Indexing, Retrieval and Information Specialists

34. No matter which method of full text capture is employed the information put into the full text system must be described and represented for retrieval purposes. The description process consists of identifying each item entering the system as to origin, authorship, content and other bibliographic information. After the item of information entering the system has been accurately and completely described, the item can be located by means such as subject or keyword selection.

35. The most important element in an information storage and retrieval system is the index. Creating an index involves making decisions as to the depth of indexing, the degree of open-endedness and the level of expertise required to classify the material. Index strategies may be hierarchical or they may use keywords without hierarchical significance. Often combinations of these two approaches are used. Hierarchical systems have the advantage of providing a structure for information, but suffer from difficulties in making modifications as new categories arise. The use of hierarchical classification for storage purposes (based on primary characteristics), coupled with keyword cross-references to permit greater flexibility in retrieval (based on secondary characteristics), is fairly common in the information field. Indexing strategies should be as open-ended as possible, in view of the gradual shift of interests that occur.

36. The main purpose behind indexing is to provide better retrieval capability. There are three basic methods by which retrieval may be performed:

- Whole text scanning. This primary method requires no indexing, only access to complete texts. A query is formulated and documents are manually scanned from end to end. Only one person need be involved and thus no translation problem is present. This would require an intelligent searcher in order to be able to interpret what the document says in terms of his query. A variation of this would be to pass on the query to an information specialist who then becomes the searcher, thus reducing the amount of time spent by the ultimate users in hunting for information. In order to provide adequate service, however, it is imperative that such information specialists be intelligent and knowledgeable in the subject matter.

- Keyword File Scanning. Because human scanning is limited in speed, a secondary method of retrieval is usually adopted. Each document before being stored is given a preliminary scanning by an indexer who selects words from the text that are to represent the content. These keywords are placed in a file together with the document address in the store. A query is again formulated, then a keyword file is scanned. Documents which appear to match are then selected and only those need to be scanned in detail. To search the file, the searcher must match the file keywords with search words he has formulated. But because of the richness of language, there is no certainty that the words the selector chooses will be the same as the words selected by the indexer. There will be a range of related words that may all represent the same or closely similar topics.

...6