

Searching Techniques in the E-Environment: An Overview

D.Kirana kumar¹and Mallikarjun S Manta²

¹Koshys Institute of Management Studies, ²Symbiosis Institute of Business Management, Bangalore – 560100, Karnataka, India.

E-mail: kannada_3044@yahoo.co.in, m.mallikarjun@yahoo.co.in

Abstract - Nowadays availability of Information over the internet is massive. To make maximum use of these information and to avoid disturbance while pointing the relevant information, the user should have search skills that include being familiar with search techniques. The use of search techniques and strategies aims at helping the user retrieve relevant and required quality information. In this paper, information search strategies and search techniques, importance of search preparation and the application of search techniques are discussed.

Keywords: Searching techniques, E-Resources

I.INTRODUCTION

In the recent years the society is entering into the paperless information environment. Information is accessed through the Internet, online databases for journal articles and books, institutional repositories for thesis and dissertations. Due to information explosion and application of Information Communication Technology (ICT) in all the field of knowledge, huge quantity of information is available over online. As a result of this information

explosion, seeking the desired information itself is a big task. Many search mechanisms retrieve a plethora of information resources, but majority are less effective and lack comprehensiveness.¹ In the digital environment search techniques help the information seekers to have successful and focused searches so as to, save time and effort while undertaking online information searches. Starting point of searching is depending on the type of information required. The user may be looking for scholarly information, practice guidelines, information on a certain field of study, such as advertisement management or general information. Depending on the information required, the user may avail the services of general search engines such as Google, Yahoo, AltaVista or specialized search tools such as the map.yahoo.com for maps for driving. A bibliographic database such as Compendex from Elsevier is most important bibliographic database for engineering. It covers all aspects of engineering and manufacturing and includes summaries of journal articles and conference papers.

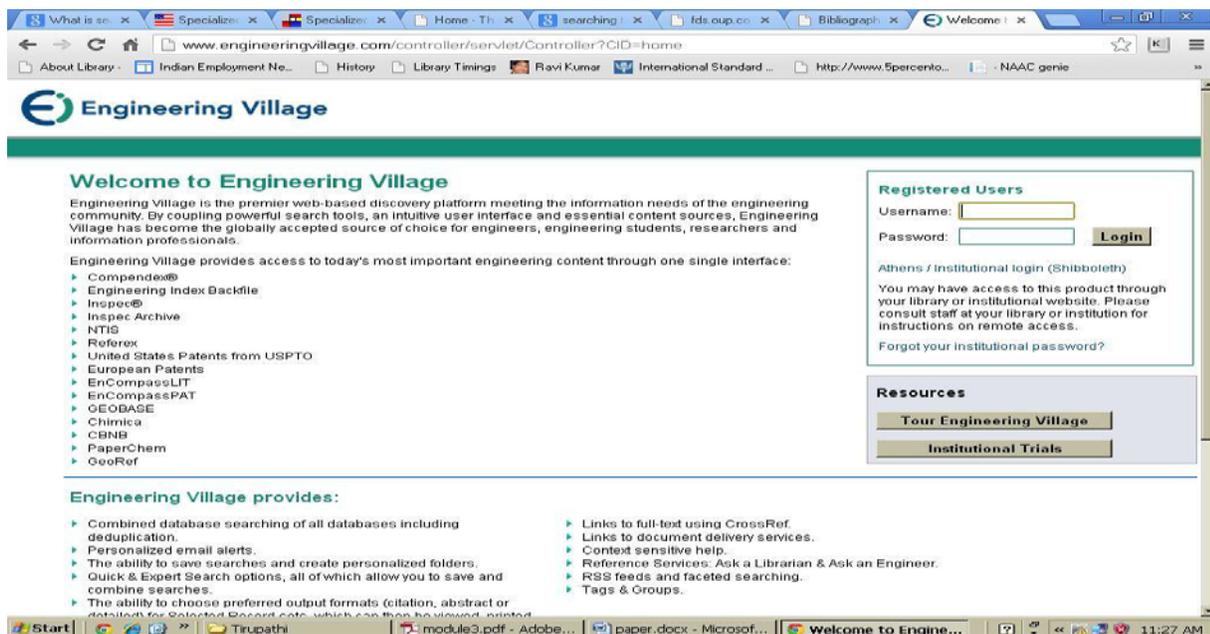


Fig.1 Compendex

A search is the organized pursuit of information; somewhere in a collection of documents, Web pages, databases and other sources, there is information that the user wanted to find. Searching is a powerful phenomena, that allows the user to easily find information in a web/database. Most searches work by asking the users to enter a keyword. The search engine then looks through the entire web/database, and returns to the user with all the pages/fields from web/database that contain the keyword user typed in. Some searches are more sophisticated which allows the users to enter more than one word to customize the search.

