UNIT - 5

Information Retrieval System

Location	Learning Outcome	Knowledge Evaluation	Performance Evaluation	Teaching and Training Method
• Class room, library or computer laboratory.	• Information Retrieval System: Concept.	• Meaning of IR system. • Functional aspects & types of IR system. • Key characteristics and features of IR system.	• Explain concept, characteristic, features and functions.	• **Interactive lecture:** Discuss concept of information retrieval system. • **Activity:** Visit computer laboratory and show students characteristics of IR on internet.
	• Search Strategy: The Action Plan.	• Search strategy: basic steps. • The searching process. • Methods to perform a search on the internet and finding information precisely.	• Explain Information searching process conducted by user and explore various methods to improve search.	• **Interactive lecture:** Discuss search strategy, searching process and methods. • **Activity:** Visit computer laboratory and perform the search on the basis of strategy and methods. .
	• Search query formulation process for information retrieval on internet.	• Web information search process by search engines.	• Able to formulate search query and information retrieval on Internet.	• **Interactive lecture:** Discuss search query formulation and conducting search on internet. • **Activity:** Visit computer laboratory & show students formulation search query and conducting search.
	• Search technique and methods used.	• Know the varieties of ways and methods to search information on the web.	• Able to identify different technique and methods used for refining search.	• **Interactive lecture:** Discuss different methods or techniques for refining search.

				• **Activity:** Visit computer laboratory & show students methods and techniques of refining search and their results.
	• Practical.	• Enhance skills of searching the required information precisely using search process and techniques.	• Identify the different technique and methods used for searching and retrieving information from web. • Able to follow basic step to conduct searches to retrieve information.	• **Interactive lecture:** Discuss process of using different web tools. • **Activity:** Visit library and ask students to perform searches using techniques discussed to retrieve information & also refine search and find the difference in results.

5.1 Information Retrieval System: Concept

The concept of information retrieval system is self-explanatory from the terminological point of view and represents itself as a 'system which retrieves information'. The term "information retrieval" coined in 1952 and gained popularity in the research community from 1961 onwards. IR systems are concerned with two basic aspects:

a)   How to store information
b)   How to interpret its structure

One may simply denote such a system as one that stores and retrieves information. As a system, an IR system is therefore composed of a set of interacting components, each of which is designed to serve a specific function for a specific purpose and all these components are interrelated to achieve a goal. The concept of IR thus pre-supposes that there are some items of information, which have been keeping organized in any suitable order for easy retrieval.

According to Lancaster, "an information retrieval system does not inform (i.e. change the knowledge of) the user on the subject of his inquiry. It merely informs him of the existence (or non-existence) and where about of documents relating to his request".

According to Spark Jones, an IR system is a set of rules and procedures, for performing some or all of the following operations:

a)   Indexing (or constructing of representations of documents);
b)   Search formulation (or constructing of representations of information needs);
c)   Searching (or matching representations of documents against representations of needs);
d)   Index language construction(or generation of rules of representation).

So, information retrieval collectively is defined as a "science of search" or a process, method, and procedure used to select or recall recorded and/or indexed information from files of data.
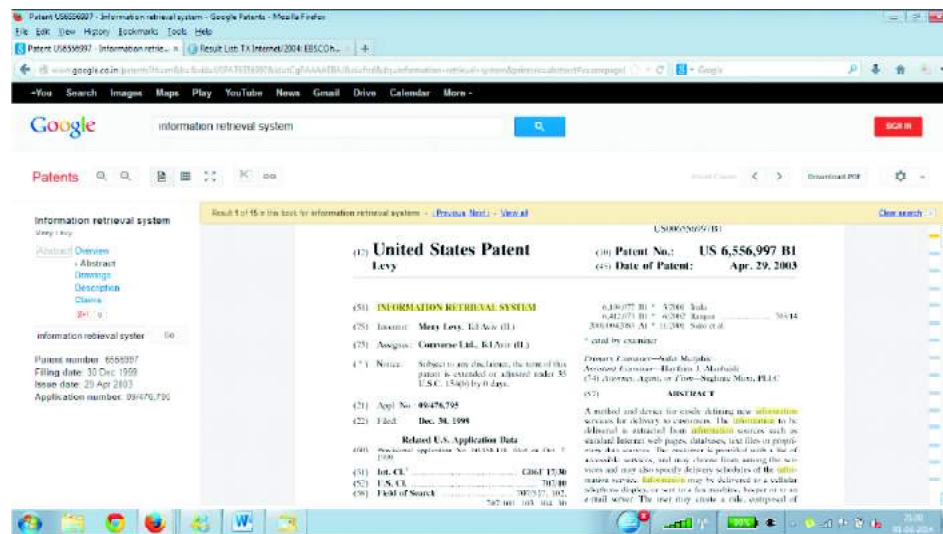
INFORMATION STORAGE AND RETRIEVAL

Figure 5.1: Information Retrieval System

5.1.1 Components of an IRS

Lancaster [1979] mentions that an IRS comprises of the following components:

a)   **The Document Subsystem**

b)   **The Indexing Subsystem**

c)   **Vocabulary Subsystem**

d)   **Searching Subsystem**

e)   **User-System Interface**

f)   **The Matching Subsystem**

(a)   **The Document Subsystem:** it involves the location, selection, ordering and receipt of source materials for collection. This process emphasis on two aspects: Currency of information and Completeness of information, both consist of the following tasks :

•   Determination of current and probable future requirements of potential users of IRS.

•   Formulation of a policy acceptance of source material as defined by subject coverage, publication type, or other criteria.

•   Comparison of available or incoming source materials with policy to determine which shall be included in the IRS.

(b)   **Indexing Subsystem and Vocabulary Subsystem:** This system is for naming subjects in the way we have described called an indexing language and/or any other language. It consists of two parts (a) Vocabulary and (b) Syntax.

If we use term as they appear in documents without modification, we are using natural language. In this process, we face problems arising out of:

•   The use of words/terms; or else.

•   Use of the word order (syntax).

For example, "child psychology" may express as "psychology for children". Therefore, a

88

controlled vocabulary is used. Vocabulary control involves the establishment of relationships among analytic, often an arbitrary basis, but most of them based on the prediction of those relationships that may facilitate identification of all source materials that have been indexed.

For example: Instead of **Children's libraries,** we use **libraries and children**

A controlled vocabulary is a part of an artificial indexing language. The notation of a classification scheme is an example of this artificial language. Therefore, it is necessary to know that all classification schemes controlled by the thesaurus based relationships. The thesaurus is a tool that shows explicitly the relationship among the words it contains. The relationships may be of synonym terms, Specific to generic (BT/ broader term), Generic to specific (NT/ narrower term) and General or non-specific relationship (RT/ related term).

(c) **Searching Subsystem:** Searching subsystem is one of the major subsystems of an information retrieval system. In this subsystem, at the beginning users' queries are being received and interpreted by the search system, then appropriate search statements are formulated, and the actual search (i.e., matching queries with the surrogates of information resources file) is conducted with a view to retrieving the required information.

(d) **User-System Interface:** The receiver of information bearing documents becomes a source, encoding the message in form of an inquiry when we discover any information in our store, which appears to match the inquiry, and we can pass them to the enquirer, who can decide whether they match his requirements.

(e) **The Matching Subsystem:** It matches the document representation against request representation that is when documents relevant to query have located, a match has achieved. Search engine acts as a giant matching device. The matching subsystem has no direct influence on effectiveness of the complete system. It plays a great role in overall system efficiency.

