

A Specialized Library Index Search Computer

B. KESSEL† AND A. DELUCIA‡

INTRODUCTION

THE NEED for mechanizing library information searches has become apparent during the past two decades. The phenomenal increase in the size and number of library establishments in conjunction with the requirement for greater speed in servicing information requests have been key contributory factors in focusing attention on this situation. A large variety of developmental and commercially available devices have been used for putting library card catalogs into forms more amenable to automatic searching. It was only a matter of time before many researchers became aware of the advantages of digital computers in mechanizing this operation. Government and commercial organizations, by writing new programs, were able to adapt those computers which were already available to them.

Early in 1958, a program was initiated by the Rome Air Development Center for the design and fabrication of an Index Searcher that was to be tailored specifically to the needs of library documentation. The Index Searcher was to contain only those logical functions that would be of value in library information searching. The Searcher was to be developed primarily as a research vehicle for use in studying various index and information retrieval approaches. The basic design was to include facilities to allow the Searcher to be used as a fully operational device. In April, 1958 a contract was awarded to the Computer Control Company with delivery of the Index Searcher to occur early in 1959.

Design concepts of the Index Searcher have evolved in part from a knowledge of the Minicard Selector.¹ However, the Index Searcher will be used in library situations where index data and graphics material are stored separately, rather than together as on Minicards.

The Index Searcher will search through large volumes of index data serially and print out the identification of those documents, reports, or graphic materials that satisfy the requirements of the search criteria. A consideration of the problems of library mechanization indicated that the following features should be included in the Searcher design: 1) high-speed searching of the index data; 2) the ability to reproduce all or part of the index library cheaply and quickly; 3) a minimum of delay in effecting the search beyond the setting up of the search criteria; 4) the capability for handling a wide variety of index and classification schemes; 5) ease of operation; 6) flexibility in permitting frequent updating of

the file; 7) search for more than one question at a time; and 8) a growth potential allowing for relatively efficient use of the Searcher either singly or in groups as the size of the library increased. The system design of the Searcher as a special purpose system, was to result in an information handling capability that could be matched in the general purpose computer field only by a considerably larger and more expensive machine.

GENERAL DESCRIPTION

The Index Searcher uses magnetic tape as the storage medium for index data. An index entry is made on the tape for each document, report, or other piece of physical material which is to be made available for rapid automatic searching. An entry might consist of anything from a title to a complete text, depending upon the storage and recovery system to be used. In typical applications the entry made for a given document will consist of a document number, title, author, date, and several descriptors that define the subject matter of the document.

The descriptors can be grouped into sets that are designated as phrases. The value of this feature can be illustrated by considering the indexing of a technical paper that describes a machine using transistorized logical circuits and a magnetic core shift register storage. Listing of just the four descriptors for transistorized, logic, magnetic cores, and shift registers could result in the false selection of this document during a search for magnetic core logical circuits. Use of phrase boundaries in the proper places assures that the descriptors will be properly associated with each other during the searches.

The standard machine word for the Searcher is 42 bits in length. This consists of seven alpha numeric characters of 6 bits each. The system also includes provision for handling double and triple length words so that it can accommodate clear text as well as coded index data.

A typical document index entry might consist of twenty machine words, making a total of 840 bits of information.

Search criteria are specified in terms of question words plus logical connectives to group the question words into question phrases and to group the question phrases into complete questions. The question words are stored in the internal memory of the Searcher. The memory has a capacity of 20 machine words.

The type of comparisons to be made between question words and the document index entry words is specified individually for each word by plugboard wiring. The specification can be for "equality," "less-than," "greater-than," or any combination of two of these types of comparison. For example, a question might specify that

† Computer Control Co., Inc., Framingham, Mass.

‡ Rome Air Dev. Center, Griffiss AFB, Rome, N. Y.

¹ J. W. Kuipers, A. W. Tyler, and W. L. Myers, "A Minicard system for documentary information," *American Documentation*, vol. 8, pp. 246-268; 1957.



a document is required on a particular subject that would be identified by descriptor equality comparisons; published after 1956, identified by a "greater-than" comparison; and with a security classification no higher than confidential, identified by an "equal-to or less-than" comparison.

The question words are grouped into question phrases by means of plugboard-connected logical circuits. Fifteen phrase elements are available for composing up to 15 different question phrases. Two or three phrase elements can be cascaded to make a larger phrase than can be handled in one logical element. Each phrase can use any desired combination of question words, in either assertion or negation form, as inputs, and each question word can be used in as many of the 15 different phrases as desired.