II. TYPES OF SEARCH QUERIES

Mr. Andrei Broder identified Three types of search queries namely

1. Navigational
2. Informational
3. Transactional

Navigational query-is a search query entered with the intent of finding a particular website or webpage. For example, a user might enter "youtube" into Google's search bar to find the YouTube site rather than entering the URL into a browser's navigation bar. In fact, "facebook" and "youtube" are the top two searches on Google, and these are both navigational queries.

Informational queries-An informational query indicates the searcher is looking for specific information. For example, "effects of FDI on retailing", "Digital library of India" or "Open Source Library Automation software's".

Informational queries tend to be broad. The Informational query doesn't tend to be site specific.

Transactional search query- is a query that indicates intent to complete a transaction, such as making a purchase. Transactional search queries may include exact brand and product names (like "samsung galaxy note 2") or actually include terms like "buy," "purchase," or "order." In all of these examples, you can infer that the searcher is considering making a purchase in the near future. ³

III. TYPES OF SEARCHES

There are different search options depending on the level and amount of information that is of interest to the user. To search is to carefully look for something and effective searching requires planning. Good searches are planned not accidental.

1. Simple search
2. Advanced search
3. meta search

A *simple search*-is when the user uses some keywords to perform a quick information search from a database or from a search engine. A simple search may retrieve a huge amount of search output that may take time to sieve through for any relevant information item. In simple searches, browsing rather than focused searches is done. In simple searching one or two concepts can also be used. The user may use simple searches when not sure of the type of information required and when the topic is not focused on any area.

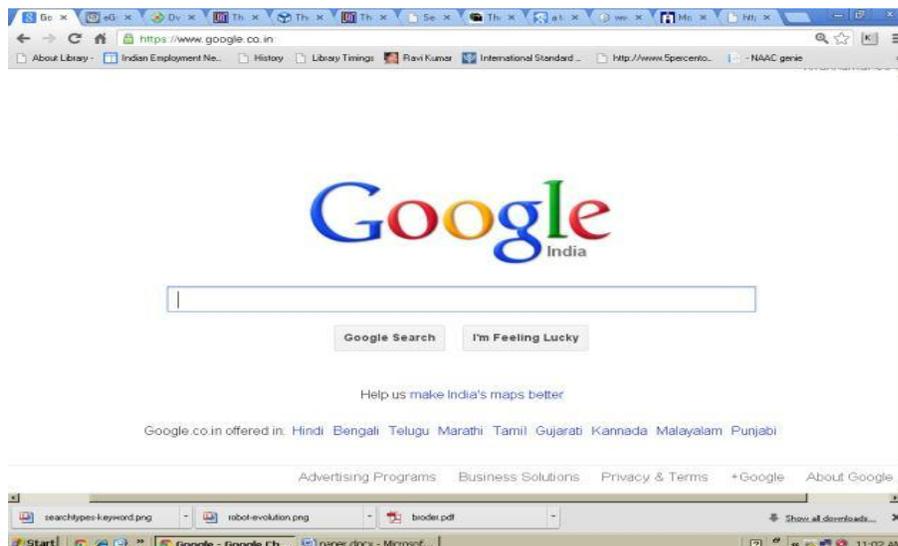


Fig.2 Simple Search

A topic such as Library professional in Africa can cover any period, can affect, technical or non-technical, male or female and can be from English, French or Portuguese

speaking Africa. This simple search can also retrieve information on Africa not related to Library professional. the search results may be overwhelming and the user would

be required to apply search techniques to focus the search for relevancy.

