accessible by the many excellent indexing and abstracting services such as *Index Medicus* and *Excerpta Medica*. These services make it possible for users to locate recent articles on a particular subject or by a particular author without having to search the tables of contents or the indexes of the many journals

* Presented at the Seventieth Annual Meeting of the Medical Library Association, New York, New York, June 3, 1971.

which might contain this information. *Index Medicus*, for example, provides detailed subject access to the major medical and biomedical journals by depth-indexing all of the articles in these journals as to subject content. *Excerpta Medica* includes abstracts that the reader may scan to get an idea of the specific subject matter of the articles abstracted. Similar services are provided by *Chemical Abstracts* and *Biological Abstracts*, and *Current Contents* does a good job of indexing in the life sciences.

Unfortunately, monograph information is not nearly so accessible. One of the major indexes to book literature is the card catalog. This tool provides access to the books in a particular library by author, title, and subject. If the reader knows the author or title of a book, he can quickly locate the book by checking the card catalog. If, however, he is looking for information on a particular subject his task is not this simple. The subject index of the card catalog in most medical libraries is organized in terms of either *Medical Subject Headings (MeSH)*, or *Library of Congress Subject Headings (LC)*, or a combination of these two controlled subject classifications. Some libraries have their own unique subject classifications or employ the Cunningham, Boston, Dewey, or similar specialized subject classifications. In any case, the books are classified only according to their general subject matter without regard for specific subject content of the various chapters or subdivisions. A book entitled *Drug Trials for Headache* would probably be found in the card catalog only under the subject heading HEADACHE—DRUG THERAPY (using the *MeSH* classification). The reader who was interested specifically in the dibenzazepine drugs for headache would then have to go to

IN-DEPTH INDEXING FOR ON-LINE RETRIEVAL

the shelf, get the book, and examine either the contents page or the index to determine whether or not this specific drug was discussed anywhere in the book. Even though there is a *MeSH* subject heading DIBENAZEPINES available, the book would not be listed under this subject in the card catalog unless the discussion of the use of this drug for headache represented the major content of the book. Some of the cards in the catalog may contain detailed contents notes, but this type of help is not available for all the titles. For these reasons, the card catalog alone is not adequate for locating information on specific or detailed subject requests.

Other subject indexes to books such as the *National Library of Medicine Current Catalog*, where the reader may find books by general subject, and the *Subject Guide to Books in Print*, which lists books under general subject, are valuable tools. However, none of these tools provides the answer to the problem of trying to locate all the information on a very specific subject or on the interrelationship between subjects that may be discussed within the contents of a book.

In 1964, the Columbia-Harvard-Yale Project took a step in the direction of an answer to this problem by having the catalogers at the Yale Medical Library increase the depth of their subject cataloging, assigning an average of 10.4 *MeSH* subject headings to the monograph titles being cataloged (1). This gives the reader many more points of access to a particular book than are available as a result of conventional cataloging procedures.

However, providing a detailed subject index to monograph literature requires that the monographs be depth-indexed chapter-by-chapter and assigned subject descriptors in much the same way that each issue of a journal is indexed article-by-article by the journal indexing services. These subject descriptors must then be correlated with the specific part or parts of the book to which they apply to eliminate the problem of false coordination of subjects. For example, if KIDNEY DISEASES is the subject of one chapter of a book and BACKACHE the subject of another chapter in the same book but there is no common discussion of the two subjects, then the person who is looking for information on the relationship between KIDNEY DISEASES and BACKACHE might be misled into thinking that the book did contain this

kind of information. Further, this information must be converted into machine readable format so that a computer generated index could be printed and/or the data made accessible via a computerized information retrieval system.

The opportunity to undertake a project of this dimension was presented with the inception of the State University of New York (SUNY) Biomedical Communication Network in 1965 (2). While the largest data base of the SUNY Network, the first on-line retrieval system for medical literature, would be five years of the MEDLARS journal data obtained on tape from the National Library of Medicine, it was one of the primary objectives of the Network to provide the same kind of access to monograph literature. Thus it was decided that, as a pilot project, the monograph collection of the Upstate Medical Center Library would be indexed in depth and made available for searching by users via the SUNY Network on-line terminals.

