

# INFORMATION RETRIEVAL

Kazi Sayeda Momtaz (Sharmin)

Retrieval is the operation by which information are selected from a collection of Informative materials. According to Allen Kent retrieval is defined as looking through or exploring thoroughly in order to find something. A collection of information may be of many forms :

such as (1) A library Contains, Books and Journals

(2) An Abstracts, Index, Bibliography etc.

Here a library catalogue, or index to a journal or a book can be called retrieval device. But retrieval is not limited to the preparation of a catalogue or an index. According to B.C. Vickery, 'Retrieval means more than this; its essence is the selection of document from a store. The operation is not completed until the selected documents are in the hands of the user'. Therefore, in a library, retrieval includes operations for selecting documents, arranging it, constructing and using the catalogue and making the library material available to the user.

The process of retrieval takes place in every search for a document. Pauline Atherton has defined the operation of retrieval in the following way :

(a) Word retrieval—in which we identify the words that will adequately describe the information sought;

(b) Reference retrieval—in which we identify references that are probably pertinent to be inquiry;

(c) Document retrieval—in which their actual documents are located; and

(d) Data retrieval—in which the sought information is extracted.

All these retrievals are information retrievals. The retrieval process includes the following operation :

(a) Recording of information

(b) Analysis of information

(c) Storage of information

(d) Question analysis

(e) Delivery of information to the user.

The increasing volume of technical literature and growing demands of information have led the develop-

ment of modern retrieval operation system.

All information, if it is to be available for retrieval, has to be held in some sort of store, which we will refer to rather vaguely as a 'data store'. The problem is how to organize this data store so that any required information may be obtained quickly and accurately, and also how to present this information in a useful form. We are mainly interested here in the use of computers for information retrieval, and will approach the subjects with this in mind.

One requirement for a computer-based information retrieval system is that it should be simple to use. It would be possible for the inquirer to explain his problem to an expert, who would then formulate the question in some obscure command and submit it to the system. The answer to the question should immediately or shortly be typed out by the system user onto the same terminal, or if it is very lengthy sent to a high-speed printer instead. Alternatively the user's request might be punched on cards or paper tape and the user would receive his printout of results sometime later. These two methods of operation we refer to as 'On-line' and 'batch mode' respectively. In either case we require the user should be able to formulate his own request.

In order to instruct a computer to search a file, we need some formal way of specifying our request. This request might be punched onto cards or paper tape, or typed in from an on-line terminal. If we want a list of the names of all those records containing the title = "Information" we have to type following way

List all that title = "Information"

Two points should be made at once. Firstly, the command we use for search requests is just for illustrative purposes, it is not modelled on any particular system. Secondly, few systems will provide all of the search strategies and

matching operations.

With records which have no internal order or structure, there are three basic operations for combining search clues and this is called 'Boolean search logic'. The first is the 'or.' operation or logical sum, which enables us to find records containing any of various attributes, thus

List all author = "Hossain" .or. title = "Information"

In this case a record is only retrieved either one condition or both condition will fulfil. The second operation for combining search clues is the 'and.' operation or logical product:

List all author = "Hossain" .and. title = "Information"

In this case a record is only retrieved if it contains both or all of the clues mentioned.

Here we refer list which author equals Hossain and title equals Information. We should remember that both condition must be true. Only when both condition fulfil then we can find our required information otherwise not. The logical product is a particularly important operation. When two or more clues are combined in this way, we say they are coordinated with one another.

The or. and .and. operations are diadic, that is why they stand between two search clues. The third operation is monadic, that is to say it operates only on the clue immediately following it. It is the '.NOT.' operation or logical negation:

List all .not. title = "Information"

This would cause retrieval of all those do not contain title equals Information. In fact the .not. operation is almost always used together with some other operation. In a particular it may be linked with a product to give a logical difference.

List all mercury .and. .not. toxic

Thus the request list all MERCURY .AND. .NOT. TOXIC which lists all documents concerning mercury except for those also on toxicity.

The search result may be seen on the screen or printed on hard copy. Searching by one field will show the whole record. ❖

The English pages are sponsored by Computerline