Advanced search: Advanced searching is the use of techniques that help to define the information that is being searched. In advanced searching, some filtering is done to reduce the amount of items retrieved. Filtering refines the search for relevancy i.e precision value of the search is increased. In advanced searching the user is able to apply

multiple search fields that can help to broaden or narrow the search depending on the topic and the search strategy. The use of advanced searching helps the user :-

1. Apply filtering
2. Reduce number of items retrieved
3. Apply multiple search fields

Advanced searching tools are available in most search tools such as eGyanKosh.

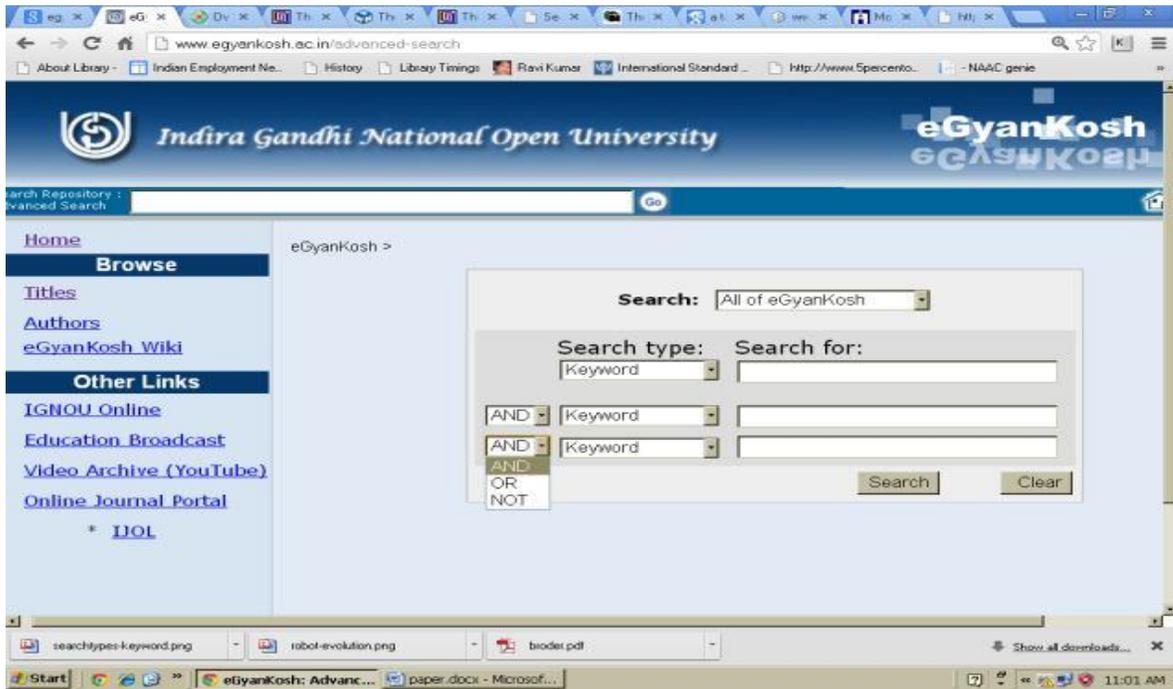


Fig.3 Advanced search

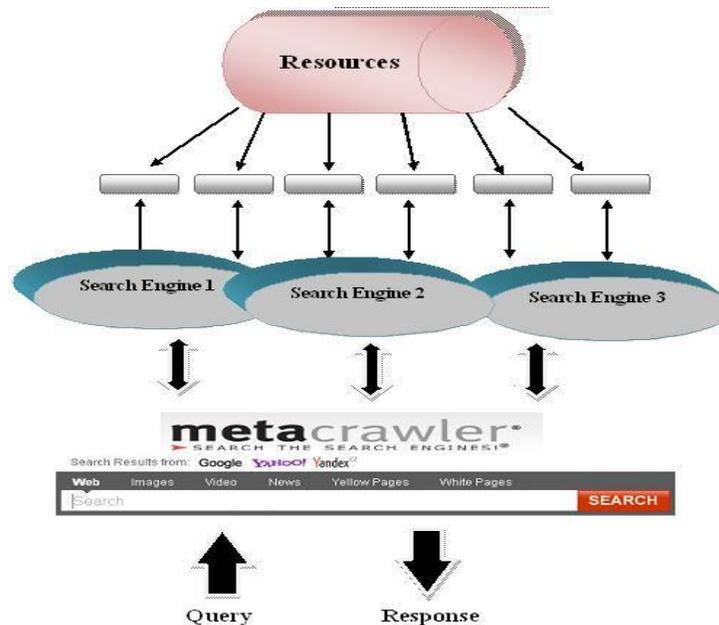


Fig.4 Meta Search

Meta search

A Meta search is when the user uses a variety of search tools simultaneously. This is done with the use of Meta search engines such as Dogpile, Search.com; Metacrawler and Vivisimo. These Meta search engines searches many search tools such as Google, Yahoo, and MSN among others. Searching from a Meta search engine is conducted in the same way as in single search engines. A search can be a simple search or an advanced search. The only difference between Meta search engines and single search engines is that the search results will come from different search engines simultaneously.⁴

Use of Keywords

Keywords are the words and phrases that are used to closely describe the topic or subject by the author or through indexing. The keyword search method is the most flexible type of search to use in almost any database because it allows the user to search for any terms that may appear anywhere in a record. This flexibility is the greatest advantage of using this type of search. The disadvantage is that the search results may display records or pages that have nothing to do with the user’s query. In using keywords, it is important to consider related terms such Sources or Resources, variations in word spelling such as US and UK versions such as mailbox or postbox, plural or singular versions, synonyms such as nascent or emerging, or use of controlled vocabulary that gives alternative terminology to a word e.g. Cancer is referred to as ‘neoplasm’ in controlled vocabulary terminology.

