

### Information Retrieval

138. Computers have become very common tools for providing a means of searching large volumes of information for specific items or groups of items. Two modes of operation are possible, batch and on-line. In batch-mode retrieval, requests are processed at regular intervals with no participation by the user while retrieval is taking place. The on-line retrieval is processed at the time the request is made and provides an interactive capability by which the user can modify or change his requests, depending on responses from the computer files. This latter capacity has been found to be particularly important for subject-oriented files where many different paths may be used in the search for information. Where a batch-system request, which requires three or four revisions in order to narrow a field down to a manageable size or to a relevant group, could give answers in a time-frame measured in days, an on-line approach could do the same thing in minutes. The costs for this much superior performance are more sophisticated "software" and a number of terminals, proportionate to amount of usage and number of users. With CRTs a new dimension in presentation of information is possible. For example, a system storing machine-compatible full-text (electronic, video or video/micro storage) can display any or all selections of text from any desired item. Similarly, full texts of abstracts can be presented. In those systems where only keywords and identification pointers are stored, linked to the location of the item(s) or class of items desired, it is possible first to call up statistics of the numbers of documents in categories as a basis for modifying requests. By interacting with the data in this way it is possible to narrow down a search or widen its boundaries to provide maximum returns. This capability greatly enhances the value and use of information stores, especially in an environment where more and more emphasis is being placed upon such requirements as immediate response and immediate access. The U.S. State Department, the Swiss Foreign Office and the EEC all use the interactive on-line retrieval technique, which is becoming the standard procedure for searching large-scale and complex files.

### The Thesaurus Problem

139. A decision to index material by means of keywords carries with it the need to decide whether to use a thesaurus. A thesaurus, in this context, is a selection of words which, chosen for relevance to the subjects to be covered, are related to each other in hierarchical subject groupings, the very specific or narrowly interpreted words falling at the bottom of the hierarchy of each group and those with a broader meaning at the top. The structure of the current Records Classification Guide resembles that of a thesaurus as described here. The question of a thesaurus arises because the structure or framework it provides is useful in a keyword-based storage and retrieval system in a number of ways. For example, the indexer is assisted by having a subject reference tool which channels the description of each item into the most relevant keyword groupings. This demands that the sense of the item being described be fully understood in order to select the correct descriptions, particularly important in material which may not mention key-descriptors explicitly and which would be poorly identified by simply selecting words from the text itself. With respect to storing the descriptors the use of the thesaurus is also an advantage in that it considerably reduces the complexity of operations within the computer. Descriptors of newly-indexed material can be more easily updated into computer-resident files and hierarchical word groupings more readily assist retrieval searches.