

## Experience with Microcomputer Management of a Personal Medical Literature Collection†

David G. Bjoraker, M.D.\*

*Microcomputers may be conveniently and effectively used to manage personal medical literature collections. A bibliographic data base consisting of the author, title, and source for each reference is entered and stored on magnetic tapes or disks. Document categorization and keyword assignment prior to entry is optional, thus allowing clerical personnel to maintain both the literature collection and the bibliographic data base. Subject searches using an array of search terms formulated by the medical user permit retrieval of relevant bibliographic citations. Current alphabetical and keyword-in-context listings of the documents in the collection may also be produced. The programming approach uses string processing and is within the capability of programmers with limited experience.*

Most physicians depend on their personal medical literature collection as a primary information source.<sup>1</sup> Computerized methods<sup>2-5</sup> of organization and retrieval of reprints have surpassed manual approaches in convenience and efficiency. In a survey by *Physicians Microcomputer Report*<sup>6</sup> personal literature files were ranked as the fifth most popular application of computers by physicians. However, the computers often used have been large and expensive data-processing installations<sup>2,3</sup> rather than the now popular desktop microcomputers.<sup>4,5</sup>

When microcomputers are used, compact structuring of data is essential. Premium random access memory and limited magnetic media storage capacity are optimally utilized only if the number of bytes that are used for formatting and processing or left unused are minimal. If sequentially read magnetic tape media are used for storage, then the time required to read data from tape to volatile memory will also affect usefulness. This interval must be no longer than necessary for the user's task to be performed efficiently.

---

*From the Department of Anesthesiology, University of Michigan Medical School, Ann Arbor, Michigan.*

\* Address all correspondence to Dr. David G. Bjoraker, 6307 Cocoa Lane, Apollo Beach, Florida 33570.

† This paper was presented at the First Annual Conference of the American Association for Medical Systems and Informatics, October 1982, and is published here with the permission of AAMSI.

A system is described that has been in use for 5 years and now manages four independent medical literature data bases. The bibliographic citation for each document in the collection is entered in a standard form, edited when necessary, and stored. The entire collection of citations may be printed according to a variety of formats or may be searched to retrieve selected citations.

## METHODS

### Hardware Characteristics

A Hewlett-Packard 9845B desktop computer was used to develop the system. This computer included a CRT display, an on-board 80 character-per-line thermal printer, and dual tape cartridge drives. The CRT and printer were essential, but there were several options for magnetic storage media. Only a single drive was necessary for tape cartridges, tape cassettes, or floppy disks, but with dual drives operation of the system was more convenient and storage media could be duplicated to prevent data loss from system failure. The computer's operating system allowed the entire content of a tape or disk to be located and read by the program without user intervention. The storage medium selected must be able to store at least 100,000 bytes of information. The capacity of the tape cartridges (Scotch DC-100) used for program and data storage was 216,000 bytes. The maximum volatile memory required at any time was 47,000 bytes, although 32,000 would have been sufficient if volatile memory was limited.

An enhanced BASIC language was used for programming. For a high level language to be used for this type of programming, efficient string processing commands, including a method to rapidly locate substrings, was necessary. The ability to lock out unused portions of the keyboard, automatic loading and running of a program with power-on, and a real-time clock were also desirable features. The high speed of a dot matrix printer was advantageous.

### Software Characteristics

All portions of the program were interactive. The user responded to prompts displayed on the CRT with a standard typewriterlike keyboard and did not need any programming knowledge. At power-on a task list was presented and a selection was made that corresponded to one of the following programs.

**Data Input Program.** The input for each document consisted of three elements: author, title, and source. The entry format was selected to be that of a major journal in the user's field. Detailed bibliographic formatting rules and examples were included in the program for display whenever the operator desired. Each element could contain up to 320 characters. To minimize memory requirements, trailing blanks were trimmed on entry. A one character end-of-element or end-of-entry marker was then added, and the string was concatenated to other entries to form a data string.