The difference between the use of plural and singular in search terms can produce varying results as can be seen by this search:

Effects of FDI on retailing– yahoo= 2,790,000, Google= 1,450,000, Bing= 27, 40,000

Effect of FDI on retailing– yahoo= 2,830,000, Google= 1,910,000, Bing= 27, 90,000

The use of singular “effect” instead of plural “effects” has increased the amount of retrieved information significantly in three search tools, Yahoo, Google and Bing⁵

Controlled Vocabulary

The use of controlled vocabulary is to give uniformity and consistency to the indexing of the literature. The controlled vocabulary ensures consistency in presentation of information for related topics. It also tries to standardize the language used by authors. The Medical Subject Headings (MeSH) is a distinctive feature of MEDLINE- the foremost biomedical literature database. It is the most used controlled vocabulary or thesaurus for biomedical literature. It is used in the Medline database. MeSH is used in the preparation of Medline and a search tool through PubMed. In Medline/Pubmed, only records that have gone through the indexing process are assigned MESH terms. Those records that are not Medline records may not have the MeSH terms and searching through MeSH will exclude such records.

Case Sensitivity

Electronic information can be presented in upper and lower case. It is advisable to control the use of upper and lower case while searching for information. Some search terms such as names of people and places may require the use of capital letters. However, the use of uppercase will retrieve only those words that are presented in uppercase. Using lowercase at all times helps retrieve information whether the information is presented in capital letters or not. Users should know when to use upper and lowercase letters. The use of lowercase should be preferred to avoid missing out useful information.

Use of Abbreviations

Some search terms are commonly presented as abbreviations. For example the use of HIV instead of Human Immunodeficiency Virus may affect the end results. The user has to be vigilant on the effect of such abbreviations on the outcome of the search.

Search Result of Human Immunodeficiency Virus— Google= 9,510,000, Yahoo=1,160,000, Bing= 1,170,000
Whereas Search Result of HIV-- Google= 44,200,000, Yahoo=5,280,000, Bing= 5,260,000

