Notes

17

SEARCH TECHNIQUES: BASIC AND ADVANCED

17.1 INTRODUCTION

Searching is the activity of looking thoroughly in order to find something. In library and information science, searching refers to looking through records thoroughly in order to find desired information. You have already studied retrieval tools like catalogues, indexes, etc., for retrieving information. In this lesson, you will learn need and ways of searching organized information for retrieval purposes. You will also be exposed to fundamental aspects of search techniques for information retrieval from electronic sources.

17.2 OBJECTIVES

After studying this lesson, you will be able to:

- define search techniques;
- explain organization of words in a dictionary;
- use dictionary, numeric and numeric/classified techniques for arranging and retrieving library material;
- define search engines;
- identify search process and design a search query;
- know the role of search operators;
- define Boolean logic;
- understand types of search;

Notes

- define, explain and differentiate the field based and full text search with examples.

17.3 SEARCH TECHNIQUES

The search technique is a mechanism through which one can find relevant information from information systems. The information system may be in-house or online. In-house information system is one where information is stored within the scope of an organization for retrieval purposes. The online information system is a system where electronic information sources are stored remotely and are accessible through a communication mechanism. Most of the online information systems are compatible with World Wide Web (WWW) and are accessible through internet. The in-house information system may have information sources in both printed and electronic form. Thus, storing mechanism and search techniques are two different aspects. We will discuss these two aspects of storage and retrieval of information.

INTEXT QUESTIONS 17.1

1. Define search technique.
2. What is online information system?

17.4 STORAGE MECHANISM

In-house information systems and online information systems are designed to store specialized information or information on a particular theme or subject. Such systems provide their own search mechanisms and a set of guidelines to find specific information. In library and information centers, information is available in both print and electronic form. Given below are some of the storage mechanisms and their role in information retrieval:

(i) Dictionary Arrangement

(ii) Numeric Arrangement

(iii) Classified Arrangement

17.4.1 Dictionary Arrangement

Dictionary arrangement means an arrangement where words are organized in alphabetical order of the language. The alphabetical order is the sequence based on the position of a particular alphabet in the script of the language. For

example, the English language uses roman alphabets and the order is A, B, C, D,…….Z. Here, the alphabet "A" is at first position, "B" at second and likewise "Z" at twenty sixth position in the sequence. Therefore, the words in dictionary arrangement are organized as per the sequential position of alphabets. For example,

Action
Ante
Apple
Art
Catalogue
Classification
Search
…….

Here, the first four words start with "A" but their positions are fixed as per the position of second, third or fourth letter. This is followed by another set of two words starting with "C". Hence, the words starting with "C" have been given position after the words starting with "A". Following this process, the words are organized in this arrangement. This mechanism of arrangement is followed for arranging entries in catalogues, which have words as access point. For example, author, title, subject, etc.

17.4.2 Numeric Arrangement

The numeric arrangement is the arrangement where numbers are organized in ascending or descending order. For example,

123.45
234.15
234.51
435.21
541.23
……

Here, you find that all the numbers have same set of five digits, i.e., 1 to 5, but as per their numeric value, these are organized in ascending order and sequence has been made. In libraries, that follow Dewey Decimal Classification system, you will find that the books are arranged in numeric order on the shelves.

17.4.3 Classified Arrangement

Most of the libraries organize their books on the shelves as per the call number

Notes

of books. The call numbers are the combination of class number, book number and collection number. These three numbers may be numeric or alphanumeric as per the scheme, followed by the library. Hence, retrieving books from the shelves becomes easy when we understand the numeric, alphanumeric or classified arrangement. For example, a few call numbers based on DDC scheme have been arranged below as they are arranged on shelves.

321.4 RAM
370.1150954 DEM
370.1523 DES
371 ILL
371.3078 KEM
371.32 NIS
371.397 GRE
371.926 BRA

Another example of information retrieval following these arrangements is taken from a book index. You might have noticed that almost all books have an index at the end. The book index is a list of words/terms alongwith page numbers on which those appear in the text. Depending upon the size and nature of the book, the terms in the book index are organized either in dictionary or classified order. After understanding these arrangements, you can find information on a topic from the book easily.

