

INFORMATION RETRIEVAL - TECHNIQUES AND STRATEGIES

VAMBARA PAVANI

Guest Faculty, Department of Computer Science
RGUKT-IIIT
NUZVID
Andhra Pradesh

BHARAT VEERLA

Guest Faculty Department of Computer Science
RGUKT-IIIT,
NUZVID,
Andhra Pradesh.

Abstract

Information has become the most significant source of our day-to-day life. Information available on internet may create some confusion among its users because of its diversity. In order to get proper and exact information from internet, users need to know the effective techniques and strategies. This paper focuses on the different information retrieval processes and techniques which help users to get the required information and also to save their valuable time.

Keywords: Information retrieval, Search process, Search strategies, Retrieval techniques.

Introduction

Data is an observed fact and when any meaning is assign to data or when it is processed, it becomes information. Information reduces uncertainty and it is communicable. Information leads to confirmation but when information is processed and internalized it becomes knowledge. Today, plenty of information is available in print and non-print format which leads to information overload. Internet is one of the vital sources of information. Sometimes it becomes impossible for the users to understand the available information or they do not understand where to find it or they are unable to access the right information. These factors leads to their stress, delay in decision making, waste of time etc.

Information Retrieval

One of the best examples of information retrieval system (IRS) is library system where information is stored, processed, organized and retrieved on demand of its users.

H.P.Luhn first applied computers in storage and retrieval of information. Different types of information retrieval systems have been developed since 1950's to meet in different kinds of information needs of different users. Information retrieval system offers different search approaches those deals with three basic aspects. These aspects are as follows.

- Information storage and organization.
- Information representation.
- Information access.

Objectives of Information Storage and Retrieval/Information Retrieval System (ISAR/IRS):

- To provide information to the user in least time with least efforts.
- To act as facilitator between information and user.
- To provide non-ambiguous search results through proper indexing.
- User friendliness.

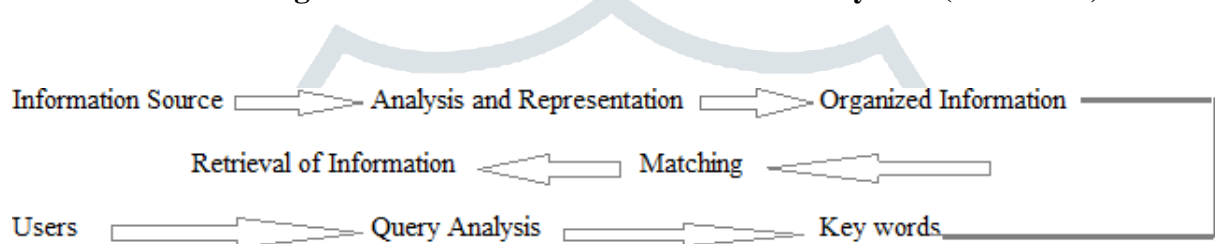
Functions of Information Storage and Retrieval/Information Retrieval System (ISAR/IRS):

- To identify sources of information relevant to the areas of the target user community,
- To analyze the contents of the sources.
- Represents contents of analyze sources that will match queries.
- Analyze user queries that will match with database.
- Retrieve the information that is relevant.
- Necessary adjustments in system based on feedback from the user.

Components of Information Storage and Retrieval/Information Retrieval System (ISAR/IRS):

According to Lancaster the major components of an information retrieval system are....

- Document selection sub-system.
- Indexing sub-system.
- Vocabulary sub-system.
- Searching sub-system.
- Matching sub-system.

Outline of Information Storage and Retrieval/Information Retrieval System (ISAR/IRS):**Kinds of information retrieval system:**

1. Offline Search: In offline search, users can get the required information with or without the help of computer and internet for example: libraries, CD-ROM etc.
2. Online Search: means the search of a remotely located database through interactive communications with the help of computer and communication channel. Online databases can be access through vendor or directly. For example: OPAC, Databases, Internet etc.

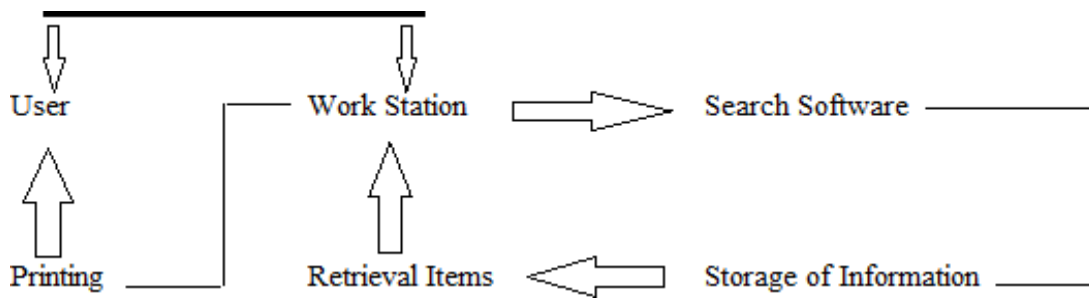
Requirements for online search system:

- Workstation or computer.
- Internet connection (dial up or broadband).
- Internet service provider such as VSNL.
- Search software.
- Storage of information (in-house collection or databases).