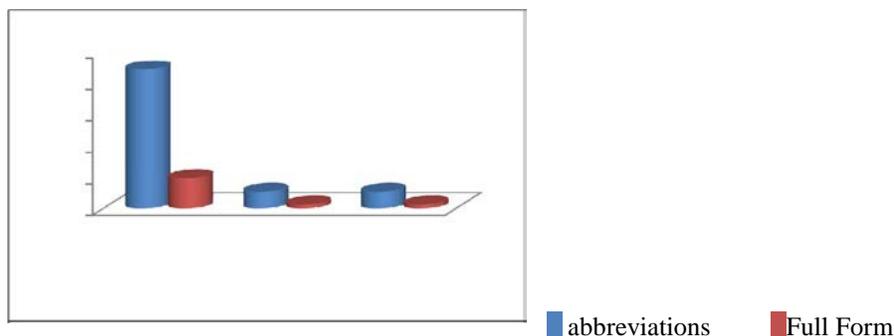


Fig. 1 Abbreviations V/s Full form

IV. SEARCH STRATEGY

This is a plan that helps the user to articulate the type of information that is required. A search strategy is a plan that guides the user to answer some questions such as:

1. What is the purpose of the information?
2. Is there a time frame for the required information?
3. Is the information required general or specific?
4. What sources would best retrieve the required information, general or specific tools?
5. Would the information be specific to a certain geographical location?

A search strategy relates to having good guidelines that will lead to a successful search output. Having a search strategy helps the user to:

- a. Define the topic e.g. Malaria treatment in infants in Africa.
- b. Break the topic into concepts or keywords such as: malaria, treatment, infants, Africa.
- c. Use search techniques such as Boolean operators to

- d. refine the search
- d. Try out the search strategy and refine it as necessary for better results
- e. Identify the appropriate search tool and search techniques
- f. Decide on whether to start with a general search engine such as Google or a specific tool such as Medline/Pubmed, or a more specialised database such as Cochrane
- g. Have alternative choices such as starting from print-based information sources such as reports or grey literature.

Search Term Selection

A search term is any word or phrase that is significant or meaningful to the topic. Usually, a search term is a noun or an adjective, and sometimes a verb that are concrete concepts. Before choosing search terms, searcher should articulate his/her information need. Having a clear idea of the scope of the topic and what, if any, aspects of the subject area searcher wish to focus on make it easier to search for information.

Examples of related words			
	International	multinational	worldwide
Standards	Policies	guidelines	protocols
Environment	Ecology	wildlife	ecosystem
Protection	Securing	conservation	regulation

Search Terms to Avoid

Some words entered into the search statement may prevent the searcher from retrieving effective search results. The following types of terms are best avoided when selecting the search terms.

Stop words are words that have a high frequency of use in the English language, called stop words because they will slow down or invalidate a search. Many databases and search engines are programmed to ignore these words.

Examples:

Articles (a, an, the)

Prepositions (of, on, in, with)

Conjunctions (but, however)

Common words are often terms so heavily used in a given database that they will retrieve too many records to be useful. For example, using the search term "education" in the *Education Full Text* database would retrieve almost every record in that database.

Abstract Words Or Concepts can be difficult to describe search terms, such as "causes of" or "impact on." These are a reflection, analysis, and evaluation of your research rather than a concrete, descriptive term. Use of these typically generates invalid search results.

Some of the abstract terms to avoid include:

Advantages Of Search Strategy

- a. Saves time
- b. Helps retrieve relevant information
- c. Can be stored for later use.

Requirements for a Successful Information Retrieval

- a. Defining the search problem
- b. Developing a search strategy
- c. Using search techniques
- d. Learning how to use a search tool e.g. PubMed
- e. Using correct spelling
- f. Considering other forms of information sources such as print and audio-visual
- g. Evaluating the retrieved information
- h. Using the information appropriately

V. SEARCH TECHNIQUES

Search techniques are routes of using search terms in seeking the required information from search tools. Search tools are many, e.g. Online Public Access Catalogues (OPAC), general search engines, search directories and portals as well as online databases or deep web. To get good

search results, it is necessary to use search techniques. The following are some of the most common search techniques that are applicable to various searching tools.

Boolean Logic

- a. Parenthesis/nesting
- b. Phrase searching
- c. Truncation
- d. Wildcards
- e. Field searching
- f. Fuzzy searches
- g. Proximity searches
- h. File types

Boolean logic

Boolean Logic is a logical relationship of search terms. It is named after the British mathematician George Boole (1815-64). Boolean logic is used to narrow or focus the search or to broaden the search depending on the information that the user wants to retrieve. Boolean logic helps the user to achieve highest precision value in the search results and it also assists in quick retrieval of information. The logic uses the following operators

- a. || OR - Broadens the search
- b. + AND – Narrows the search
- c. NOT – excludes unwanted words or concepts to focus the search

Note: It is also important to keep in mind that most Search Engines require that

Boolean operators be typed in capital letters.