The initial nonnumeric characters for the source element were assumed to be the Index Medicus abbreviation of a journal. The characters were matched against a list of journal abbreviations concatenated into a string. If no match occurred, the operator was alerted so that incorrect abbreviations or spellings were identified and corrected. A sub-

sequent search for a journal was successful only if the correct abbreviation was originally entered. If the document source was not a journal, then the rejection of the source element could be overridden with a single key-stroke. The year of publication was included in the source element.

When the data string length of 21,200 characters was exceeded, it was stored on tape and a new data string was started. Each tape contained 10 data strings. A maximum of a 300-character reserve was maintained on each string to allow for some lengthening during subsequent entry editing. Routine duplication of data tapes for backup was mandatory since the time investment for data entry was great.

The selected length of the data string was dependent on the storage medium and the computer's operating system. The length selected for the 9845B was based on programming convenience when using the tape file catalog. For disk storage media, short data strings may minimize access times since a data string may be located directly without serially passing other data strings, as was done during transfer from a tape. The practical minimum data string length is determined by the maximum number of individual file locations that the operating system can catalog for the storage medium.

On occasion a document title did not contain useful keywords or was ambiguous or misleading. When this was noted, keywords were appended to the title element during entry, but separated from the title by a third marker character. The user of this system could elect to append the author-designated keywords routinely or they could be included as a fourth element of each entry. However, when a person knowledgeable in the subject area must review each article to designate keywords, the benefit of a system requiring only clerical skills for data entry is lost. One of the four current data bases had a category label appended to each title and the document was filed by this category. In the other data bases the documents were filed alphabetically by first author.

**Journal List Editing Program.** The journal list was the string of Index Medicus journal abbreviations<sup>7</sup> used in screening the source element during entry. The journal list editing program allows for additions, deletions, and corrections of journal abbreviations and for printing of the list. The abbreviations were also stored in the form of a concatenated string. The matching of the first portion of the source element to the journal list was done by scanning the journal abbreviation string for a matching substring. Journal abbreviations were added to the list only if the data bases included more than one article from the journal.

**Search Program.** All entries that contained a specific word or character group from any portion of the entry could be retrieved by the search program. Recovered entries were displayed on the CRT, displayed with operator option to print, or printed directly. Up to four words or word fragments could be linked such that a retrieved entry would contain all the terms in the set. The specificity of the search was thereby increased. These additional words were linked by the Boolean "AND" operator. Up to four sets of terms could be searched for simultaneously. The sets were linked by the Boolean "OR" operator. The Boolean "NOT" operator could also have been used, but it was not implemented. Its function would have been to limit retrievals in an already highly specific data base. Search printouts included the time and date of the search and the size of the data base when the search was conducted.

Ingenuity was rewarded when selecting search terms. For example "cardi" recovered entries containing both the words "cardiac" and "cardiovascular." Including a year or a journal as a search term further increased specificity of the list of documents re-

trieved. Preceding an author's name with the end-of-entry marker caused only citations for which the author was *first* author to be retrieved. Preceding the author's name with a space resulted in retrieval of citations in which the author was *other* than first author. Alternate spellings such as "anaesthesia" for "anesthesia" had to be included. Examination of retrieved articles frequently suggested additional terms for the search.

**Edit Program.** If the operator mistyped an entry, the line could be corrected until a carriage return committed the entered line to concatenation to the data string. Errors recognized later in the entry process could also be corrected immediately, since the location of the incorrect element was obviously near the end of the database. The entry program could scroll each element of each entry into position for correction or acceptance of its current form.

When an error was discovered during later printout or display of an entry, the storage location was generally not known. By entering a unique set of characters, all data strings were sequentially loaded and scanned until the entry containing the characters was located. The user then could eliminate or edit that entry or edit adjacent entries. If the user had prior knowledge of the tape and string location of the error, only that portion of the data base needed to be scanned. Capability to delete entries was not specifically provided, since editing an existing entry by replacing it with a new entry served the same function. Programming for this task combines the sequential loading and scanning of data strings used in the search program with the editing option of the data entry program.

**Printout Program.** A list of entries could be printed in bibliographic form according to one of the following orders: (1) order of entry into the data base, (2) alphabetically by first author, (3) alphabetically by title, or (4) alphabetically by source. Alphabetical sorting required reordering of the elements of the entry such that the element determining the order was first. Sorting required each entry to be compared with a previously ordered string of entries and inserted into the ordered string at the proper point.