Complete search criteria questions are made up of question phrases by plugboard-connected select logical elements. The same flexibility exists here as in combining words into phrases. Ten question elements are provided, resulting in the ability of the Searcher to simultaneously search for documents meeting ten different search criteria.

Searching consists of scanning through the complete document index tape and comparing the contents of each index entry with the question words and logic stored in the Searcher memory and plugboard connections.

The result of a successful search is a print-out of the document numbers of those document index entries that have met the search criteria. An identifying number printed beside each document number shows which of the several search questions is answered by that document. A picture of the Index Searcher is in Fig. 1.

FUNCTIONS

The Index Searcher has six different modes of operation. Listed in the order in which they would be used, these are 1) Document Insert, 2) Regenerate Tape, 3) Question Insert, 4) Search, 5) Print, and 6) Edit. Each of these will be described after a brief reference to the system block diagram shown in Fig. 2.

The major blocks making up the Searcher, and their functions, are as follows:

- A) Magnetic Tape Unit—stores and scans document index data.
- B) Tape Buffer—serves as a buffer to and from the Magnetic Tape Unit and Flexowriter.
- C) Flexowriter—serves as a punched-paper-tape reader for document and question insertion, and as output printer for searching.
- D) Word Input Buffer—accumulates magnetic tape information frames to form complete machine words.
- E) Print-Out Buffers—store tape data which is to be printed out if the document is a desired one.
- F) Word Storage Buffer—stores complete magnetic tape index words for comparison with question words.

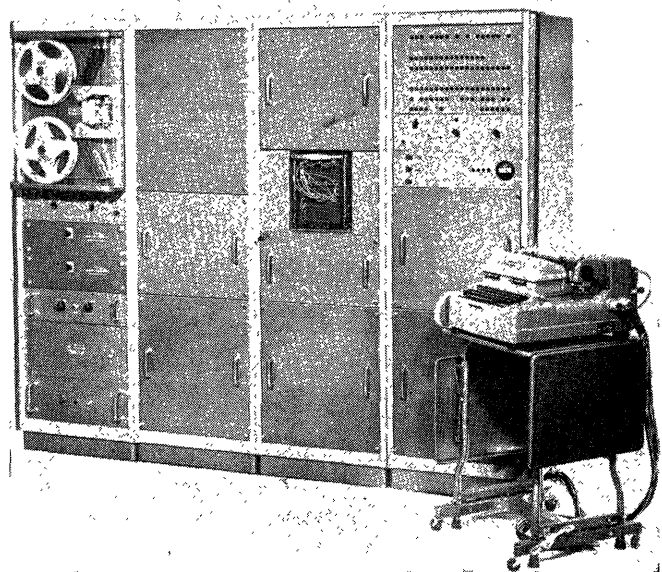


Fig. 1—Index searcher.

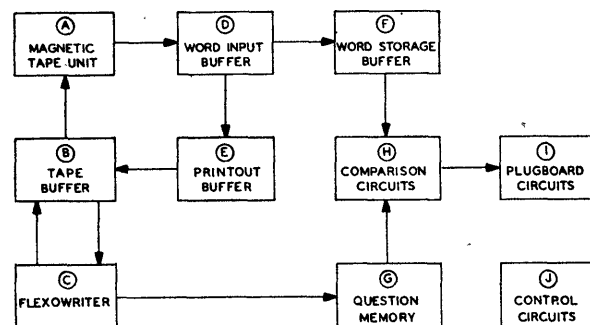


Fig. 2—Block diagram.

- G) Memory—stores question words.
- H) Comparison Circuits—compare tape index words with question words from memory.
- I) Plugboard Circuits—use results of word comparisons to make complete question comparisons.
- J) Control Circuits.

The actions of the Searcher in each of its operating modes follow.

Document Insert

This mode is used to place new document index entries into the Searcher's magnetic tape storage. Primarily it is a punched paper tape to magnetic tape conversion process. New index entries are submitted to the machine in the form of punched paper tapes. Paper tape frames are accumulated in the tape buffer until the buffer is full. Those contents of the buffer which represent complete index entries are transferred to the magnetic tape as one block of tape data. Any partial entry left in the buffer is then completed with the next paper tape information to arrive, and is followed by more documents. The process continues automatically to the end of the paper tape. Each magnetic tape block contains an integral number of complete index entries. The actual lengths of blocks on the tape are variable.