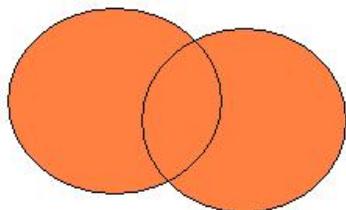
The OR operator is the default conjunction operator. This means that if there is no Boolean operator between two *Note:* The NOT operator **cannot be used with just one term**. For example, the following search will return no results:

NOT "Automated Library"

As a workaround QUOSA from Elsevier adds one word **exclusionpattern** to each and every document it indexes. As a result to find all documents that do not contain phrase "Automated Library" searcher can use this query:

Difference in the search results

Automated Library OR Digital Library = **105,000,000**



Automated Library NOT Traditional Library = **19,500,000**

terms, the OR operator is used. The OR operator links two terms and finds a document if either of the terms exists in it. The symbol || (two bars) can be used in place of the word OR. For example, to search for documents that contain either "Virtual Library" or just "Library" specify the search expression as follows:

"Virtual Library" Library or
 "Virtual Library" OR Library
 AND (or the && symbol)

The AND operator finds documents where both terms exist anywhere in the text. The symbol && (two ampersands) can be used in place of the word AND.

For example, to search for documents that contain "Automated Library" and "Digital Library" specify the search expression as follows:

"Automated Library" AND "Digital Library"
 + (plus sign)

The "+" operator (known as the required operator) requires that the term after the plus sign exist somewhere in a document.

For example, to search for documents that *must* contain "Digital" and may contain "Library" specify the search expression as follows:

+Digital Library
 NOT (or the ! symbol)

The NOT operator excludes documents that contain the term after NOT. The symbol "!" (Exclamation point) can be used in place of the word NOT.

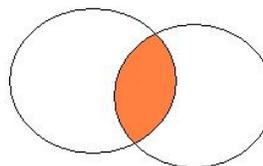
To search for documents that contain "Automated Library" but *not* " Digital Library" searcher has to specify the search expression as follows:

"Automated Library" NOT "Digital Library"
 exclusionpattern NOT "Automated Library" - (minus sign)

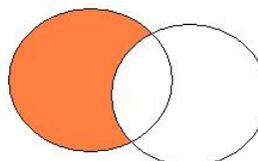
The "-" (minus sign) or prohibit operator excludes documents that contain the term after the minus symbol.

For example, to search for documents that contains "Automated Library" but *not* "Digital Library" specifies the search expression as follows:

"Automated Library" – "Digital Library"



Automated Library AND Digital Library = **9,580,000**



Parenthesis/Grouping/ Nesting

This search technique is used to force the order of how information is retrieved. For example the retrieval of information related to diabetes prevention and control can be searched as following.

VI.DIABETES AND (PREVENTION OR CONTROL)

The keywords in the brackets will be searched first and then matched with the results of the keyword outside the bracket. In the absence of brackets all the keywords will be searched together from left to right. The use of brackets controls how the search is executed. Results returned through the use of parenthesis are more relevant. Parenthesis uses more than one Boolean operator and is used to perform complex searches.⁷

The user is advised to check if the search tool in use allows the use of Parenthesis/ Grouping

Phrase Searching

This technique surrounds the search keywords in quotes "...". It instructs the search engine to search only words that appear side by side inside the quotes. This method narrows search results significantly leading to more relevant results. An example of a phrase search: Working Conditions of "Librarians" in Karnataka would only retrieve records focusing only on Librarians..

A search example from Google on Working Conditions of "Librarians" in Karnataka as a phrase retrieves about 531,000 results while as a search statement on Working Conditions of Librarians in Karnataka 35,600,000 results. There is huge difference in the amount of records retrieved without the use of the quotation marks.