5.1.2 Functions of an IRS

The Major Functions of an IRS are:

- To identify the sources of information relevant to the areas of interest of the target users' community.
- To analyze the contents of the sources (documents).
- To represent the contents of the analyzed sources for matching with the users' queries.
- To match the search statement with the stored database.
- To retrieve information which are relevant.
- To make the necessary adjustments in the system based on feedback from the users.

5.1.3 Features of Information Retrieval Systems

Based on the functional aspects two broad categories of Information Retrieval System can be identified: I) In-house and II) Online.

In-house information retrieval systems are set up by a particular library or information centre to serve mainly the users within the organization. One particular type of in-house database is the library catalogue. Online public access catalogues (OPACs) provide facilities for library users to carry out online catalogue searches, and then to check the availability of the required information source.

Online information retrieval systems is one which have been designed to provide access to remote database(s) to a variety of users, via a computer terminal directly interrogate a machine readable database. The main features or characteristics of online IRs are:

- The terminal can be remote.
- Time sharing so several users can be online at one time.
- Information is communicated instantaneously.

There are a number of public as well as commercial databases available in the market and there are a number of vendors that handle this sort of service. With the development of optical storage technology, another type of information retrieval system appeared on CD-ROM (compact disc read-only memory). Information retrieval systems based on CD-ROM technology are available mostly on a commercial basis, though there have been some free and in-house developments too. Basic techniques for search and retrieval of information from the in-house or CD-ROM and online information retrieval systems are more or less the same, except that the online system links to users at a distance through the communication network.

Recent developments in computer and communication technologies have widened the scope of online information retrieval systems. The Internet and World Wide Web have made information accessible to anyone virtually anywhere with the help of an appropriate equipment. This has led to the concept of a digital global library system, where information can generate and made available in electronic form on the Web from any corner of the world. This of course involves a number of technical and management issues that need to be resolved in order to make the global digital library concept a reality.

5.1.4 Types of IR System

The IR system may be of different types depending upon the search conducted by the information seeker. As a whole it can be categorized namely as:

- **Reference Retrieval System:** Information related to specific questions is retrieved.
- **Document Retrieval System:** Information can retrieve by the attributes of documents such as author, title, subject, and so on. Nowadays, complete texts are also retrieved, so called text retrieval system.
- **Fact Retrieval System:** The specific data or facts are retrieved (viz., numerical databases).
- **Knowledge Retrieval System:** Is a rule-based system in which there is a knowledge base with capability for knowledge acquisition and an inference engine.

Review Questions

A. Fill in the blanks

1. Information retrieval system is a system of _____ information.
2. The document subsystem involves the location, selection, ordering and receipt of _____for collection.
3. The matching subsystem matches the document representation against _____ _____ representation.

B. Multiple Choice Questions

Tick the correct answer

1. Retrieval of information related to specific questions is known as:
   (a) Document retrieval system

        (b)   Reference retrieval system

        (c)   Fact retrieval system

        (d)   Knowledge retrieval system

2.   What are the two categories of information retrieval system?

        (a)   In-house and online

        (b)   Off-line and book

        (c)   Online and journals

        (d)   None of the above

C.   Short Answer Questions

1.   What do you mean by the IRS?

2.   What do you mean by indexing subsystem?

3.   What do you mean by searching subsystem?

4.   What do you mean by a document retrieval system?

Checklist for Assessment Activity

Use the following checklist to see if you have met all the requirements for assessment activity.

Part -A

1.   Understood concept of information retrieval system.

2.   Understood components of information retrieval system.

3.   Discussed the features of information retrieval system.

4.   Identified different functions of information retrieval system.

5.   Identified different types of information retrieval system.

Part-B

1.   What is information retrieval system?

2.   Identify different components of information retrieval system.

3.   Discussed features of information retrieval system.

4.   Identified different functions of information retrieval system.

5.   Identified different types of information retrieval system.

Part-C

Performance Standards

The performance standards may include, but not limited to:

Performance Standards	Yes	No
Able to explain information retrieval system.		
Able to identify different components of information retrieval system.		
Able to discuss features of information retrieval system.		
Able to identify different types of information retrieval system.		

5.2 Search Strategy: The Action Plan

Searching is an important function of an IRS. It may be search for information or for documents, such as books, periodicals, patents, theses, etc. Essentially, it is a problem solving exercise. The process of searching in an information retrieval system can be manual or on-line. In any case, the searcher needed information of the user verbalized and translated into a query posed to a search engine. At the end, the matching results produced by the searching process. The search result is heavily dependent upon the correct understanding of the users precise needs and translation of the need into the form of a query. Therefore, a cognitive aspect of interaction between human and search system has always been established to drive out the exact result. A diagram of an ideal search process has given below in the Figure-5.1.
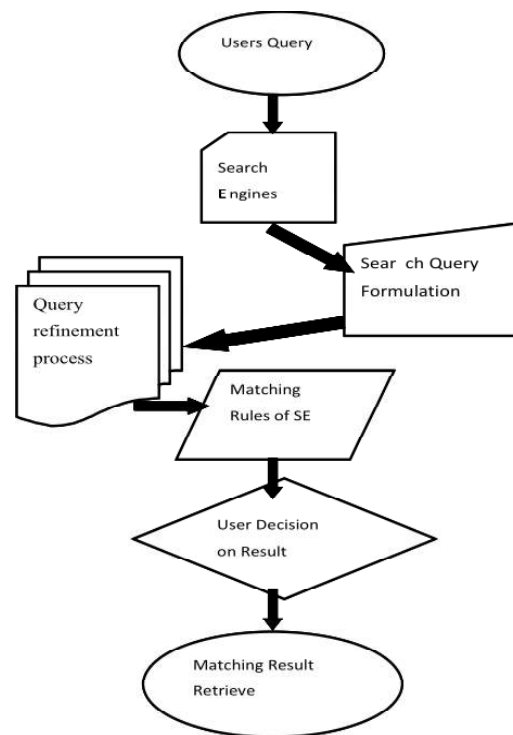

Figure 5.2: Information Search Model

5.2.1 Search Strategy and its Prerequisites

As stated above search strategy is the action plan, which is drawn to conduct a search. It encompasses several steps and levels of work in information retrieval process. Many issues need to be considered while formulating an appropriate search statement. These are:

a)   The concepts or facets to be searched and their order.

b)   The term(s) that appropriately represent(s) the search concept.

c)   The feature(s) of the retrieval system concerned.

d)   The measures to be taken in revising a search statement.

Developing a good search strategy requires knowledge about the nature and organization of target database(s) and the exact needs of the user. Knowledge of the users exact requirement can greatly affect the actual search and retrieval process. In some cases, the user may want only a few relevant items on a given topic, in this case the task of searching will obviously be limited. Conversely, the user may wish to obtain all the relevant items (obviously with as small a number of non-relevant items as possible), in this case the search must be exhaustive. Thus, an information search may fall into one of the following three categories [Chowdhury, 2004];

a)   **High Recall Search :** When the user needs to find out all the relevant items on the stated topic.

b)   **High Precision Search :** When the user needs only relevant items, i.e., as small a number of non-relevant items as possible.

c)   **Brief Search:** When the user wants only a few relevant items as opposed to all the relevant items.

5.2.2 Search Strategy : The Basic Steps

While there are several methods of searching which depends on the information need, file organization and storage media. The basic steps involved in the search process are:

- **Recognise and State the Need** (Includes; developing a query statement and Specify the requirements for the search).