INTEXT QUESTION 17.2

1. Define dictionary, numeric and classified arrangement.

17.5 SEARCH ENGINE

Searching information from the electronic or digital media is different from the print media. When information is stored in electronic or digital form, user interface is provided to find relevant information from the system. This user interface is a software, which has provisions to accept keywords or terms, representing required information to conduct the search. It brings the result of the search in the format defined in the software. The software meant for searching information from the information system is known as search engine. Hence, we can define a search engine as 'a software, meant for searching information from electronic or digital information domain, on the basis of input given by a searcher that displays the result in user friendly format'.

The input to the search engine is known as search string or query. The query may be a single term or a set of terms representing the information one is looking for. The search engine searches information based on the query and provides a list of sources which match the query. The list is displayed in a format, designed by the search engines. Depending upon the nature of the search engine, the list may contain brief description of information sources, on the basis of which, the searcher may decide to acquire or refer to full record or not.

You might have searched Online Public Access Catalogue (OPAC) of your library or *Library of Congress Online Catalog (LCOC)* or *PubMed* as well as *Google* or *Yahoo* on internet. All are the search engines.

INTEXT QUESTION 17.3

1. Explain a search engine.

17.6 SEARCH PROCESS

The search process is a set of functions which are performed for searching the relevant information effectively. The process follows some basic steps to conduct search and get desired results. These steps are as follows:

(i) Recognise and State the Need

(ii) Development of Search Strategy

(iii) Execution of the Search Strategy

(iv) Review Search Results

(v) Edit Search Results

(vi) Evaluation and Feedback

17.6.1 Recognise and State the Need

It is important for an information professional or searcher to understand the need and the purpose of a search. Information on a topic may be needed for general knowledge, research and development, or for any other purpose. After understanding the need and purpose of the search, a query statement should be developed.

There should be an agreement between the information seeker and the searcher on the search requirements. This agreement leads to formulation of effective search strategy for relevant and effective result.

Notes

17.6.2 Development of Search Strategy

The development of the search strategy includes conceptual formulation of query, translation of conceptual formulation into the language of keywords, descriptors or facets, identification of synonym and associated terms, etc. The concept of facet analysis(PMEST), given by Ranganathan as well as the concept of specific subject can be used as an effective tool for designing a query. After this, it is important to select the information domain to be searched like, the OPAC of a library, database or likewise, depending upon requirements.

The search string or query, is the combination of terms, keywords or descriptors, which represent the information. As search strings contain vocabulary, the linguisic features and their implications on the search and retrieval of information have to be analyzed. Here, three aspects, namely, Syntactic Value, Semantic Value and Boolean Operators are to be understood.

a) Syntactic Value

The syntactic value of a search string deals with the kind of formula or connecting symbols through which keywords or terms are connected to represent the concept to be searched by the search engines. We will try to understand the syntactic value of a query by this example. There are two terms, say "poetry" and "Indians" connected by two different connectors,"among" and "by'. Each gives a different meaning, as follows:

(a) 'poetry among Indians'means 'What is the status of poetry among Indians?" Or 'What is the approch of Indians towards poetry?"

(b) 'Poetry by Indians'means, poetry composed by Indians.

b) Semantic Value

The semantic value of a search string deals with the meaning of the string in the context of the required informtion and the interpretation by the search engine. For establishing the meaning of the concept to be searched and understood by the search engines, we use operators as connectors of keywords as permitted by the search engines.We can understand the semantic value of a query through two examples given below:

(i) The query 'contribution of Indian society in mathematics' means the contribution of Indian society in the field of Mathematics.

(ii) The query 'contribution of mathematics in indian society'means contribution of Mathematics in shaping Indian society.

These two examples give us clear perception of the semantic value of a query. The same set of keywords and connectors give different meaning when written in different order.

c) Boolean Operators

Boolean Operators are simple words (AND, OR and NOT) used as conjunctions to combine or exclude keywords in a search. These are used to connect and define the relationship between the search terms. Thus, resulting in more focused and productive results.These three terms are widely accepted by the designers of the search engines. They have well defined meaning while used as operators in information search. The three operators of Boolean logic are the logical sum (+) OR, logical product (x) AND, and logical difference (-) NOT. All the information retrieval systems allow the users to express their queries by using these operators. Let us now understand the implications of these three operators.