PROJECT DESCRIPTION

Several basic policy decisions were made before the indexing actually began. The difficulty of drawing a representative sample of the library collection resulted in the decision to depth-index all books in the collection with a publication date of 1962 or later, excluding works such as textbooks, laboratory manuals, atlases, and directories. (Textbooks and lab manuals are, by their very nature, expected to cover all the relevant aspects of a general subject and were thus not suited to depth analysis.) The rest of the collection would be indexed in depth by breaking down the content of each book into any number of meaningful subdivisions, depending on the unique nature of the subject matter. In most cases, this breakdown would follow the actual chapter division of the book as determined by the author, but the indexers would have the freedom of combining two or three chapters that contained identical information or of further subdividing a chapter if several distinct subjects were discussed within it. All papers in volumes of conference proceedings or symposia were to be indexed individually.

Medical Subject Headings (MeSH) was chosen as the control vocabulary for the indexing project, with the stipulation that LC subject headings could be assigned when the *MeSH* vocabulary was found inadequate by the in-

JANET EGELAND

dexers. It was anticipated in advance that the need for LC subjects would be particularly important when indexing the behavioral science

portion of the collection. It was also planned that indexers could institute new subject headings if the subject to be indexed was not repre-

	INDX	NLM	STXT	KP1	KP2	EDIT	UPDT
Tag Description	JE		VJM		√SL		

020	Supp.	Type of Entry	Listing	Illus.	Map	Form	Level
		A Personal F Uniform Ⓢ Title Z Corporate	X Bib D Rep B Abs E Cat C Ind F Dir G Atlas	X Present Ⓐ Mono dr B Col dr C Mono ph D Col ph	X Map A Port B Map & Por	⊗ Conf A Essay B Outline C Thesis D Q & A	A Pop B Text C Lab

030	Language			Publication				
	Class	Lang 1	Lang 2	Key	Date 1	Date 2	Place	Publisher
	S	ENG		S	1968		MDBA	WW

040	Edition		Copy Statement		
	No.	Word	U	D	H
			1		

05 WG 300\$ 20 ACUTE myocardial infarction: \$ proceedings of a symposium sponsored by the University of Edinburgh in September 1967; 25 edited by D. G. Julian and M. F. Oliver. 30 Baltimore. % Williams and Wilkins, % 1968. %
 40 xvi, 343 p. illus.
 65 SBN 443-00587-7.
 70 Myocardial Infarct—congresses.
 72 Edinburgh. % University. 71 Julian, D. G., % ed. % \$ Oliver, M. F., % ed. % iv.
 74 Symposium on acute myocardial infarction.

01 70-360140

- Tag Description
- 000 Document Number
 - ⓪10 LC Call Number
 - ⓪50 UMC Call Number
 - 060 DMC Call Number
 - 070 HSL Call Number
 - 080 UR Call Number
 - 090
 - 100 Main Entry
 - 110 Author Sort
 - ⓪200 Title
 - 210 Title Sort
 - ⓪250 Remainder
 - ⓪300 Imprint
 - ⓪400 Collation
 - 500 Series—add
 - 510 Series—no
 - 600 Notes—store
 - ⓪650 Notes—no
 - ⓪700 Subject Tracing
 - ⓪710 Personal Tracing
 - ⓪720 Corporate Tracing
 - 730 Uniform Tracing
 - ⓪740 Title Tracing
 - 750 Series Tracing
 - 940 Library of Congress Suffix
 - 950 UMC Accession Number
 - 960 NLM Citation Number

FIG. 1.—State University of New York Biomedical Communication Network Conversion Input Form.

IN-DEPTH INDEXING FOR ON-LINE RETRIEVAL

sented adequately in either *MeSH* or *LC*. This was, however, to be done with much discretion and only after the proposed new heading had been discussed and approved at the monthly meetings of indexers. The indexers were required to precede all except *MeSH* headings with a special code so that an analysis of the

need for vocabulary supplementation in a project of this type could be made at a later time. The codes were as follows:

- 2—LC heading
- 3—Heading originating at Upstate
- 4—Author name (for conferences)