For alphabetical sorting, a second machine with two additional tape drives was interfaced such that the program could access the entire data base without operator intervention. In the absence of sufficient drives to mount all magnetic media or sufficient volatile memory to store all data at one time, operator intervention during sorting was inevitable and alphabetical printouts were a time-consuming and frustrating task. All printouts were labeled with the date in order to allow subsequent identification of the most recent listing.

## Recent Modifications

The printout program had a fifth formatting option added, the IBM Keyword-In-Context (KWIC) Index.<sup>8</sup> This index listed a document alphabetically by each significant word in the title element. The example in Figure 1 illustrates that sufficient additional terms from the title are included such that the keyword at the beginning of the second column may be considered in context. The third column is the first author of the document. Since documents were filed by the first author, the user could retrieve documents on a subject without performing a computer search of the data base.

An additional program section for generation of reprint request cards was added. The user interactively entered the three elements of the bibliographic citation for the requested article together with the author's address. After approval of the entered data,

ERAPY DURING	CORONARY ARTERY SURGERY	*VASODILATOR TH	KAPLAN JA
INFUSION DURING	CORONARY-ARTERY SURGERY	*NITROGLYCERIN	KAPLAN JA
A AND MORTALITY IN THE	CRITICALLY ILL	*COLLOID OSMOTIC PRESSUR	RACKOW EC
NITORING GOALS FOR THE	CRITICALLY ILL PATIENT	*PHYSIOLOGIC MON	BLAND R
	CYANIDE POISONING - A CHALLENGE	*	BERLIN C
PULMONARY EDEMA	*ACUTE	CYANIDE POISONING COMPLICATED BY LACTIC	GRAHAM DL
		DEATHS (EDITORIAL)	-
	TIC	*ANAESTHETIC	SAVAGE BJ
ID-FILLED CATHETERS	*	DELIRIUM	BLANTZ SA
SURE-VOLUME		*THE PREVENTION OF POSTANESTHE	SABAWA KI
		DETERMINATION OF FREQUENCY RESPONSE FROM	BURKI NK
		DIAGRAM REVISITED	MILLER RD
		*THE VENTRICULAR PRES	THOMPSON DS
		DISTRESS SYNDROME	LAWSON DH
		*ADULT RESPIRATORY	GRAHAM DL
NISM BY NEOSTIGMINE	*	DIURETIC-INDUCED HYPOKALAEMIA PANCURONIU	RACKOW EC
GENERAL ANESTHESIA IN	*	DOGS: A NEW LOOK AT AN OLD TECHNIC	SHNIDER SM
THERAPY	*	*IN	SHNIDER SM
ACIDOSIS AND PULMONARY		DRUG THERAPY REVIEWS: INTRAVENOUS FLUID	LAGLIA AB
INDICATOR OF PULMONARY		EDEMA	
AND CHILDBIRTH: AN		*ACUTE CYANIDE POISONING COMPLICA	
ENDURING MARRIAGE	*	EDEMA AND MORTALITY IN THE CRITICALLY IL	
LIVER DISEASE	*SPINAL	ENDURING MARRIAGE	
		*EPIDURAL ANESTHESIA	
		EPIDURAL ANESTHESIA AND CHILDBIRTH: AN	
		EPIDURAL HEMATOMA AFTER LUMBAR PUNCTURE	

Figure 1. Example of the IBM Keyword-In-Context (KWIC) index based on significant words in the title element of the bibliographic citation. The author column permits direct recovery of the document. "\*" indicates the first word in the title.

a postcard (Request-A-Print, Institute for Scientific Information) was printed, and the bibliographic information was added to the data base with a notation indicating the date of the request. When the reprint arrived, the notation was removed, if the article was to be filed. The entire citation was removed, if the article was to be discarded. A separate, current list of pending reprints, which included the author address, was also maintained.

