Functional Organization of Data in the RCA BIZMAC System

A. D. BEARD W. K. HALSTEAD J. F. PAGE

BASIC to the data-handling pattern in the RCA (Radio Corporation of America) BIZMAC system is the highly flexible arrangement of data on magnetic tape. This arrangement was established after careful study of actual business data and processes to be handled by the system. It was found that business data have two major characteristics:

(a). Various parts of the data composing a record differed from each other in their maximum length.

(b). Each part of the data might vary from time to time from nothing up to the maximum length called for by the characteristics of the data.

The usefulness of a business machine system in a given application ultimately can be expressed in terms of cost of performing the job. The costs of entering information into the system, storing it, and retrieving the information represent important items in the over-all cost. Whatever can be done to reduce the number of manual key strokes at the input, to reduce the amount of magnetic tape used to store a given volume of information, and to reduce the time of passing through this information in processing will produce large dividends in holding down the cost of a job.

Arrangement of Data on Tape

Data in the RCA BIZMAC system are represented by binary-coded characters consisting of 7 bits each, 6 information bits plus a parity bit. Inclusion of the parity bit is one part of the RCA BIZMAC accuracy control system. It is used as a means of insuring accuracy in the transfer of data. It is chosen to be a one or a zero so as to make an even number of binary ones in every character.

Characters include the decimal digits, the letters of the alphabet, punctuation and certain other special marks, and a group of control symbols used in organizing data in the system. The code assignments permit alphabetic, numeric or alphanumeric sorting.

Data to be processed by the system are inscribed through one of several input devices and are recorded on magnetic tape mounted at tape stations. The 7 bits of each character are recorded simultaneously onto 14 parallel channels on the tape. Each bit of the character is recorded on two separate channels providing an added increase in the reliability of the tape recording.

Groups of one or more characters having some particular significance, such as a numerical quantity, an alphabetic name, a stock number, etc., are called items. An item is always preceded by a control symbol called an item separator.

All characters are recorded on tape, and read from tape serially so that the characters making up an item follow one another in sequence from the most significant to the least significant. Character spacing on magnetic tape is 125 characters per inch. The tape speed is 80 inches per second, giving a data rate of 10,000 characters per second.

One or more related items are called a message in the RCA BIZMAC system. A message is delineated by always having a "start message" symbol as the first character and an "end message" symbol as the last character. Each message usually consists of a number of discrete items of information, some of which may or may not be present each time. Each of these items can vary in character length from one to some sensible maximum. The maximum for one item may be different from the maximum of another. The use of item-separator symbols allows the use of variable item lengths, but does not preclude the use of fixed item lengths.

It is important to distinguish between two separate concepts included in the term "variable item length." First, this implies that different items of information may have different maximum lengths. This may be referred to as the nonstandard maximum item length. Each item will have some maximum possible length. Secondly, variable item length implies that each item of information may have a varying number of characters from zero (not present) up to this assigned maximum length in any particular message.

The items of a message follow one another in sequence, each being preceded by an item-separator symbol. In every message of a given type the n^{th} item always has the identical connotation, e.g., date. Therefore, a count of the item-separator symbols, starting with the first which follows immediately after the start-message symbol, permits identification of the meaning of any item.

If any particular item of information should not occur in certain messages, its item-separator symbol only is recorded on the tape in its proper sequence, except that it too may be omitted if the items that follow it are missing also. In this case, the end-message symbol may follow immediately after the last active item. This avoids writing an unnecessary number of consecutive item-separator symbols at the end of a message. Thus, with such a system, a further gain is achieved by arranging the most frequently used items at the front of the message.

Fig. 1 illustrates the layout of a typical message on magnetic tape. The message consists of a start-message symbol, followed by an item-separator symbol, and the characters of the first item, the succeeding items (each preceded by an item-separator symbol) and an end-message symbol in that order.

It is frequently desired to read messages individually from magnetic tape and to record messages individually on tape, stopping for a period of time between messages. Therefore, sufficient blank tape must be provided between successive messages to allow stopping and starting the tape between messages without missing characters. The design of the tape station is such that 5/8 inch is adequate for this purpose.

Fig. 2 illustrates the difference between systems which use either the fixed word, the nonstandard maximum, or the variable word used in the BIZMAC system. In the fixed-word system each word must be equal in length to the maximum word expected to be processed. In the example, this is 12 digits. In the second example (nonstandard maximum) a saving of digits is realized by requiring each item or word of a message to be equal only to the maximum length expected for that particular word. In the example chosen this is seven digits. The variable word used in the BIZMAC system however, requires only the meaningful digits for any item.

Influence of Data Pattern on System Design

It should be clear from the preceding sections that the variable item and mes-

A. D. BEARD, W. K. HALSTEAD, and J. F. PAGE are with the Radio Corporation of America, Camden, N. J.

sage length concepts lead to extraordinary savings in tape space, and hence in tape-file processing time. In a large inventory-control operation, ratios of maximum to average message lengths have been encountered which have exceeded 5 to 1. This ratio, in fact, applies to the main inventory file composed of approximately 250,000 messages.