The indexing of all monographs was to be

MAIN HEADING	*subheading	CHAPTER NOS.	CHECK TAGS	CHAPTER NOS.
PREGNANCY			CASE REPORT	
MYOCARDIAL INFARCT		0	CLINICAL RESEARCH	
ACUTE DISEASE		0	HUMAN	0
CORONARY CARE UNITS		1,2,3,4	FEMALE	
MELTZER, L.E.		1,11	MALE	
SLOMAN, G.		1,10	INFANT, NEWBORN - 1 mo	
LOWN, B.		1	INFANT 1 - 23 mos	
SHILLINGFORD, J.P.		1	CHILD, PRESCHOOL 2-5 yrs	
KILLIP, T.		1,2,9	CHILD 6 - 12 yrs	
MCDONALD, L.		1	ADOLESCENCE 13-18 yrs	
JULIAN, D.		1	ADULT 19 - 44 yrs	
CROSS, E.B.		1	MIDDLE AGE 45-64 yrs	
ORGANIZATION AND ADMINISTRATION		2	AGED 65 +	
HOSPITAL ADMINISTRATION		2	ANIMAL EXPERIMENTS	6
EDUCATION, NURSING		2	CATS	
EDUCATION, MEDICAL		2	CATTLE	
FROMMER, P.L.		3	DOGS	6
EQUIPMENT AND SUPPLIES		3,4	GUINEA PIGS	
HOSPITAL DESIGN AND CONSTRUCTION		3	MICE	
MONITORING SYSTEMS		3	MONKEYS	
SAMUEL, E.		4	RABBITS	
RADIOLOGY		4	RATS	
MOUNSEY, P.		5	IN VITRO	
ARRHYTHMIA *complications		5	COMPARATIVE STUDY	
ARRHYTHMIA *classification		5	HISTORY OF MEDICINE	
ARRHYTHMIA *therapy		5	BIOGRAPHY	
MYOCARDIAL INFARCT *complications		5,7	HISTORICAL WORK	
PROGNOSIS		5	Ancient	
LOWN, B.		6	Medieval	
VENTRICULAR FIBRILLATION		6	Modern	
ELECTRIC STIMULATION		6	15th Cent	
BOUVRAIN, Y.		7	16th Cent	
TACHYCARDIA *complications		7	17th Cent	
LAWRIE, D.		8	18th Cent	
VENTRICULAR FIBRILLATION *classification		8	19th Cent	
RACIAL AND ETHNIC GROUPS			20th Cent	
ARRHYTHMIA *drug therapy		9,10	1901 - 1919	
MYOCARDIAL INFARCT *drug therapy		10	1920 - 1939	
PROPRANOLOL *therapeutic use		10	1940 - 1949	
GEOGRAPHIC HEADINGS			1950 - 1959	
DIPHENYLHYDANTOIN *therapeutic use		10	1960 -	
HEMODYNAMICS		10		
QUINIDINE *therapeutic use		11		
QUINIDINE *blood		11		

NOTE: INDEXING WAS CONTINUED ON ADDITIONAL SHEETS

CHAPTER INDEX			
01: 3-44	11: 123-25	21: 229-32	
02: 47-52	12: 131-7	22: 237-42	
03: 57-62	13: 138-46	23: 243-50	
04: 65-9	14: 154-60	24: 251-8	
05: 73-9	15: 169-73	25: 266-71	
06: 84-90	16: 179-83	26: 277-82	
07: 94-7	17: 191-5	27: 285-92	
08: 97-102	18: 201-8	28: 299-303	
09: 106-13	19: 208-14	29: 304-9	
10: 115-22	20: 217-24	30: 313-16	

Fig. 2.—Depth Indexing Form

done in accordance with the principles set forth by the National Library of Medicine in the *MEDLARS Indexing Manual* (3). The *MEDLARS Manual* and all other indexing instructions and aids distributed by NLM were to be used as the authoritative sources for all indexers in the proper use of *MeSH* headings and sub-headings and in interpreting the indexing policies applicable to the fourteen broad categories into which the subject matter is divided.

A two-part Conversion Input and Indexing Form was designed for use in the indexing project. The first sheet (Figure 1) contains a xerox copy of the catalog card for the book to be indexed. The format of this sheet is basically the same as that used by the Library of Congress MARC I Project. The card is coded by the indexer according to the tags listed under the column entitled Tag Description. The various boxes at the top of the sheet contain basic descriptors which are to be circled by the indexer if applicable to the book, as well as the coded language and publication information. The in-