- **Develop the Search Strategy** (This is the query formulation steps, the specification of the conceptual entities and their relationships (based on the query statement) and select the source to be searched and arrange them in a search sequence and translation of the conceptual query into the language of each source or system).

- **Execution of the Search Strategy** (For executing the search, one has to have knowledge of the data structure and search mechanisms of the sources to be searched).

- **Review Search Results** (In reviewing search results, the criteria mentioned namely precision, recall, avoidance of redundancy, novelty, numerical precision, validity, accuracy and display are considered. It helps user to get the matching results or else provide scope for edit/modify the search for further results).

- **Edit Search Results** (By editing it is meant that the transformation of the search results into an usable format. This may involve arranging the results in a well- organised package, Highlighting the important entitles, Adding more information to the entitles and reformating of information to suit the users requirement or need).

- **Evaluation of the Results by the User** (This involves participation of the searcher and/ or the user to quantify his/her search results to use. Obiviosly the quality of the search is the dominant criteria on which one bases searcher' judgement).

However, searching is not a linear process that processes smoothly from one step to the next. There is a certain degree of overlap among the functions. The sequence of steps enumerated above may be followed for better results. At this point, it would be appropriate to mention that the real need of the user (which may be different from the expressed need) may surface only after he has scanned through the search result. In searching for information, the searcher is guided by his/her perception of the need expressed by the user. This perception, in its turn, is governed by the searcher's knowledge of the user and his background, the working situation giving rise to the need, his own knowledge of the subject of the search, the availabilty of information or entitles in general and individual sources-their coverage and searchability, etc.(Soergel,1985).

5.2.3 Search Strategy and the User

In web context, in fact search is pursued by the user; with the intention, intelligent and decision to meet the challenges of information retrieve. We can take the example of descriptions given by Broder, who, categorizes the intentions or purpose framing web query and conducting the search into three categories namely (3) clauses with respect to user intention. These are:

- **Navigational:** The purpose of such queries is to reach a particular site that the user has in mind either because they visited it in the past or because they assume that such a site exists. The immediate intention for such type of query is to reach the particular site at the first attempt.

INFORMATION STORAGE AND RETRIEVAL

- **Informational:** For such type query, the intention is to acquire information assuming it would be present on that webpage or related to that may available. The purpose of such queries is to find exact information assumed to be available on the web in static form so user need to interact or read out the content for getting out result.

- **Transactional:** The intention is to perform some web-mediated activities; the purpose of such queries is to reach a site where further interaction will happen. This interaction constitutes the transaction defining these queries. The main categories of such queries are finding various web-mediated services, downloading various type of file, accessing certain databases etc.

   Therefore, depending upon the approaches of users it would be essential for information designed to evaluate the behaviour patterns of information seekers before architect information for use in web.

5.2.4 Search Strategy: The Basic Steps

Step-I :   **State what you want to find?**

-   At the beginning, the user must choose some keywords, subject description, and/or sentences, what he/she wish to find one.

Step-II :   **Identify the keywords, terms, Phrase**

-   Next, underline the main concept in statement.

   For example; I want to find information on popular methods for losing weight.

Step-III :   **Select Synonyms and Variant words from the List**

-   Listing of synonyms, alternate spelling of each keywords.

   E.g., Popular – Common, Favorite;

   Method – Technique, Ways;

   Loosing – Loss, reduce

   Weight – fat, diet.

Step-IV :   **Combine Synonyms Keywords through operators (OR, NOT, AND)**

-   E.g. (Popular OR common OR Favorite).

-   Use truncation methods (popular*)

-   Popular AND (way OR methods).

Step-V :   **Check your Spelling**

-   In order to get the right results for the query.

5.2.5 Search Strategy: How to Improve

To improve the result of a search, one should follow different search technique (as discussed in 5.4 columns below). These techniques can be operational only when searchers review the query terms. It has been discussed above in search strategy action plan processes (5.2.2) as it may be broadened or narrowed down the search on the basis of the review.
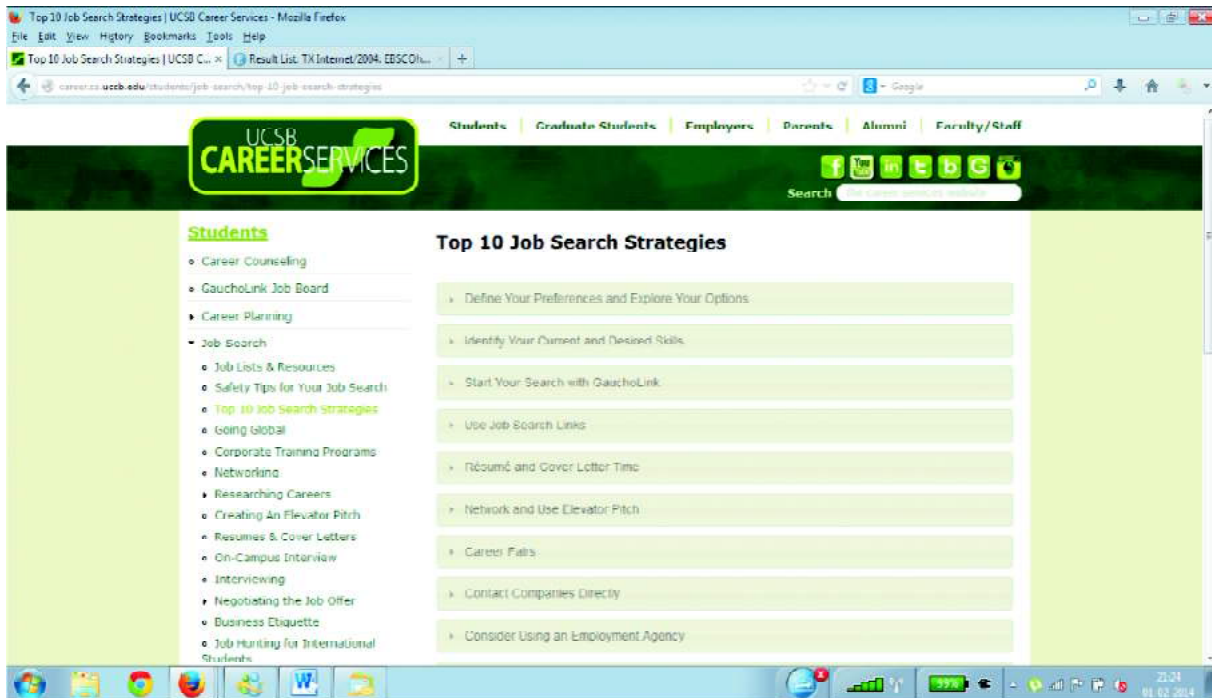
Figure 5.3: Search Strategy

Review Questions

A. Fill in the blanks

1. At the end of the search process, matching ⸺⸺⸺⸺⸺⸺ produced.

2. Developing a good search strategy requires knowledge about the nature and organization of target ⸺⸺⸺⸺⸺ and the exact needs of the users.

B. Multiple Choice Questions

Tick the correct answer

1. Finding out all the relevant items on the stated topic is known as:

    (a) High precision search

    (b) High recall search

    (c) Brief search

    (d) None of the above

2. Specify the requirements and developing query is known as:

    (a) Recognize and state the needs

    (b) Develop search strategy

    (c) Review search result

    (d) Evaluation of result

C. Short Answer Questions

1. What do you mean by search query?

2. What is needed for developing a good search strategy?

3. What is the statement of the need?

4. What do you mean by the development of a search query?

5. What are three different searching intentions of users?

Checklist for Assessment Activity

Use the following checklist to see if you have met all the requirements for assessment activity.