OR Operator: The OR operator allows the searcher to specify alternatives among the search terms. When a string is created using OR operator, the search engines retrieve all those resources where any of the terms or keywords connected with 'OR' exist. For example, if we create a search string like, 'student OR education' and search it, then the output of the search will be a list of references of all those resources, available in the system, where either student or education exists.

AND Operator: The AND operator is used to combine two or more terms. When a string is created using AND operator, the search engine retrieves all those resources where all the terms or keyword connected with 'AND'exist. For example, if we design a search string like, 'student AND education'and search, then the output of the search will be a list of references of all those resources, where student and education, both the terms exist.

NOT Operator: The NOT operator is used to exclude the term from a set of resources. For example, if we create a search string like 'student NOT education' and search, then the result of the search will be a list of references of all those resources, available in the system, where term student exists but not education.

You can understand the implications of boolean operators with the graphic representations. Here 'A' represnts a set of students and 'B' represents a set of education in Fig. 17.1

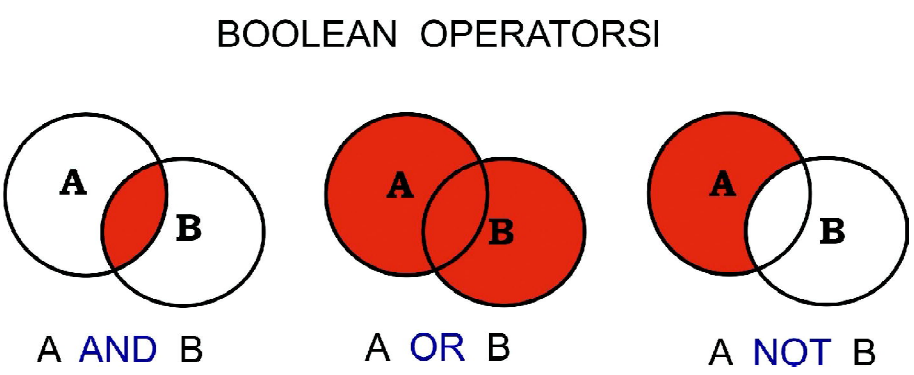

Fig. 17.1: Graphical presentation of the implications of Boolean Operators

INTEXT QUESTION 17.4

1. Define search strategy.

17.6.3 Execution of the Search Strategy

The searcher should have knowledge of data structure adopted by the information system that stores data before executing a search. The system based search engines are designed to search information in a database according to its architecture. Like in OPAC, if we put a query as 'Tagore, Rabindra Nath' and search in author field, then only those records will be retrieved and displayed from the database which have been authored by him. But, if we direct the same query into the title field, then those records will be displayed, in which 'Tagore, Rabindra Nath'appears in the title or a part of the title. This means that the references of materials, written on 'Tagore, Rabindra Nath'will be listed in result.

Depending upon the need and purpose of the search and expertise of the searcher, the search may be conducted using the features of the search engines. Hence a searcher should know the types of search and implications to get effective output. The types of searches are:

a) Field Based Search

b) Full Text Search

c) Truncation Search

d) Proximity Search

e) Limiting Search

f) Range Search

g) Simple Search

h) Advanced Search

(a) Field Based Search

The search conducted on a particular field of the database to get required information is termed as field based search. As you are aware, the complete information of catalogue is stored in different fields in a bibliographic database. If you wish to search an author, direct the search engine to author field or if you wish to search through title or subject, direct the search engine to title or subject field.

If 'Sen, Amartya' is searched in Author field, then the result will show the works authored by him. While, the same search when executed on Title field, the result will show the works on him. For example, a search was conducted on Library of Congress Online Catalogue (LCOC) putting 'Sen, Amartya' as a search string and selected 'author/creator' field to be searched. The result given by the search engine was a list of 7 documents when author/creator was searched, while it gave a list of 157 documents when title field was searched. The images of the search and the results are given below in Figs. 17.2 to 17.5.