Obviously, these concepts of data pattern influence the design of the data-processing equipment. The special symbols which indicate start message, end message, and item separator are used as sentinals to convey to the processing equipment the meaning of the accompanying information, and to control equipment action in accord with previously established programs or machine settings. In this sense, the special symbols replace character counts so commonly used in fixed-word-length computing systems.

In the sorter, for instance, the selection of the sorting (or merging, or extracting) criterion is based on a recognition and counting of item-separator symbols, rather than characters, beginning with the item separator immediately following the start-message symbols. (Under certain circumstances, the criterion can be a part of an item, in which case it is necessary to count characters within the designated item to arrive at the start of the criterion.)

In the computer, the high-speed memory is laid out on the basis of the maximum number of items in the incoming, and outgoing, messages, and of the maximum assigned character count for each item. Since that much memory must be available anyhow to handle the maximum-maximum case, fixed memory assignments are made, since they simplify programming and speed up processing.

The computer "read-in" instruction has associated with it a series of addresses giving the starting (leftmost) positions in the memory for consecutive items on the input tape. In this manner, the compressed information on the tape can be spread out to permit fixed-address programming for the remainder of the computer processing job. Since the series of read-in addresses does not have to be monotonic, the read-in process can be used simultaneously for editorial rearrangements, preparing for subsequent printing, for example.

For the "write-out" operation, the reverse requirement exists, at least in those instances where a file or intermediate storage tape is created. In these instances, it was found most expeditious to precede the write-out instruction by a "compress" instruction. This shifts all meaningful information close together in one area of the high-speed memory, eliminating all unnecessary space characters and, if so programmed, also suppressing all zeros which may have built up to the left of the most significant character in some results of arithmetic operations.

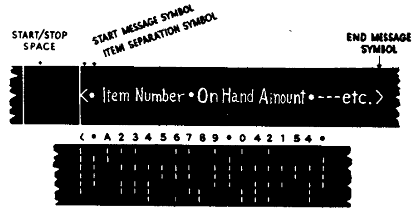

Fig. 1. Organization of data on magnetic tape

Another peculiarity of the compression of information on tape is that left-justified and right-justified numbers are not readily distinguished from each other. An example of left-justified numbers exists in the Dewey Decimal System of book cataloguing where places of equal significance are measured from the left end of the number. Alphabetic lists (telephone books) are also usually left justified. Right-justified numbers, typically, are dollars-and-cents amounts, or other integral numbers in columns which are to be added. Few, if any right-justified alphabetic items exist in practice.

In the computer, right justification becomes important in connection with arithmetic operations. All items are read left-justified from tape to high-speed memory, i.e. with the most significant digit immediately to the right of the item separator. Arithmetic operations, on the other hand, proceed from right to left, and the memory for the numerical items involved is addressed at the right-most location, where the least significant digit would appear if the item were long enough to occupy all allocated memory positions. If the actual item is shorter, space characters appear in the remaining memory positions to the right. In carrying out an arithmetic operation, the computer skips over any such unused position, and then works with the numerically significant digits, automatically "lining up" the least significant digits of the two operands. Since the computer only works with the significant digits of a word, considerable savings in computer time are realized. This concept of variable operation time in the computer is discussed in a later paper (Bensky *et al.*).

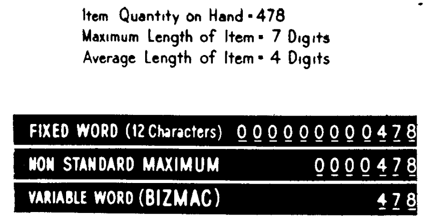

Fig. 2. Variable word data organization on tape

The use of variable word organization on tape:
- Minimizes tape storage requirements
- Minimizes the number of tape handling mechanisms required
- Minimizes the time required to pass information to and from magnetic tape

The foregoing examples show the extent to which the major equipments have been specially designed to accommodate the basic data organization. This was thought to be particularly important in connection with the computer.

In some of the peripheral devices, compromises have been struck in favor of equipment simplification. In the electromechanical printer, for instance, item-separator symbols are used to cause horizontal tabulating of information in a line, but right justification of an item can be obtained only by previously inserting on the tape the required space symbols to the left of the numeric digits. This is done in the computer. Similarly, in the card transcriber, some unnecessary space symbols, derived from blank card columns, are recorded on the magnetic tape.

Although the system is designed to take full advantage of variable item and message lengths, with the restrictions described above, this does not preclude the use of fixed lengths where convenient. The card transcriber for example, produces fixed item and message lengths, since data read from punched cards are arranged in fixed fields. After transcription, this information may be read directly into the computer memory and compressed as desired.

Summary

In summary, the variable item and message lengths concepts lead to a highly efficient tape-storage system.

Extremely significant savings in tape space and hence tape-file-processing time are achieved. The quantity of manual input operations is held to a minimum.