Trends in Engineering and Technology (NCTET-2K17)
International Journal of Advanced Engineering Research and Science (IJAERS) *Special Issue-3*
https://dx.doi.org/10.22161/ijaers/nctet.2017.25 *ISSN: 2349-6495(P) | 2456-1908(O)*

Information Retrieval systems and Web Search Engines: A Survey

R. Arun Kumar, M. A. Jabbar, Y.V. Bhaskar Reddy

[1]Department of CSE, Vardhaman College of engineering, Hyderabad-501218
Email: arunravula12@gmail.com
[1]Department of CSE, Vardhaman College of engineering, Hyderabad-501218
Email: jabbar.meerja@gmail.com
[1]Department of CSE, Vardhaman College of engineering, Hyderabad501218
Email: yaramala.vijay@gmail.com

Abstract:
Information retrieval systems (IRS) are field concerned with retrieval of information. A search engine is the application of IR techniques. A web search engine is a tool to find information on the www. Search engines are updating their index to the World Wide Web. In this paper we review search engines and their architectures in brief and suggest some guidelines for the users.

1. Introduction:

An information retrieval system (IRS) is a field concerned with retrieval of information. Information retrieval includes a wide range of information. Applications of information retrieval involve multimedia documents with structure, significant text content, and other media. Dimensions of IR are listed in table 1.

Table 1. Some dimensions of information retrieval [1]

Examples of Content	Examples of Applications	Examples of Tasks
Text	Web search	Adhoc search
Images	Vertical search	filtering
Video	Enterprise search	classification
Scanned documents	Desktop search	Question answering
Audio	Peer to peer search	music

Relevance is an important concept in information retrieval. Search engines are constantly building and updating their index to the World Wide Web. Spiders are used to crawl the web and fetch web pages. The words used in these web pages are added to the index along with where the words came from.

Web search engines, must be able to crawl, and should provide respond to millions of queries .To engineer a search engine is a challenging task. Web Search engines index millions of web pages. The World Wide Web Worm (WWWW) had an index of 110,000 web pages [2]. With the increasing number of users on the web, and due to automated systems which query search engines, top search engines are handling hundreds of millions of queries per day.
Today, most of the search engines are based in the U.S. The users search documents by keywords. However, there are other search engines in other languages such as Chinese, Korean, and Japanese.

Problems faced by users when facing search engines.

1. The users generally do not know how to Search.
2. The user cannot perform advanced searching
4. Many users only look at the first page [3]

Guidelines helping users to search.
1. Clearly specify the words
2. Provide as many particular terms as possible
3. Some search engines are specialized in some areas.

2. Search Engine Architectures

https://dx.doi.org/10.22161/ijaers/nctet.2017.25 *ISSN: 2349-6495(P) | 2456-1908(O)*

This section discusses architecture of search engines.
Architecture is designed to satisfy the goals. Goals of a search engine
• Effectiveness: retrieve the relevant set of documents possible for a query.
• Efficiency: process queries from users [4]

Actions performed by the search engine

1. looks for the keyword in the index
2. Web crawler search for the information.
3. search engine shows the relevant web pages after web crawler finds the pages

Major functions supported by search engines

1) Indexing process 2) Query process.
The indexing builds the structures, and the query process produce documents in ranked order