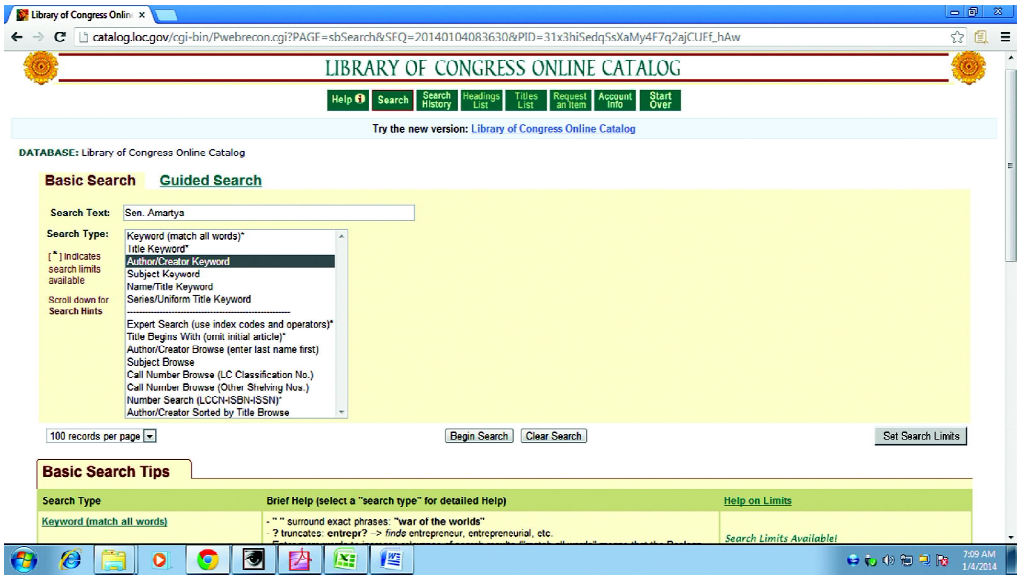

Fig. 17.2: Search on LCOC

Fig. 17.3: Search on LCOC where Sen, Amartya was searched in author/creator field