Truncation

Truncation is the process of using an asterisk mark (*) while searching. The purpose of truncation is to broaden the search results. Through the use of truncation at the root of the word, the search tool searches all word variations after the asterisk eg. the use of * with Child* will retrieve child, children and childhood. In the example below,

- Child* retrieves 16,150,000,000 records
- Child retrieves 1,530,000,000 records
- Children retrieve 2,410,000,000 records
- Childhood retrieves 255,000,000 records

Depending on the root word and the place where the asterisk is placed, use of truncation can sometimes retrieve some irrelevant results. Each search tool treats the use of search techniques differently.

The user is advised to check if the search tool in use allows the use of truncation.⁸

Wildcards

Wildcards are characters that are used to assist in searching for information. Wildcards are used to represent one character or letter in a word. Wildcards are especially useful in situations when the searcher is not sure of the correct spelling of a word. It is advisable to check the wildcard symbols used by each search tool or search engine. The most commonly used wildcard is "?". Below is an Example:

- An?mia will retrieve all records that contain either anaemia or anemia
- Colo?r will retrieve all records that contain colour and color

However most search engines do not currently support the wildcard facility

Field Searching

An electronic record is presented as a field. Within a bibliographic database, there are data fields for the title, author, affiliation, journal or book title and language, among others. The user can use any of these fields to retrieve the required information.

As an example, a user can search for an article by a certain author on a specific topic such Dublin Core use in libraries: a survey by Carolyn Guinchard. This is a focused search through the author field. The user can also focus a search through the use of the title, subject or publication type such as journal article. In most search tools, there is also the possibility of limiting the searches to such fields as year of publication, language, affiliation, review articles, age, gender or just the abstract of the article. Field searching, therefore, focuses on specific information that the user may require. Field searching is mostly available through the advanced searching mode of most of the search tools.

Fuzzy Searches

A fuzzy search can be used to find words similar in spelling. To create a fuzzy search, add the "~" (tilde) symbol at the **end** of a *single-word* term. For example, to search for a term similar in spelling to "roam," specify the fuzzy search as follows:

roam~

This search will find terms such as "foam" and "roams."

Proximity Searches

A proximity search can be used to find words that are within a specific distance to other words. To create a proximity search, add the "~" (tilde) symbol at the end of the words.

For example, to search for the words "heat" and "shock" within 10 words of each other in a document, specify the search as follows:

"heat shock"~10

File Types

File type search can be used to find specific kind of files such as .PDF, files, or Word documents, or Excel spreadsheets, for example, use the "filetype:" operator. Example:

Working Conditions of Librarians in Karnataka "PDF"

V.CONCLUSION

Searching the World Wide Web/Online database successfully is the basis for many of our information tasks today. Search engines are thus increasingly being relied upon to extract just the right information from a vast number of databases. . The database/ web search engines are limited in terms of coverage, recency, how well they rank query results, and the query options they support. Access to the growing body of scientific literature on the publicly indexable web/database search engine is limited by the lack of organization and because the major search engines do not index Postscript or PDF documents. We have discussed several fruitful search techniques that will improve access to general and scientific information, and greatly enhance the utility of the needed information.

REFERENCES

- [1] Understanding metadata. (2004), Bethesda, USA: NISO. Available at <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>
- [2] Wilson, Jim (2005). "What is Search" available at http://www.nasa.gov/search/tips/SE_INT_What_is_Search.html
- [3] Andrei Broder (2002), "A taxonomy of web search" IBM Research <http://www.sigir.org/forum/F2002/broder.pdf>
- [4] R. Lempel, and S. Moran. Predictive caching and perfecting of query results in search engines. In WWW, pp 19-28, 2003
- [5] M. Marin, F. Ferrarotti, M. Mendoza, C. Gomez and V. Gil-Costa. Location cache for Web queries. In CIKM, pp. 1995-1998, 2009.
- [6] Baker, Joe, "Basic search Tips and Advanced Boolean Explained" available at <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Boolean.pdf>
- [7] Using AND, OR, and Not (Boolean Operator), Mathewson-IGT Knowledge centre, University of Nevada, Reno Available at <http://knowledgecenter.unr.edu/help/using/boolean.aspx>
- [8] Sandeep Kumar Satapathy, Shruti Mishra and Debahuti Mishra (2010). "Search Technique Using Wildcards or Truncation: A Tolerance Rough Set Clustering Approach", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 1, No.4, pp 73-77