Part -A

1. Understood search strategy.

2. Identified basic steps of search strategy.

Part-B

1. What do you mean by search strategy?

2. Explain different steps of search strategy.

Part-C

Performance Standards

The performance standards may include, but not limited to:

Performance Standards	Yes	No
Able to explain search strategy.		
Able to explain different steps of conducting search.		

5.3 Search Query Formulation Process for Information Retrieval on Internet

While user focuses on "search", the search and matching function are undergoing through the following four phases; that is

a. Document Processor

b. Query Processor

c. Search Matching Function

d. Ranking and Feedback

Each of these four phases may cause the expected or unexpected results that considered getting when they use a search engine.

a. **Document Processor:** During this phase, the robot/spider of a search engine crawls the mentioned query and allow user to locate information by specifying its relevancy. In addition to this, the phase of search query formulation, conflation procedure of Term Stemming (e.g., works, worker, worked, etc. into a single root "work"), and deleting Stop Word (like a, an, the) takes place. Then it extracts the list of terms for the next phase of query processor.

b. **Query Processor:** Query processing has six possible steps includes – tokenizing query stream, parsing to use operator (as such AND, OR, NOT), Index documents by sorting and calculating term weight value. Then, it proceeds to make a reverse index, which allows user to view the particular document with a comprehensive list of terms with their term weights. However, the process of computation during this phase has a myth, that is "a list of documents that contain a particular word would be much more useful, rather than a list of words for a particular document". Thus search system designers must choose "what is most important for their users", that is the time or quality. Mostly, the publicly available search engines choose time over quality, which leads to search too many documents against the query.

c. **Search Matching Function:** This is the process of computing to find out the relevancy of a document, by comparing the term weight or similar co-efficient term related to relevant document retrieved from web resources.

d. **Ranking:** Ranking is the analysis phase which analyses retrieved result and find out its relevancy to that particular query. It could be evaluated through the percentage of Recall and Precision value.

5.4 Search Technique and Methods Used

There are multiple methods to apply, to get a refined result from a query. These techniques or methods are available to conduct searches effectively so that, the maximum relevant information is retrieved according to users need. Some common techniques used for searching almost all types of search engine include:

1. Keywords or String Search
2. Phrase Search
3. Subject Search
4. Boolean Operator Used (AND, OR, NOT)
5. Truncation
6. Proximity Search
7. Limiting Search
8. Range Search

5.4.1 Keyword or String Search

Keyword search is the simplest form of search facility, offered by a search system. In this mode, the system searches the inverted file (the index) for each keyword/term forming the search expression. The search terms can be entered through the keyboard, or selected from an index or vocabulary control tool, such as lists subject headings or thesauri. More than one keyword, forming a search expression has to be combined using the Boolean or proximity operators (discussed below).
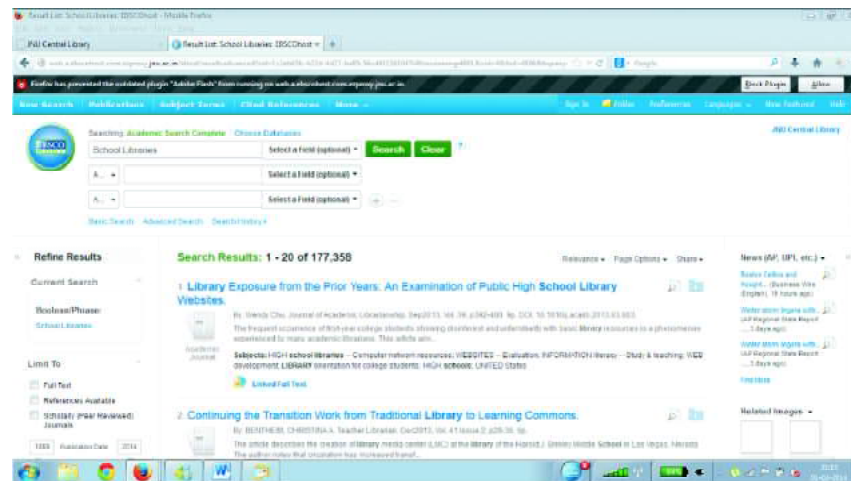
Figure 5.4: String Search

5.4.2 Phrase Search

In a phrase search, the system searches for the entire phrase rather than each individual keyword forming the phrase. Phrase searches can be conducted only on those fields that are phrase indexed. If the index file comprises only single terms, then phrase search cannot be conducted, except using the proximity operators (discussed later in this Unit). Usually, a few fields in a database are phrase indexed; hence, the phrase search can be conducted only in those fields. To search exact phrases, the phrase is being put under quotation mark. For example: **"e-learning", "library management"** and so on.
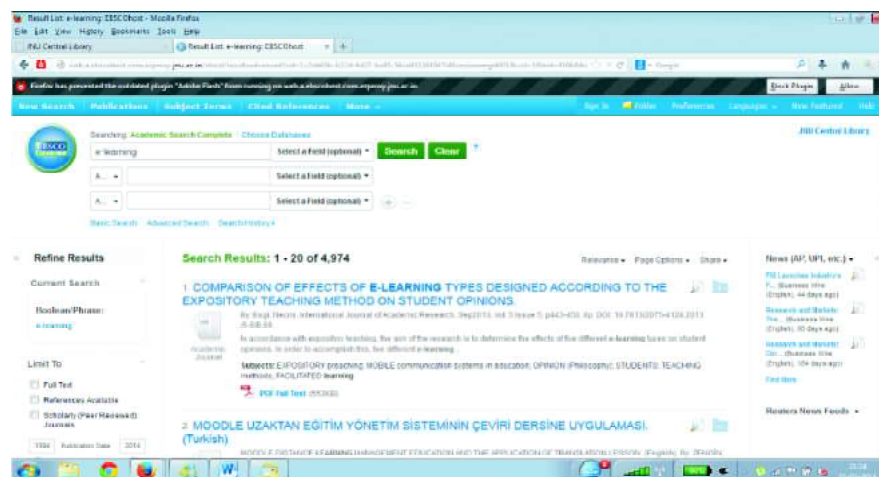

Figure 5.5: Phrase Search

5.4.3 Subject Search

Many bibliographic information retrieval systems provide two types of search facilities, i.e. keyword search and subject search. A keyword search allows users to enter one or more keywords pertaining to their query. Depending upon the requirements, the user can choose these keywords and combine them using different operators. There are a number of search operators that can be used to combine several keywords for formulating a search expression. The search keywords can appear anywhere, or in one or more chosen fields, in the records of the database.

A subject search is the one that allows the user to submit a subject expression that reflects his/ her information requirement. Such a search is conducted on the subject field that contains one or more assigned subject headings entered by the indexer while creating the record in the database. Thus, a record will be retrieved only when user's subject search expression exactly matches the subject heading assigned by the indexer.