Regenerate Tape

This is a simple magnetic-tape to magnetic-tape routine, which allows the file tape to be duplicated as insurance against loss of file data through accidental damage to the tape on the Searcher. This requires an additional magnetic tape unit that is not part of the Searcher as originally built, although space has been left for it in the racks.

Question Insert

This mode transfers question words from punched paper tape to the Searcher Memory. This is accompanied by insertion of a question plugboard that specifies the nature of the comparison to be made for each word and the combination of the question words into phrases and complete questions. The plugboard connections can also specify one or two words per selected document to be printed out in addition to the document number. A simplified symbolic representation of the plugboard and its circuits is shown in Fig. 3.

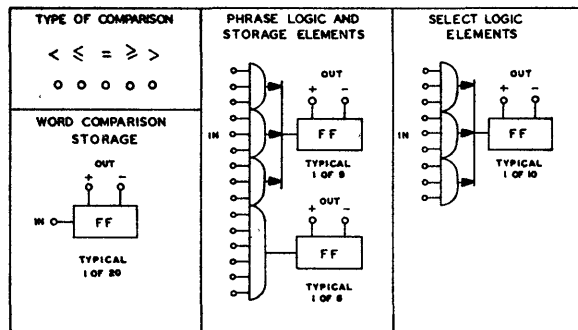


Fig. 3—Simplified representation of plugboard circuits.

Search

This is the primary operating mode of the Searcher. This mode performs the scanning of stored document index entries in search of those that meet specified question criteria. The tape can move in either the forward or reverse direction to accomplish this search. While the tape is scanned, tape frames are accumulated into complete machine words, and are compared with the twenty stored question words. A plugboard word storage element remembers a successful comparison with any of these words until an end-of-phrase designation occurs in the tape data. At that time the word storage outputs are sensed in the phrase element logic to determine whether any complete question phrase criteria have been satisfied. If so, a phrase storage element remembers this fact as scanning continues through the remainder of the tape index entry. When the end of tape data for the document is reached, sensing the outputs of the phrase storage elements determines whether or not a complete question criteria has been satisfied. During the scanning of the document entry, the Print-Out Buffer receives automatically the document number and two other plugboard-specified words. When a document answers a question, the Tape Buffer receives the contents of the Print-Out Buffer and a number identify-

ing the question answered by the document. This information is printed out on the Flexowriter. Further searching ordinarily continues during print-out. However, if a series of successive selections results in filling the Buffer faster than the maximum print-out rate, the tape automatically stops until adequate Buffer capacity is available for further document data. When the end of the recorded portion of the tape is reached, the tape unit automatically stops and positions itself ready for the next search in the opposite direction.

Print

This mode is used to print out entire document index entries rather than just the three words possible in a normal search. The machine operates in much the same manner as for normal searching until a selection is made. The tape must move in the forward direction. Selection of a tape index entry causes the tape to stop, reverse, reread the entire selected block into the Buffer, print out the complete selected entry, and then resume search.

Edit

This mode is used to delete unwanted index entries from the tape. Document entries to be deleted are specified by document number or other normal question criteria. Operation is similar to the Print mode up to the point of bringing into the Buffer the block of tape information containing the document to be deleted.

At this point the block is recorded in the same place it formerly occupied on the tape, but with blank characters in the position which had been occupied by the deleted entry.

PARAMETERS

The Searcher scans through magnetic tape document index entries at an effective rate of 218,000 bits per second, or about 5200 machine words per second, where each machine word contains 7 alpha numeric characters. For the typical document index entry length of 20 words mentioned earlier this amounts to 260 documents per second. While searching at this rate the machine seeks documents satisfying up to ten independent search criteria.

The Searcher uses 2400-foot rolls of one-inch magnetic tape for document index storage. One reel stores 57,600,000 bits, which is about 68,500 twenty-word documents. Uninterrupted search time for a complete reel is approximately 4.5 minutes.

The input-output rates of the Searcher are presently limited to the 10-character-per-second rate of the Flexowriter for document insertion, question insertion, and selection print-out.

Therefore, the document insertion rate, based on 20-word entries, is about 4 documents per minute, and the selection print-out rate, when printing out 3 words per selection, is approximately 25 documents per minute. A 20-word question insertion takes about 15 seconds. These rates can be substantially increased by use of high-speed paper tape reader units and high-speed printer or punch outputs.