Since the system was first developed 5 years ago, the cost of fast, large-capacity storage devices has continued to decrease. Therefore, the system is now being installed on an IBM-XT computer, which includes a 10-megabyte hard-disk drive. In addition, the popularity of maintaining a bibliographic data base has increased to the level that several commercial software packages are now available (Table 1). The lower cost and increasing sophistication of data base software suitable for microcomputers enhances the possibility that a general-purpose system will perform many of the functions incorporated into the special-purpose system described here.

## RESULTS

The three largest data bases, which cover clinical anesthesiology, hemostasis, and virology, contained 11,678 entries. Seventy-six data strings contained a total of 1,566,590 bytes of information and 35,110 bytes of markers. Nine tape cartridges were required to store the data strings, three for anesthesiology, four for hemostasis, and two for virology. Table 2 presents detailed data for memory and storage requirements for the two largest data bases in use, anesthesiology (ANES) and hemostasis (HEMO). The breakdown of requirements by citation element demonstrated the difference in content of the bibliographic citations between a data base that was largely clinical medicine, ANES, and one that was largely biomedical research, HEMO. The clinical data base had shorter author and title elements.

For the three largest data bases, 370 journals of the 2,686 journals listed by Index Medicus were included in the journal abbreviation list.<sup>7</sup> Another 39 journals not listed

Table 1. Commercially Available Bibliographic Software<sup>a</sup>

Program	Vendor
1. BIBLIOFILE	Power Tools 48 Hunter's Hill Circle Amherst, MA 01002
2. BIBLIOGRAPHY	Digital Marketing 2363 Boulevard Circle, Suite 8 Walnut Creek, CA 94595
3. BIBLIOTEK	Scientific Software Products 3171 Donald Avenue Indianapolis, IN 46224
4. DOCUMENT RETRIEVAL SYSTEM	Information Solutions-Softrend Div. Post Office Box 198 Charlottesville, VA 22902
5. GOLDEN RETRIEVER	CLASS 1415 Koll Circle, #101 San Jose, CA 95112
6. PERSONAL BIBLIOGRAPHIC SYSTEM	Personal Bibliographic Software, Inc. Post Office Box 4250 Ann Arbor, MI 48106
7. QUICK-SEARCH LIBRARIAN	Interactive Microware, Inc. P.O. Box 771 State College, PA 16801
8. SCI-MATE PERSONAL DATA MANAGER	Institute for Scientific Information 3501 Market Street Philadelphia, PA 19104
9. SEARCHLIT	MediSoft 1595 Soquel Drive, Suite 350 Santa Cruz, CA 95065

<sup>a</sup> The author makes no claims concerning either the quality and reliability of the listed software or the completeness of the list.

by Index Medicus were also included. The storage of the journal abbreviation string required 6,026 bytes, including a 1-byte marker character separating abbreviations.

Transferring the data from an entire tape required 148 seconds, plus 19 seconds to rewind the tape for each of two tracks. A search of 1,500 entries (one tape) would require

Table 2. Observed Memory Requirements<sup>a</sup>

	ANES	HEMO
Total bibliographic entries	4,815	5,071
Total characters (bytes) stored <sup>b</sup>	597,951	724,840
Total data strings in use	29	35
Characters per entry	124.2 ± .6	142.9 ± .6
Characters per author element	26.5 ± .3	32.7 ± .3
Characters per title element	64.6 ± .4	78.5 ± .5
Characters per source element	33.1 ± .3	31.7 ± .3
Entries per data string	166.9 ± 1.7	145.9 ± 1.4
Journals cited more than once	214	244

<sup>a</sup> Mean ± standard error of the mean.

<sup>b</sup> End-of-element and end-of-entry markers excluded.

186 seconds data transfer time, 9 seconds processing time, plus .33 seconds format and display time for each recovered entry.

## DISCUSSION

Other computer systems, including microcomputer based systems,<sup>4,5</sup> have been described for the management of personal reference collections. However, the simplicity of programming and operation provided by string processing has not been exploited. The system of Betrand and Bader<sup>4</sup> requires that keywords be identified during entry and that subject searches be conducted with a set of defined keywords. Van Styvendaele's<sup>5</sup> system does allow subject searches with user selected strings. However, citation entry is complex since an authorship code, document accession number, subject code, and extensive content evaluation are required. Both microcomputer systems use low-volume storage media, a 256,000 byte floppy disk and a 300,000-byte tape, respectively. However, the structure of their stored data does not optimize utilization of the limited space, as does string concatenation with elimination of excess blanks.