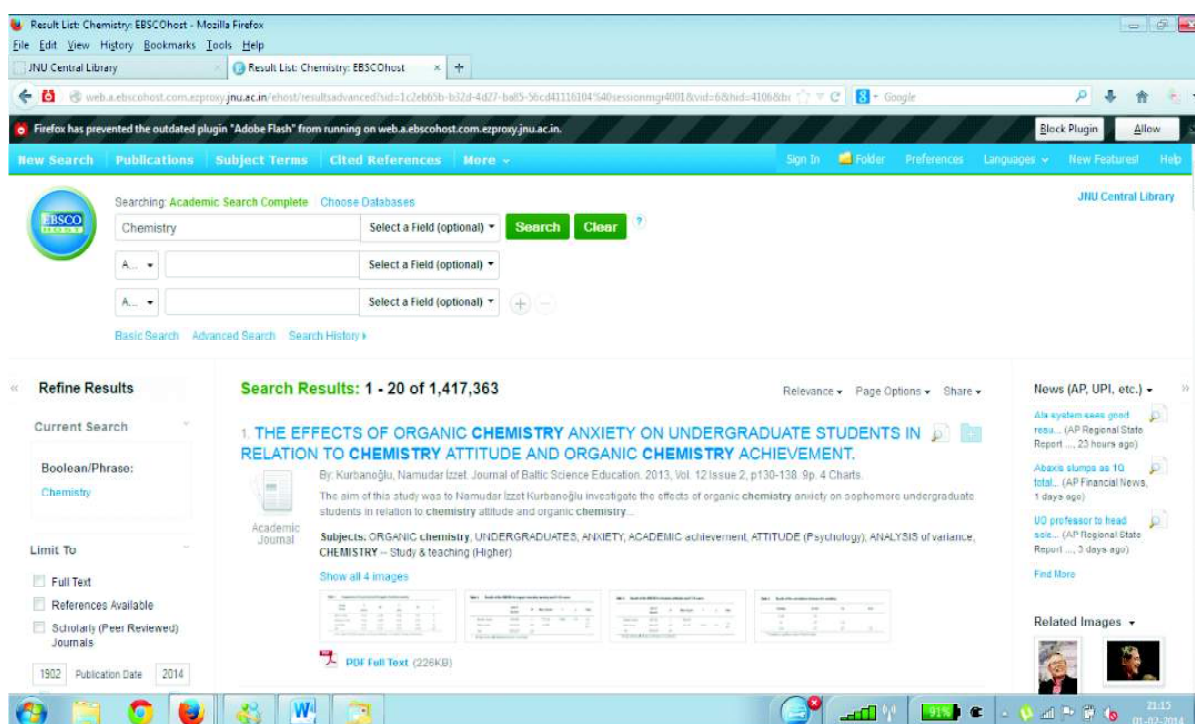**For Example: Physics, Library, Chemistry, etc.**

Figure 5.6: Subject Search

5.4.4 Boolean Search

George Boole (1815-1864) devised a system of symbolic logic in which he used three-operator viz. '+', 'x' and '-' to combine statements in symbolic form. John Venn later expressed Boolean logic relationships diagrammatically. The three operators of Boolean logic are the logical sum (+) - OR, logical product (x) - AND, and logical difference (-) - NOT. All most all the information retrieval system allows the users to express their queries by using these operators. (Choudhury; 1999).

The AND Operators: When two or more keywords/terms are combined using AND operator then the search engine retrieved this information that includes all the terms connected with this operator. Logical products or 'AND' logic allows the searcher to specify the coincidence of two or more concepts.

Example: In order to retrieve information on 'Computers and Information retrieval' the user may formulate the search statements as- (COMPUTERS) AND (INFORMATION RETRIEVAL) and the result will contain all the records having both the keywords.
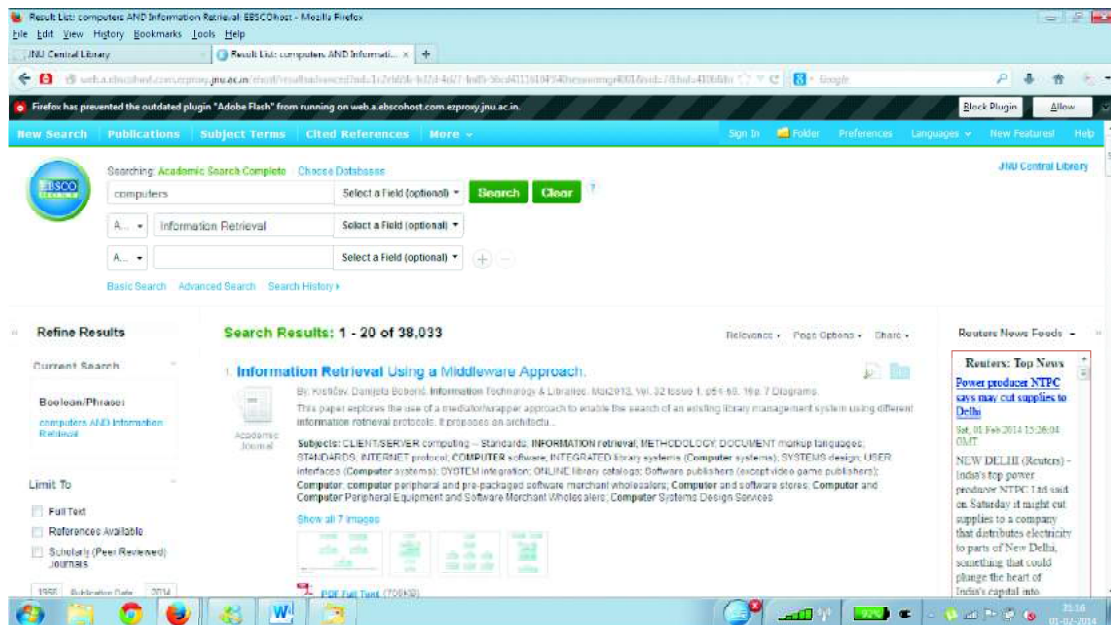
INFORMATION STORAGE AND RETRIEVAL

Figure 5.7: Search Through 'AND' Operator

The OR Operators: Logical sum or OR logic allows the searcher to specify alternatives among the search terms. When two terms are connected with 'OR' and searched then the search engines retrieve those records which have either of terms or both.

Example: The result of the search string "(COMPUTERS) OR (INFORMATION RETRIEVAL)" will contain the records having (COMPUTERS) or (INFORMATION RETRIEVAL) or (COMPUTERS) and (INFORMATION RETRIEVAL) both the set of keywords.
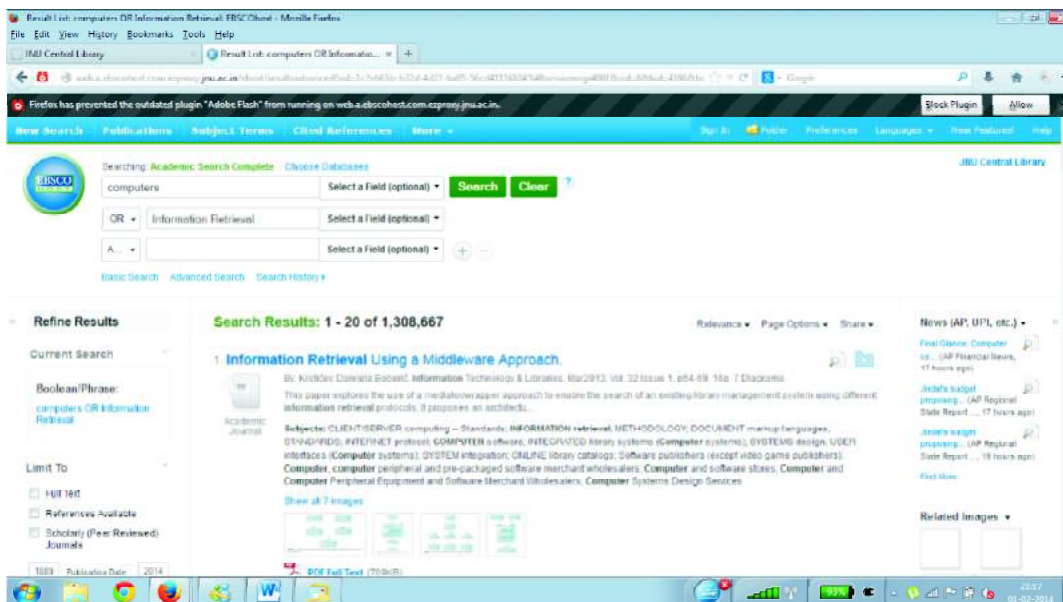