put of all this coded information into the computer system enables the user to retrieve the book by many different bibliographic information fields and to restrict the nature of the book by the descriptors listed at the top. The second sheet (Figure 2) is the depth-indexing form and is similar to that used by NLM in journal indexing. Here the indexer lists all the subject heading descriptors and the portion of the book to which they apply. If a subject heading is descriptive of the entire book, this is indicated by putting a zero in the chapter column. The indexing form contains a special column called Check Tags. These check tags are commonly used restrictors such as AGE GROUPS, HUMAN, and ANIMAL EXPERIMENTS that are to be routinely considered for every book or segment within a book. In the example shown in Figure 2, the check tag HUMAN was given a zero, indicating that it was applicable to the entire book. Notice, however, that Chapter 6 (p. 84-90) did also contain information specific to animals, dogs in particular. At the bottom of the

```

start
DO YOU WANT INSTRUCTIONS ON HOW TO ENTER A SEARCH?
no
INDICATE THE TYPE OF SEARCH YOU WANT PERFORMED.
  A. JOURNAL SUBJECT
  B. BOOK SUBJECT
  C. BOOK AND JOURNAL SUBJECT
  D. BOOK AUTHOR
  E. BOOK LOCATION

b
IF YOU WANT AN EXPLANATION ON HOW TO ENTER SUBJECTS,
TYPE IN THE WORD EXPLAIN.
OTHERWISE, TYPE IN THE FIRST SUBJECT AND, IF DESIRED, A SUBHEADING.
myocardial infarct
DO YOU WANT TO ADD ANOTHER SUBJECT TO THIS GROUP?
& propranolol
ANOTHER SUBJECT TO THIS GROUP?
no
DO YOU WANT TO BEGIN ANOTHER GROUP OF SUBJECTS?
no
DO YOU WANT TO RESTRICT YOUR SEARCH BY AUTHORS, DATE, LANGUAGE,
OR PLACE OF PUBLICATION?
no
ARE YOU INTERESTED IN QUALIFYING YOUR SEARCH BY LEVEL,
FORM, MAPS, PORTRAITS, OR ILLUSTRATIONS?
no
THE FOLLOWING CAN BE IDENTIFIED:
  A. ABSTRACTS
  B. REVIEWS
  C. BIBLIOGRAPHY
DO YOU WANT TO RESTRICT YOUR SEARCH TO ONE OR
MORE OF THESE CATEGORIES?
no
DO YOU WANT TO SEARCH RECENT ACQUISITIONS ONLY?
no
YOUR SEARCH IS BEING PROCESSED.
PLEASE TYPE IN YOUR NAME AND DEPARTMENT
john smith/medicine

```

FIG. 3.—The Programmed Retrieval Series

IN-DEPTH INDEXING FOR ON-LINE RETRIEVAL

sheet the indexer indicates the corresponding pages for each chapter or subdivision into which the book has been divided. In this way, the user will know not only that the book does contain information specific to his request but also exactly how many pages it covers. There is space for forty-one subject headings on one sheet, but in many cases additional sheets had to be attached to index those books that required many more subject headings.

In December 1966 fifty test monographs were indexed and the machine conversion and retrieval programs tested. The full-scale indexing operation got underway in January 1967 and continued through April 1969. During this period the indexing staff consisted of either two full-time indexers or one full-time and two part-time indexers. The indexing was done at the rate of approximately one book per hour or seven books per day per indexer, with an average of five subject headings assigned to each chapter or subdivision. Some books—essays, for example—required very little time to index, while such works as detailed research monographs and conference proceedings required

much more time than an hour to complete. The number of chapters designated per book ranged from 2 to 152.

By April 1969, approximately 7,200 titles had been indexed in this manner.

REEVALUATION OF INITIAL OBJECTIVES

During the process of indexing it was discovered that many books in the collection simply did not lend themselves to a detailed chapter-by-chapter analysis and could easily be handled as a single entity. In these cases, the indexer augmented the subject tracings on the catalog card by adding from five to twenty general headings on the indexing form and putting zeros in the chapter column. It also became evident that the actual quality of some of the books in terms of their value to the subject field was questionable. Rather than ask the indexer to make a subjective evaluation as to the value of a particular book, indexing remained non-selective until April 1969.