Features of online search system:

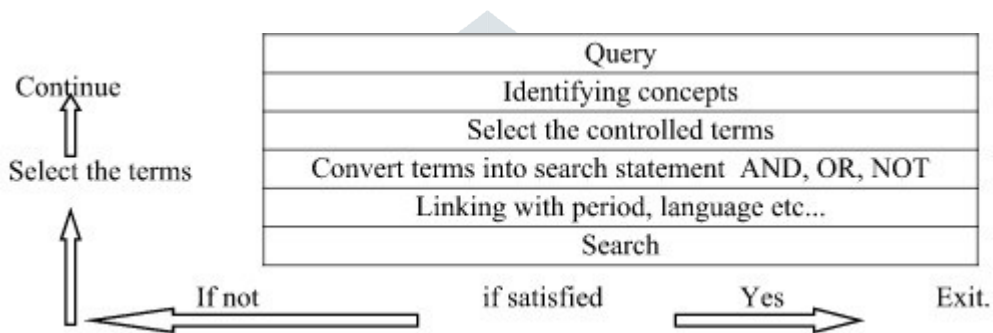
- Coverage and sources of information.
- Indexing mechanism.
- Vocabulary control.
- Searching facilities.
- Ranking techniques.
- Search modifications.
- Online search systems.

Input Search Request



Search Process:

Search process help us to get the exact information by following step by step process. This process is as follows.



Retrieval Techniques:

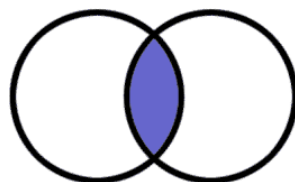
Retrieval techniques are designed to help users to locate the information they need effectively and efficiently. These techniques help users to find out the required information easily. There are two types of retrieval techniques.

I. Basic Retrieval Techniques:

1. Boolean Searching

George Boole (1815-1864) developed AND, OR, NOT Boolean operators. By using these techniques user can narrow down their search to get the required information.

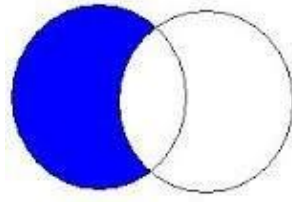
- a) **AND:** It includes addition of two different concepts for narrowing down the search. It retrieves all those items where all the constituent terms occur.



Internet

web

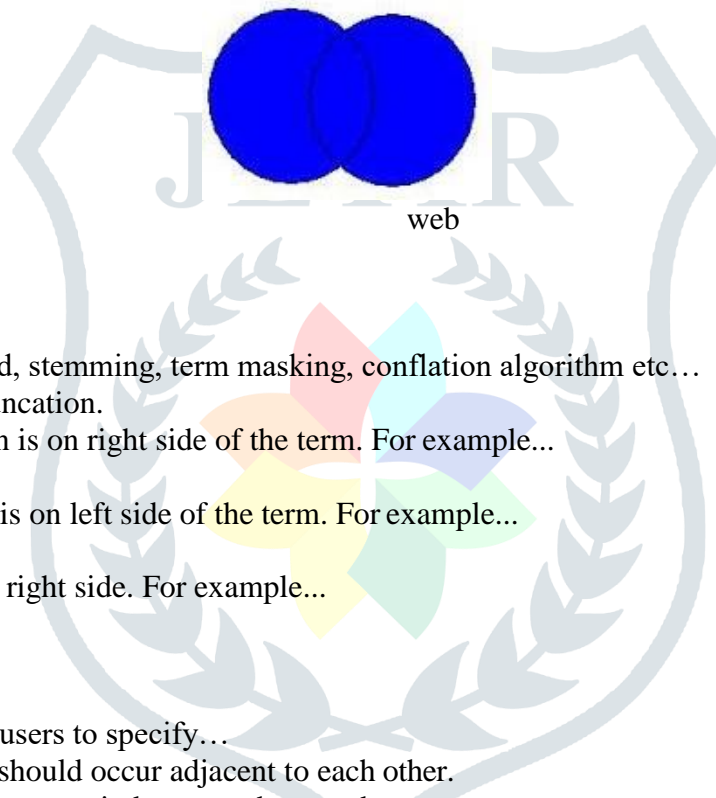
- b) **NOT:** It is separation of complex concepts into individual simpler ones. It allows users to specify those terms that they do not want to occur in the retrieval records. It excludes unwanted results. Search output will decrease with increase in NOT term.