Figure 5.8: Search Through 'OR' Operator

NOT Operators: Logical difference, or NOT logic provide, facilities to exclude items from a set of items. When two terms are connected using the NOT operator and searched then, the search engines retrieve those records which do not have the second term. Example: The search result of "(INFORMATION RETRIEVAL) NOT (DBMS)" will contain only those records which have (INFORMATION RETRIEVAL) but not (DBMS).
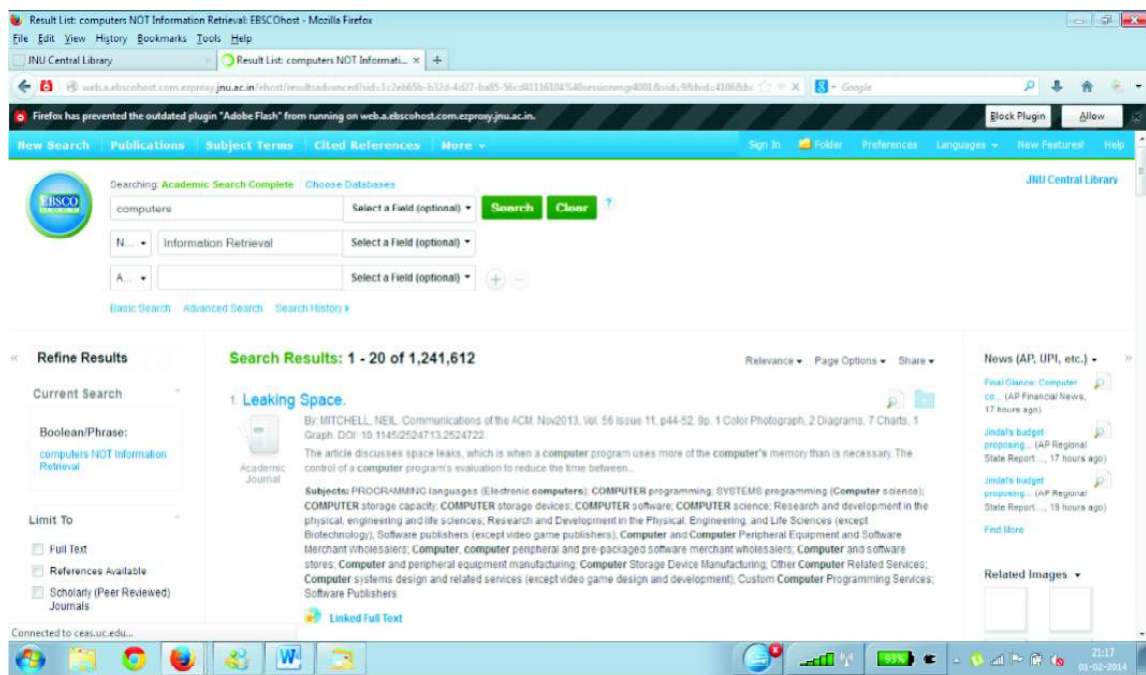

Figure 5.9: Search Through 'NOT' Operator

The graphical presentation of the Boolean logic may as follow:

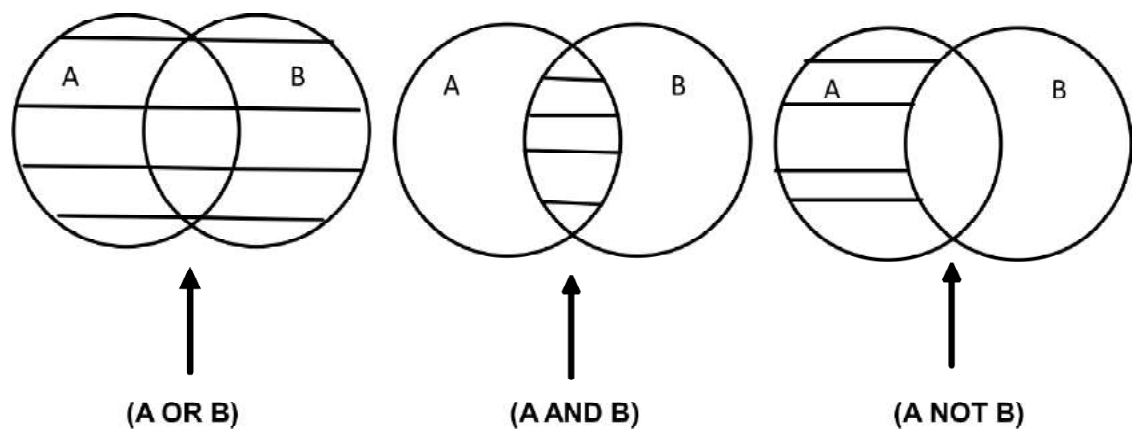**(A OR B)**   **(A AND B)**   **(A NOT B)**

Figure 5.10: Graphic Presentation of Boolean Logic

Here, A and B are two different terms and connected with different operators. The shaded areas show the search results.