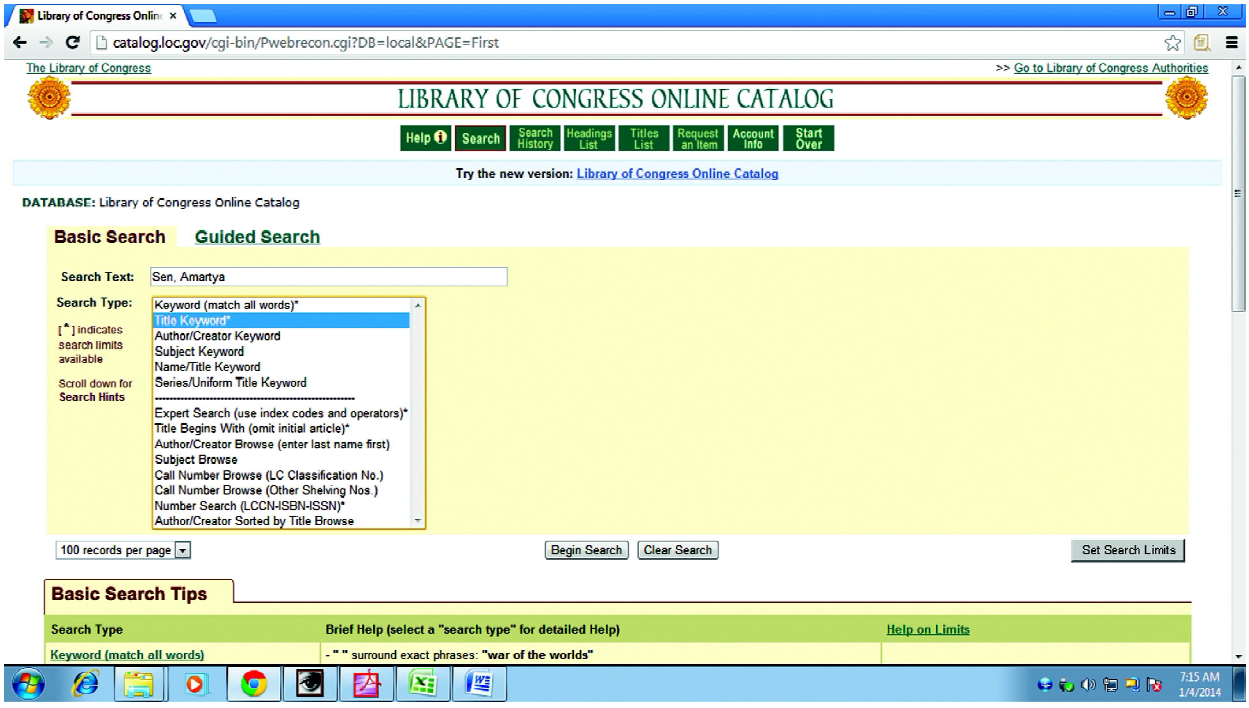

Fig. 17.4: Search on LCOC

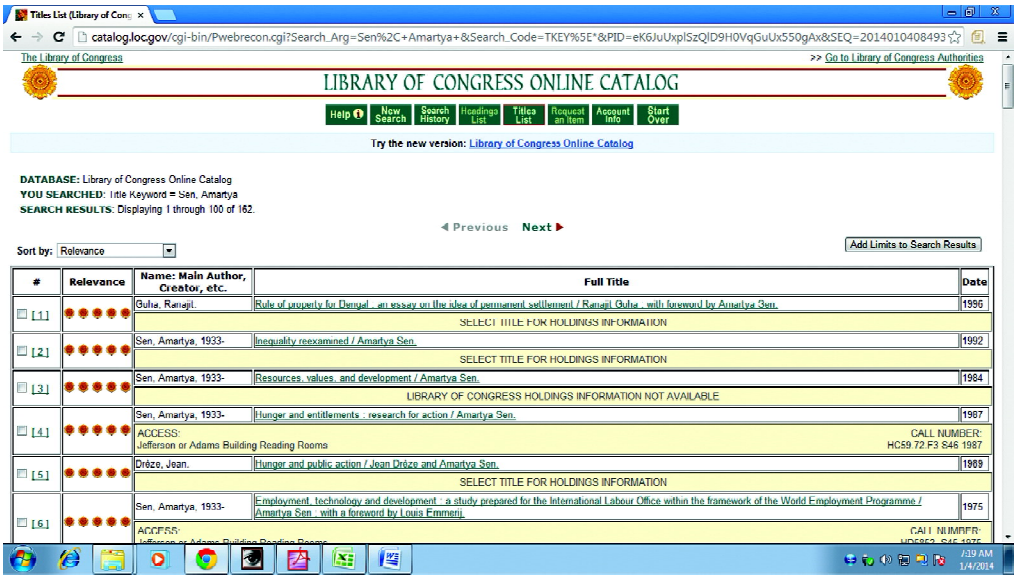

Fig. 17.5: Search on LCOC where Sen, Amartya was searched in title field

(b) Full Text Search

Full text search is a searching mechanism, which conducts the search on each and every field of the database and extracts all those records which match the query. For example, the same search (Amartya Sen) when conducted on LCOC with keyword option, which works as full text search, gave a list of 193 records. This shows that, in full text search the number of hits increased as it extracted all those records which had 'Sen, Amartya'in any fields. The search result is given in Figure-17.6.