Internet

web

- c) **OR**: The inclusion of more concepts to expand their connotation. It is used for broadening a search. It allows users to combine two or more search terms that system will retrieve all those terms that contain either one or all of the constituent terms.



Internet

web

2. **Truncation Search :**

It is also known as wildcard, stemming, term masking, conflation algorithm etc...

There are three types of truncation.

- Right truncation: truncation is on right side of the term. For example...
Lib*
- Left truncation: truncation is on left side of the term. For example...
*rary
- Simultaneous with left and right side. For example...
polymer

3. **Proximity search :**

This search facility allows users to specify...

- Whether two search terms should occur adjacent to each other.
- Whether one or more words occur in between the search terms.
- Whether the search terms should occur in some paragraph irrespective of the intervening words. Proximity search is as good as AND. It searches for the occurrence of two or more search terms but it specifies the distance between the search terms. For example...
Sun within four words after moon (sun # 2 # after moon)

4. **Field Searching :**

Document is presented by attributes such as Author, Title, Publication date, Document type, File type etc. these attributes are called as field searching. Search is possible with the help of these fields also.

5. **Case Sensitive Searching :**

For languages such as English, French, Spanish upper and lower cases makes a difference. Case sensitive searching allows pinpointing exactly how a term is represented in a query and the system. For example...

Web - World Wide Web.

Web - web woven by spiders.

6. **Range Searching :**

It is very useful in numerical searching. It is important in selecting records within certain data ranges. For example...

Greater than (>) Less than (<) Equal to (=)
 Not equal to (\neq or \nlessgtr) Greater than or equal to (\geq) Less than or equal to (\leq)

II. Advanced Retrieval Techniques :

i. Fuzzy Search :

It sounds related to truncation search but with a major difference. Truncation is intended to retrieve different forms of terms when they share some parts in common. Fuzzy searching is designed to find terms that are spelled incorrectly at data entry or query point.

ii. Weighted Search :

For weighted searching, weights are assigned to terms when a search query is composed to indicate proportionally their significance or the emphasis the user placed upon. For example...internet 3 AND web 5

Search Strategy:

A search strategy is referred to as that set of decisions and actions taken throughout the conduct of search. In other words, steps involved in searching or retrieving information, understanding of the query. For this a strategy is needed which is called the strategy of searching the query.

The conceptual analysis of the request translated into the language of the system is search strategy. Knowledge about the nature of organization and exact needs of users are the essentials of good strategy. Following are the kinds of searches.

Kinds of Searches:

1. High recall search: when the user needs to find out all the relevant items on the stated topic.
2. High precision search: users need only relevant items.
3. Brief search: users want only relevant items as opposed to all relevant items.

Types of Search Strategy:

Charles Boornie classified into following topics.

1. Building block: starts with single concept searches.
2. The snowballing: also known as citation pearl growing approach. User first conducts the search and then modifies the search query based on the results retrieved till satisfactory results.
3. Successive fraction approach: starting a search with a broad concept and narrow concept. Use of Boolean operator.
4. Most specific facet first approach: it is applied to multiple concept searches.
5. Towards a quick / convenient approach.

Preparation of Search Strategy:

1. Necessary search aids :

- System manuals and search aids.
- System and database newsletter.
- Vocabulary control devices and classification schemes, dictionary, glossaries etc.

2. Information about the query :

- Consultation of different reference sources.
- Determination of exact requirement of user.
- Pre-search interview.

3. Whether online search is necessary :

- Availability of databases on given subjects.
- Its coverage, cost and other related factors.

4 Choice of databases governed by factors :

- Sub coverage.
- Document coverage.
- Accessibility.

5 Decision regarding systems to use according to :

- Database coverage.
- Search fields.
- Search devices.
- Performance.

6 Analysis of query and selection of search terms through :

- Knowledge of the systems and files.
- Use of reference tools.

7 Planning and carrying our search through :

- Preparation of initials files.
- Narrowing down search by reducing or terms and adding and if necessary.
- Broadening the search.

8 Obtaining results and communicating with the user :

- Using most suitable format.
- Getting users response regarding the search results.

9 Recordings :

- Steps followed in the search.
- The user's responses regarding the search results.

Conclusion

Today each one of us needs information for some or the other purpose. The value of information increases in the process of its use. Library professionals should know these above search processes, techniques and strategies to provide its users right and exact information in a less time. Only by doing so, we can give proper justice to the S.R.Ranganathan's laws of library science.

References

1. Lancaster, F.W. *Information retrieval systems: characterization testing and evaluation*. New York. John Wiley & Sons. 1979. Print.
2. Large, Andrew and others. *Information seeking in the online age: principles and practices*. Munchen, K. G. Saur. 2001. Print.
3. Foskett, A. C. *Subject approach to information*. London. Library Association. 1982. Print.
4. Meadows, Jack. *Information & communication*. London. Bowker. 1994. Print.