INFORMATION STORAGE AND RETRIEVAL

5.4.5 Truncation Search

Truncation allows a search to be conducted for all the different forms of a word having a same common root. It is one of the most widely adopted methods in an information retrieval system especially in searching. It means stripping of the uncommon parts of any terms which have the same portion in common in order to facilitate the process of searching. For e.g. the truncated word COMPUT* includes all terms beginning with COMPUT, whatever may the other letters; thus it covers terms like COMPUTER, COMPUTING, COMPUTATION, COMPUTATIONAL etc. The major advantage of truncation is to decide the root term. E.g. whether to use COMPUT or COMPUTE. In the former case the number of terms covered will be many more thus, the result output may be large. In the second case of relevant items, like computing, computation, computational etc. will be ignored which will affect the result output. Some systems provide facility, it is possible to define the limits of right-handed truncation by multiple use of single character truncation or by using a numerical value. Therefore, a number of different options are available for truncation, viz., right truncation, left truncation, and masking of letters in the middle of the word. Left truncation retrieves all words having the same characters on the right hand part, e.g., '*hyl' will retrieve words like 'methyl', 'ethyl', etc. Similarly, middle truncation retrieves all words having the same characters at the left and right hand part. For example, a middle truncated search term 'col*r' will retrieve both the terms 'colour' and 'color'. A 'wild card' is used to allow any letter to appear in a specific location within a word. Right truncation and character masking or wild cards are the most common truncation search facilities available in search systems. Operators used for truncation search vary from one information retrieval system to another. Most commonly used truncation operators are *, $, !, and ?.
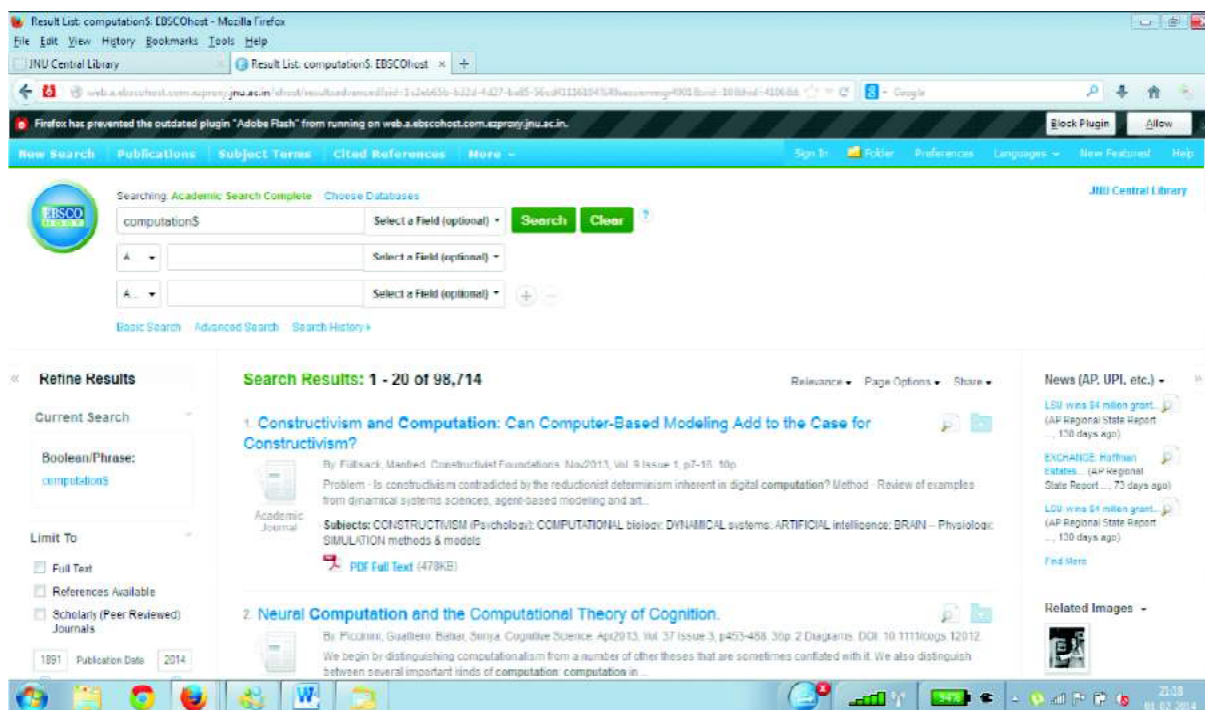

Figure 5.11: Truncation Search

5.4.6 Proximity Search

This search facility allow the users to specify (1) whether two search terms should occur adjacent to each other (2) whether one or more words occur in between the search terms (3) whether the search terms should occur in the same paragraph irrespective of the intervening words and so on. The operators used for proximity search and their meaning differ from one search system to the other. The purpose of proximity search is to refine the search statements by permitting the searcher to specify the context of a term. The following are some examples of proximity search:

- **FISH SAME CHIPS** will retrieve all records where the search terms, i.e. FISH and CHIPS occur in the same paragraphs.

- **FISH WITH CHIPS** will retrieve all records where FISH and CHIPS occur in the same sentence.

- **FISH ADJ CHIPS** will retrieve all records where FISH is followed by CHIPS in the same sentence.

- **FISH NEAR CHIPS** will retrieve all records where FISH is followed by CHIPS or CHIPS is followed by FISH in the same sentence.

Text retrieval software differs in the use of notation for representing various proximity operators. Before using any search engines, it is recommended to read the guidelines of the engine carefully and use the notations which are allowed by the engines only.
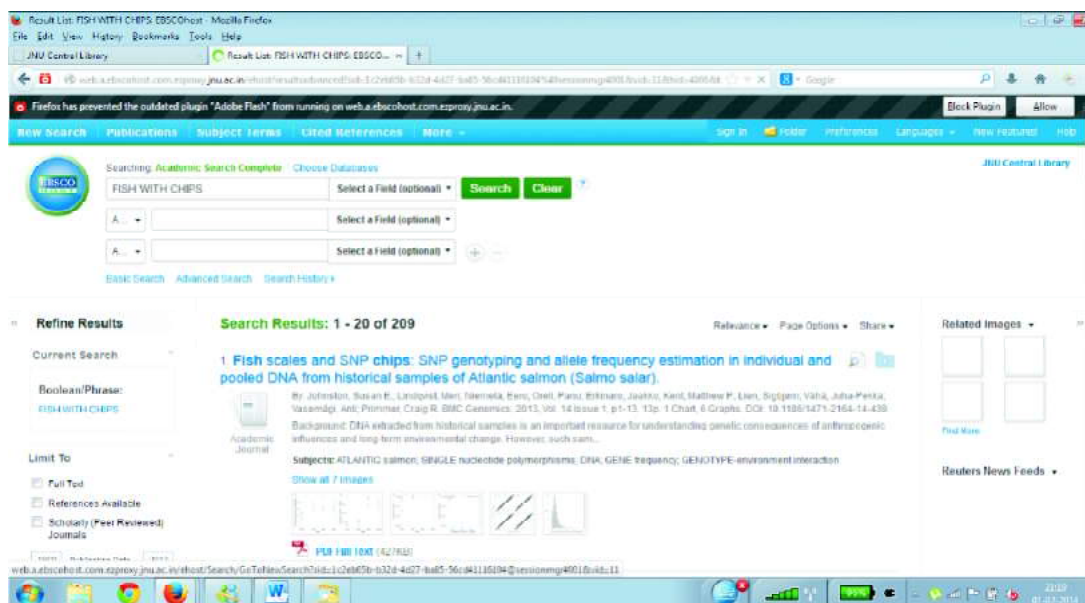

Figure 5.12: Proximity Search

5.4.7 Limiting Search

The database in a text retrieval system comprises different fields containing different items of information. The user in his or her query formulation should be able to limit the search in one or

INFORMATION STORAGE AND RETRIEVAL

more fields such as language, year of publication, type of information sources, and so on and the text retrieval software usually provides this facility. Parameters that can be used to limit a search are decided by the concerned database. This is also known as field searching, whereby the user can specify that the search terms are to be looked for in one or more fields.

Example: Gandhi in author field will bring the result containing records where Gandhi is an author but if the search directed to the title field then the result will contain the records where Gandhi is in the title means the documents written on Gandhi.
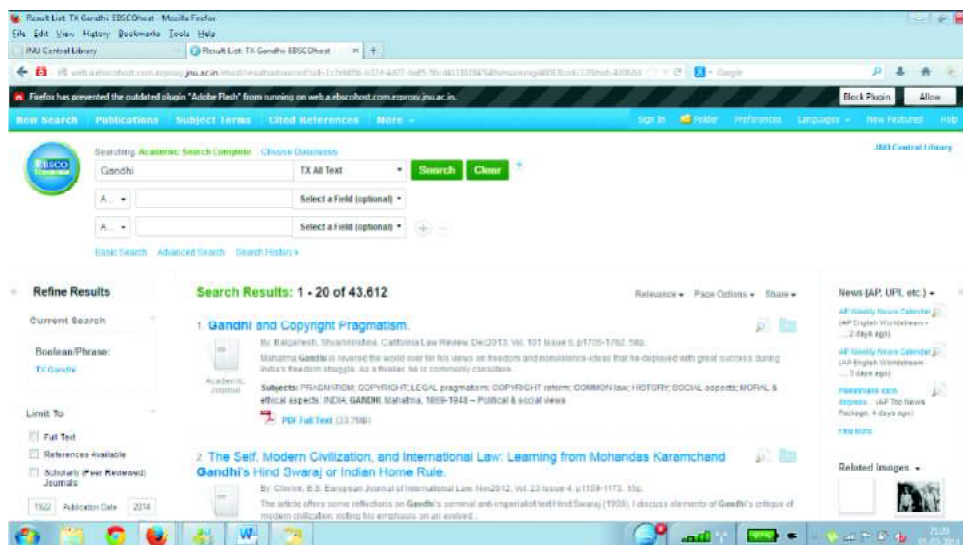

Figure 5.13: Limiting Search

5.4.8 Range Search

Range search is very useful with numerical information. It is important in selecting records within certain data ranges. The following options are usually available for range searching though the exact numbers of operators, their meaning etc., differ from one search system to another.