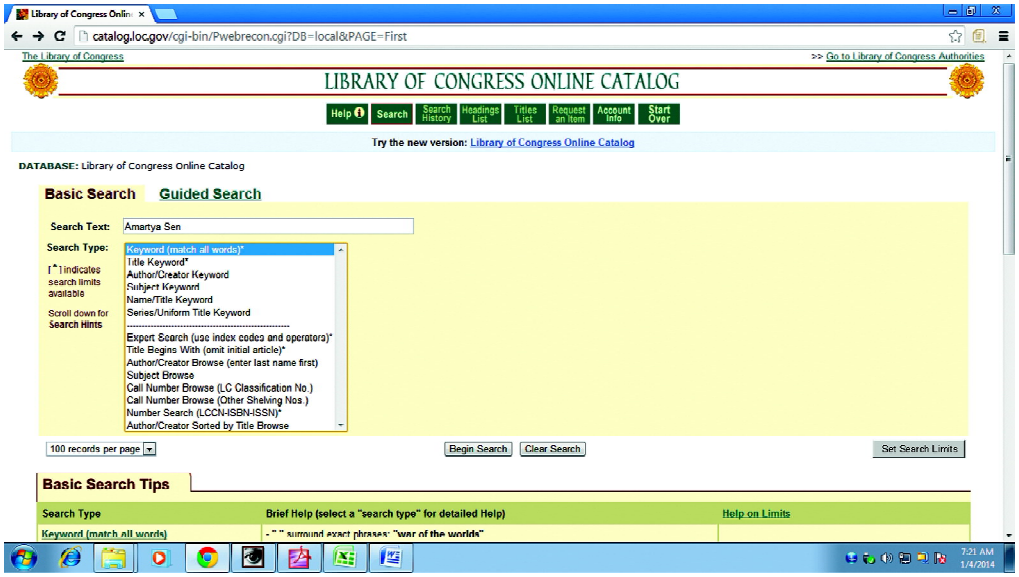

Fig. 17.6: Search on LCOC

Notes

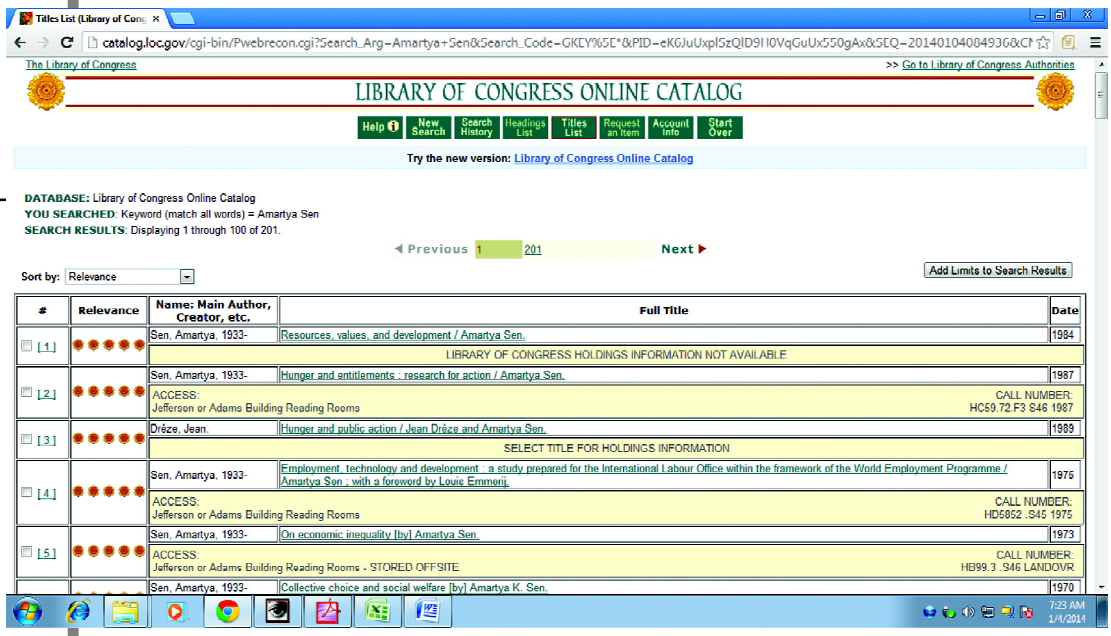

Fig. 17.7: Search on LCOC where Sen, Amartya was searched in keyword match.

(c) Truncation Search

Truncation search, is a search technique, in which, the search is conducted for different forms of a word having the same common root. It is one of the most widely adopted methods in information retrieval system. In this technique, root word is taken with truncation mark and search is conducted. For example, if we search 'India*' then all the records will be retrieved where term 'India'appears full or part of any word. The output will list all the records of the domain having, India, Indian, Indiana, Indianization or likewise.