In April 1969 a decision was made to discontinue the nonselective indexing of the entire monograph collection. It was felt that the 7,200

```
OPTION NETBOOK1,STAT;  
K001 MYOCARDIAL-INFARCT;  
K002 PROPRANOLOL;  
K003 K001 & K002;  
LIST AUTHOR,TITLE,IMPRINT;  
END;
```

```
RESULT: 0000000001
```

```
9509      B0
```

```
AUTHOR:
```

```
TITLE:    ACUTE MYOCARDIAL INFARCTION: PROCEEDINGS OF A SYMPOSIUM SPONSORED  
          BY THE UNIVERSITY OF EDINBURGH, SEPTEMBER 1967, EDITED BY D.G.  
          JULIAN AND M.F. OLIVER.
```

```
IMPRINT:  BALTIMORE, WILLIAMS AND WILKINS, 1968.
```

```
IDNUMBR: 0010381
```

```
*
```

```
DO YOU WANT TO LOCATE ANY OF THE BOOKS WHOSE CITATIONS HAVE JUST PRINTED OUT?
```

```
yes
```

```
ENTER THE 2-DIGIT LIST NUMBERS OF THE BOOKS YOU WANT TO LOCATE.
```

```
A BOOK'S LIST NUMBER IS FOUND ON THE FIRST LINE OF EACH BOOK
```

```
CITATION AND BEGINS WITH THE LETTER B.
```

```
b0
```

```
B0      *WG 300 A 189 1967
```

```
*
```

```
EOS
```

FIG. 4—Search Output

completed documents were a more-than-adequate sample upon which to perform a project evaluation of this type of indexing effort. As a result of the factors discussed above with regard to the questionable value of nonselective indexing of all types of monographs, it was agreed that from this point on only conference proceedings and symposia would be indexed—again, as a pilot project. Conference proceedings are generally acknowledged to be of considerable importance in the research literature, and the fact that they are often the most difficult to locate unless the exact name of the conference is known gave further reason for extracting their specific content for literature searching. *The Directory of Published Proceedings* does list meetings according to title, sponsor, and general subject, but the specific content of all the individual papers is not described. From April 1969 to December 1969, 800 conferences and symposia were indexed in depth, with each individual paper treated separately. An initial analysis of this part of the project indicated that these monographic works had an average of twenty-three papers or sections and that the indexers assigned an average of seven subject headings to each paper.

The three major authors of each paper from a conference were also included on the indexing form so that all papers could be retrieved by author as well as subject, a capability considered to be especially useful.

RETRIEVAL PROGRAMS

The data resulting from this indexing in depth, including the conferences, has been key-punched and loaded into the Network searching system as a separate data base called Netbook 1. This data base can be accessed by either subject or author and is now available for on-line searching via the SUNY terminals in all of the seventeen member libraries.

In order to retrieve information from the Netbook 1 file containing the 8,000 depth-indexed monographs, a user sits down at a SUNY terminal and, with the aid of a programmed set of instructions, enters a search for the subject or subjects or for the author in which he is interested. The only restriction is that the subjects he chooses must be listed in *Medical Subject Headings*, the controlled entry vocabulary.

Figure 3 illustrates the retrieval process. Suppose, as in this example, that a user is interested

in citations on the drug propranolol as a therapeutic agent for myocardial infarction. He enters from *MeSH* the two subjects MYOCARDIAL INFARCT and PROPRANOLOL (both in lower case letters), indicating that he would like citations dealing with both these subjects. He is then given the opportunity to restrict his search in several different ways. After he has answered all the questions, there is a brief wait while his search is being processed. Then the book citations that satisfy his request are printed back to him at his input terminal.

Figure 4 shows the actual output from this search on MYOCARDIAL INFARCT and PROPRANOLOL. The search returned the depth-indexed conference *Acute Myocardial Infarction* shown in Figure 2 because the indexer had indicated on the indexing form that there was a discussion of the therapeutic use of propranolol in Chapter 10, p. 115–122. Notice that the user is also given the opportunity to find the location of the book if he is interested. If the user is in the Upstate Medical Center Library in Syracuse, then the call number for the book will be printed for him so that he can go directly to the shelf and not have to check the card catalog first. In all other Network libraries, the message BØ IS AT SYRACUSE will be printed out.

At the present time, the programs that will correlate subject headings with specific pages in monographs to which they apply are not completely operational. The user can get a printed list of books that refer to the specific subject or subjects he has asked for, but the pages are not yet given with the citations; he has to make a separate request for page numbers. For this reason, no formal evaluation of the project has been attempted. The final programming will be completed in the near future, at which time an evaluation will be made to determine the usefulness of this first new computerized means of access to depth-indexed monograph literature in the field of medicine.

REFERENCES

1. KILGOUR, FREDERICK G. Mechanization of cataloging procedures. *Bull. Med. Libr. Assoc.* 53: 152–162, Apr. 1965.
2. PIZER, IRWIN H. A regional medical library network. *Bull. Med. Libr. Assoc.* 57: 101–115, Apr. 1969.
3. CHAREN, THELMA. *MEDLARS Indexing Manual*. Bethesda, Md., National Library of Medicine, 1965–.