Greater than (>)

Less than (<)

Equal to (=)

Not equal to (|= or <>)

Greater than or equal to (>=)

Less than or equal to (<=)

These operators are used to prescribe a precise condition in a given search statement.

The following examples of range search are from DIALOG:

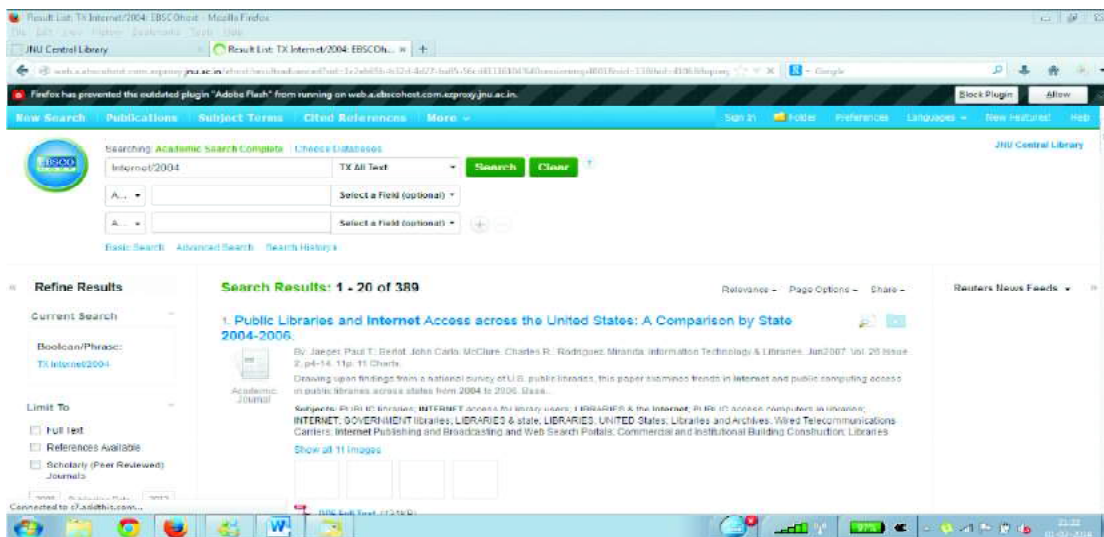| Publication Year	/yyyy	S Internet/2004
	/yyyy:yyyy	S Internet/2003:2004

Figure 5.14: Range Search

Review Questions

A.   Fill in the blanks

1.   Ranking is the analysis phase which, analyses retrieved ————————
and find out ———————— to that particular query.

2.   In the phrase search , the searches for the entire ———————— rather
than each individual keyword forming the phrase.

B.   Multiple Choice Questions

Tick the correct answer

1.   What are different Booolean operators?

(a)   "OF", "TO", "NOT"

(b)   "AND", "OR", "NOT"

(c)   "AND", "OR", "OF"

(d)   "OR", "NOT", "IN"

2.   For searching which kinds of information, range search is usefull.

(a)   Textual

(b)   Images

(c)   Numerical

(d)   Multimedia

C.   Short Answer Questions

1.   What are the different phases of search and matching functions?

2.   What do you mean by string search?

INFORMATION STORAGE AND RETRIEVAL

3.   What is a phrase search?

4.   What are different Boolean operators?

5.   What do you mean by truncation search?

Checklist for Assessment Activity

Use the following checklist to see if you have met all the requirements for assessment activity.

Part-A

1.   Gained skills of formulating search query for search on internet.

2.   Explained different search techniques and their applications.

3.   Performed search on internet using different search techniques.

Part-B

1.   How to formulate search query for search on internet?

2.   Explain different search techniques and their applications.

3.   Perform search on internet using different search techniques.

Part-C

Performance Standards

The performance standards may include, but not limited to:

Performance Standards	Yes	No
Able to formulate search query for search on internet.		
Explained different search techniques and their applications.		
Performed searches on internet using different search techniques.		

5.5 Summary

Searching is an important function of an IRS. The process of searching in an information retrieval system can be manual or on-line. The search strategy is the action plan, which is being drawn to conduct a search to get better results. The Boolean operators provide tools to deal with the linguistic aspects of the terms and create search strings to represent the logic and meaning to the search engine. The notations of operators may differ from system to system but the concept remains same. The knowledge of search techniques and different types of search help searcher finding precise and relevant information from the information system.

5.6 Exercise

1.   What is information Retrieval System in Web Information Environment, explain in brief?

2.   Describe various functions of IRS system.

3.   Write a brief note on the characteristics of IRS.

4.   Write a short note on Subject search.

5.   Give a brief description of truncation search.

6.   What are the Key Components in an IRS system, explain in details?

7. Explain in details the types of Information retrieval system and their use.

8. What are the prerequisies to conduct a search for information explain?

9. Discuss in brief each step of the search strategy.

10. Describe different methods and technique for searching information?

11. How do the Boolean operators change the meaning of a query with it permutation and combination, explain with example?

5.7 Practical

1. Test the various methods and technique of search for information retrieval through following example: Library education in CBSE School.

2. Make search string and conduct search on internet using Google, Google Scholar and Google advance on the topic "ancient history of India" .

   Note:The Ancient India may include today's Afghanistan, Pakistan, Bangladesh, Myanmar.

3. Create a five search strings using Boolean operators and conduct searches on the Library of Congress catalogue of PubMed database which is available on the Internet.

5.8 Glossary

IRS (Information retrieval system): Defined as a "science of search" or a process, method, and procedure used to select or recall recorded and/or indexed information from files of data.

Search: An important function that pursue by the user; with the intention, intelligent and decision to meet the challenges of information retrieve.

Search Strategy: Search Strategy is the action plan, which is drawn to conduct a search.

Search Technique: The methods used to refine any search query; there are many techniques used as such phrase search, keyword search, Boolean operators search, and truncation and so on.

Resource Material

(i) Broder, A.(2000). A taxonomy of web-search, IBM research, 2000 <http://www.sigir.org/forum/F2002/broder.pdf>

(ii) Chowdhury, G.G. (2004). Introduction to modern information retrieval, 2nd Ed. London: Facet Publishing.

(iii) Chowdhury, G.G. and Chowdhury, S. (2001a). Information sources and searching on the World Wide Web, London: Library Association Publishing.

(iv) Chowdhury, G.G. and Chowdhury, S. (2001b). Searching CD-ROM and online information sources, London: Library Association Publishing.

(v) Sorgel, D. (1985). Organising Information: principles of database and retreival systems, London: Academic Press.Library and information science series.

(vi) Sullivan, D. (2003). How search engines rank web pages. <http://www.searchenginewatch.com/webmasters/article.php/2167961>.

(vii) Xie, M; Wang, H and Goh, T. N. (1998). Quality dimensions of Internet search engines. Journal of Information Science, October 1998. vol. 24(5).