(d) Proximity Search

The proximity search, is a search technique, which allows the searcher to define the distance of two terms from each other. Whether, the two search terms, should occur adjacent to each other, or, one or more words occur in between the search terms; or the search terms should occur in the same paragraph, irrespective of the intervening words, etc. Different search engines use different set of operators for this purpose.

(e) Limiting Search

In limiting search technique, a searcher limits the string as per the architecture of database and searches different terms of the same string in different fields. For example, if a searcher is searching 'Development as freedom by Amartya Sen' then the string will be broken into two sub-strings, viz.'Development as

freedom'and 'Amartya Sen'. The sub-string 'Development as freedom' will be put in title field and sub string 'Amartya Sen' will be put in author filed and then search will be conducted.

(f) Range Search

Range search technique is a technique,which allows searchers to select records within certain data ranges. This technique is more suitable for numeric data search. The operators and their meaning differ from search engine to search engine. A few commonly used operators are:

- Greater than (>)
- Less than (<)
- Equal to (=)
- Not equal to (|= or <>)
- Greater than or equal to (>=)
- Less than or equal to (<=)

For example, if we put publication year 2000 >=, then the result will list all those resources which have been published 2000 AD onwards.

(g) Simple Search

Simple search is such a technique where a searcher puts keywords in a simple format without understanding the behavior of the search engine or the architecture of the database or the impact of the operators and connectors. Almost all the search engines provide the facility of using simple search technique. The simple search works on the model of Full text search discussed above.

(h) Advanced Search

Advanced search technique is a technique through which a searcher searches the information using different tools and mechanisms to get precise and relevant results. In this technique, a searcher creates the search string using operators and parameters provided by the search engine. Searching information, combining different methods discussed above, also falls in this category. Here, the scope of each and every term of the string may be defined as per facility available in the search engine. We will discuss different aspects of advance search in Lesson 18.

INTEXT QUESTIONS 17.5

1. What is field based search?
2. Explain full text search.
3. Define truncation search.
4. Discuss limiting search.

17.6.4 Review Search Results

The best reviewer of the search results is the user. But the information professionals should also review the search results on the basis of criteria given for evaluating information retrieval systems.

17.6.5 Edit Search Results

The editing of search results involves transformation of the search results into a user friendly format. This may involve arranging the results into a well-organised package, hightlighting the important entities, adding more information to the entities and reformating of information to suit the user's requirements.

17.6.6 Evaluation and Feedback

The evalution of search results involves participation of both, the users and the searchers. The quality and quantity of the results are assessed and if needed, the process may be redefined and restarted if the final result does not satisfy the users' needs.

On the basis of the search process discussed above, a simple model can be given as in Fig. 17.8.