Use of the system described has been a marked improvement over cumbersome manual techniques. Virtually unlimited cross-referencing of a document is possible, since it is not committed to any one subject category during entry. As interests change and the collection expands, redefining categories is not necessary. Useful, current document listings are readily available in correct form for inclusion in bibliographies. The retrieval of a reprint or other document, when only fragments of a complete citation are recalled, is possible with the search program. Also, instead of lending folders of reprints, either reprints *or* the printed bibliography from a search may be offered. This minimizes the risk of losing the original document.

Previous proposals for managing anesthesia literature required prior user effort to either abstract an article or categorize it before it was included.<sup>9,10</sup> In this system, maintenance of both the literature collection and the data base may be done by clerical personnel with neither medical nor computer knowledge. Medical personnel may then direct their efforts toward conducting an effective search and retrieval of relevant entries once a specific question is identified. A controlled vocabulary is not necessary since the medical user is able to structure an array of search terms using the most relevant synonyms, authors, and journals.

Although this system uses sequentially read tape cartridge magnetic media for storage, access time for searching and editing have not been a problem for the users. The search time is usually used to review documents retrieved earlier in the search or the search results are printed without the user being present. Disk storage and smaller data strings would improve access time considerably.

Many of the major problems associated with large on-line medical data bases such as MEDLINE are absent in this system. Since the thoroughness of the search and the relevance of the retrieved material are inversely related, intensive searches yield much irrelevant material.<sup>11</sup> In this system, presumably most documents are relevant to the user or he would not have ever collected them. Access protocols, command languages, and Boolean search strategies have also been noted to be problems.<sup>11</sup> However, providing the searcher with specific prompts during the search without using computer terminology

has minimized these problems. The use of a controlled vocabulary or referring to a thesaurus was not necessary.

The system described is a group of special-purpose programs to process a specific data set. However, the processing required could be accomplished by a data-management system using a flat-file data model, the most simple of data models. However, the simplicity of use and effectiveness with which the system accomplishes its described tasks has been rewarding. The equipment necessary and the programming required is sufficiently uncomplicated that the limitations of the flat-file data model and the fixed relationship of the data are easily offset.

## REFERENCES

1. Stinson, E.R., and Mueller, D.A., Survey of health professionals' information habits and needs: Conducted through personal interviews. *J. Am. Med. Assoc.* 243:140–143, 1980.
2. Gillis, C.N., Biomedical information retrieval: A computer-based system for individual use. *J. Chem. Doc.* 7:98–100, 1967.
3. Marks, R.H.L., A Fortran computer program for storage and retrieval of personal journal references. *Int. J. Biomed. Comput.* 12:283–290, 1981.
4. Bertrand, D., and Bader, C.R., Storage and retrieval of bibliographic references using a microprocessor system. *Int. J. Biomed. Comput.* 11:285–293, 1980.
5. Van Styvendaele, B.J.H., The making of an RNA phage bibliography with personal information retrieval by microcomputer. *Meth. Inform. Med.* 18:158–164, 1979.
6. *Physicians Microcomputer Report* 2(10):15–17, 1979.
7. *List of Journals Indexed in Index Medicus* 1982, National Library of Medicine, U.S. Dept. of Health and Human Services publication No. (NIH) 82–267, U.S. Government Printing Office, Washington, D.C., 1982.
8. Luhn, H.D., *Keyword-in-Context Index for Technical Literature (KWIC Index)*. IBM Advanced Systems Development Division, Yorktown Heights, New York, 1959.
9. Yeakel, A.E., An anesthesia literature abstracting and retrieval method (ALARM). *Anesthesiology* 25:760–766, 1964.
10. Petty, W.C., and Carden, W.D., A personal cross-reference filing system for anesthetic literature. *Anesthesiology* 38:498–510, 1973.
11. Doszkocs, T.E., Rapp, B.A., and Schoolman, H.M., Automated information retrieval in science and technology. *Science* 208:25–30, 1980.