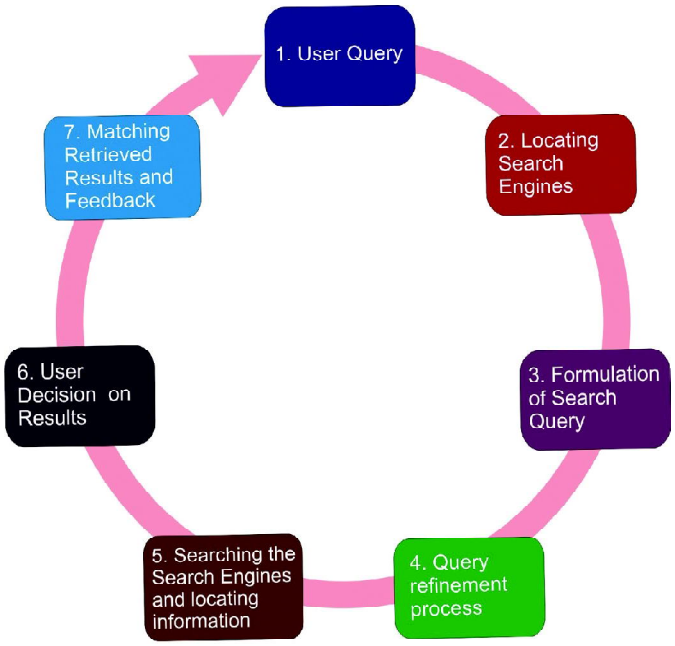

Fig. 17.8 : Search Process Model

INTEXT QUESTIONS 17.6

1. How a search result is reviewed?
2. What is the need of editing search result?

WHAT YOU HAVE LEARNT

- The standard mechanism, called information search techniques is used for retrieving information from any information system.
- The search technique is a mechanism through which, one can find relevant information from information systems. The information system may be in-house or online.
- Storage mechanism can be dictionary, numeric and classified arrangement of information.
- Search process follows a set of functions as:
  - determination of user's need of information search;
  - designing search strategy;

- selecting the information system to be searched and accordingly the search engine;
- creating search query or string using keywords and operators which, represents the semantic value of the user's requirements and the syntactic format which the engine interprets;
- conducting the search;
- evaluation of the result. If needed, again filter or redefine or restart the search process; and
- presentation of the search results in a user friendly format.

- For getting relevant and effective search results, a searcher should have knowledge of the types of searches and skills of conducting them.

TERMINAL QUESTIONS

1. Explain search techniques and their need for information retrieval.
2. List the search process and give a brief note on each of the steps.
3. Explain semantic value of a search string.
4. Explain the Boolean operators and their impact while connecting two keywords 'A' and 'B'.
5. How does an advance search differ from a simple search?

ANSWERS TO INTEXT QUESTIONS

17.1

1. The mechanism by which we find relevant information from any information system is known as search technique.
2. The online information system is a system where electronic information sources are stored remotely and are accessible through a communication mechanism.

17.2

1. Dictionary arrangement is an arrangement in which words are organized in alphabetical order of the language. In numeric arrangement, numbers are organized in ascending or descending order. Classified arrangement is an arrangement in which the words or numbers or a mix of both (alphanumeric) are firstly grouped on the basis of some characteristics. Then these are organized in dictionary or numeric or a combination of both orders.

17.3

1. The search engine is a software, meant for searching information from electronic or digital information domain. On the basis of query given by a searcher, the search engine displays the results in user friendly format.

17.4

1. The search strategy is a process of developing conceptual formulation of query, translation of conceptual formulation into the language of keywords, descriptors or facets, identification of synonymous and associated terms, etc.

17.5

1. The field based search is one where search is conducted on a particular filed of the database to get required information.
2. The full text search is a searching mechanism, which conducts the search on each and every field of the database and extracts all those records which match the query.
3. The truncation search is a technique, in which the search is conducted for different forms of a word having the same common root.
4. In the limiting search technique a searcher breaks the string as per the architecture of a database and searches different terms of same string in different fields.

17.6

1. The best way of reviewing the search result is to get feedback of the user. If the user is satisfied with the result then the search may be considered successful.
2. The result displayed by the search engines are not always in the user friendly format. Thus, editing the result in the format which user can understand becomes is necessary.

GLOSSARY

Alphabetical arrangement: Arrangement based on the sequence of the alphabet of a particular language in which terms are written.

Ascending: Increasing order

Descending order: Decreasing order

Information domain: An information system where information or its sources are stored logically for retrieval purpose.

LCOC: Library of Congress(USA) online catalogue

Numeric arrangement: Arrangement of numbers in ascending or descending order

OPAC: Open Public Access Catalogue

PMEST: Personality, Matter, Energy, Space and Time (of Colon Classification)

PubMed: Online database of US National Library of medicine covers citations for biomedical literature from MEDLINE, life science journals and online books. The citation may include links to full-text content from PubMed Central and publishers web sites.

Software: Any computer executable programme like Libsys WINISIS etc.

User interface: A tool which works between user and the system

WWW: World Wide Web is system of interlinked hypertext documents accessed via the Internet. With a Web browser, one can view web pages that may contain text, images videos and other multimedia and navigate between them via hyperlinks.

WEBSITES

http://catalog.loc.gov/help/titles.htm

http://www.ncbi.nlm.nih.gov

http://www.internettutorials.net/basic-search-techniques.asp

http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html