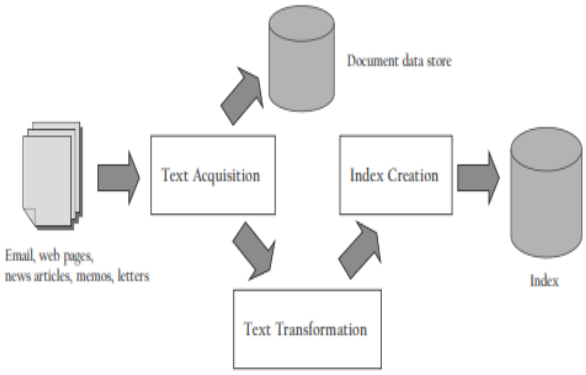


Figure 1.Index process

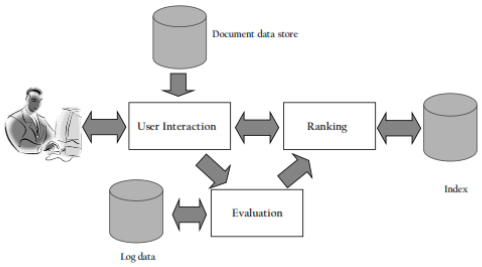


Figure 2 query process

Following are the several search engines available today:

Table 2: Search engines

Search Engine	Description
Google	Google is the most popular search engine globally.
Bing	Launched by **Microsoft.** Also delivers Yahoo's results.
Ask	It was originally known as **Ask Jeeves.**
AltaVista	It was launched by **Digital Equipment Corporation**, it is powered by Yahoo
AOL.Search	It is powered by Google.

2.1. AltaVista Architecture

The crawler sends requests to remote Web servers. The index is used to reply queries from users. Figure 3 shows software architecture of AltaVista. Query engine and user interface are in first part. The second part contains the indexer and crawler [5].

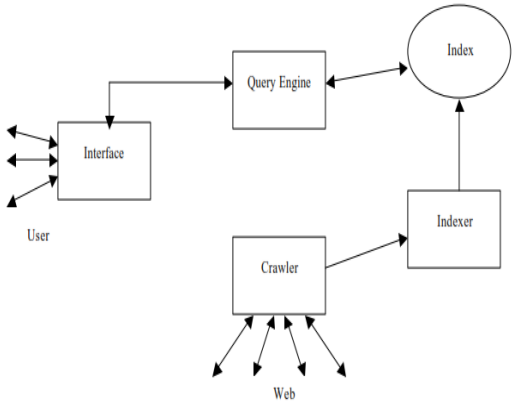
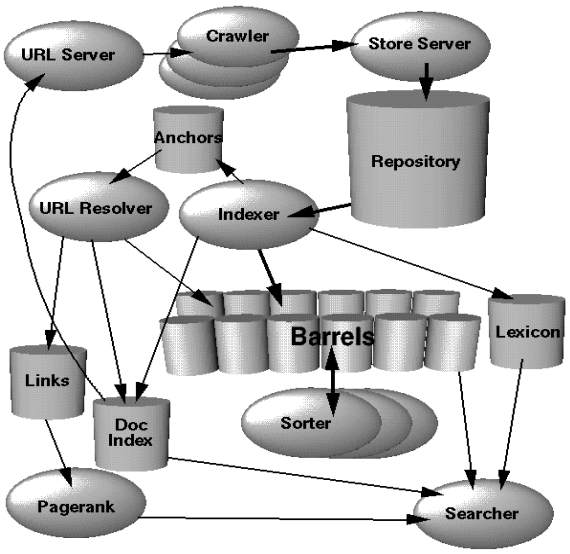


Figure 3: Architecture of AltaVista search engine.

2.2. Google

The word Google comes from the word googol, which means 10^{100}. 64.0% of searches were powered by Google. [6].
Google is written in C/C++.
Figure 3 shows architecture of Google..

https://dx.doi.org/10.22161/ijaers/nctet.2017.25 *ISSN: 2349-6495(P) | 2456-1908(O)*



2.3. Bing

Bing is Microsoft's attempt to challenge Google in the area of search .

2.4. Yahoo

Yahoo search is powered by Bing. Yahoo is still the most popular email provider.

2.5. Ask.com

Formerly known as Ask Jeeves. ASK is based on a question/answer format. It lack quality compared to Google,Bing and Yahoo.

2.6. DuckDuckGo

Have a number of advantages over the other search engines. 1) It has a clean interface 2) it does not track users 3) it is not fully loaded with ads

3. Conclusion

Information on the web is diverse in content and catering to the different information need of users is a big challenge. In this research paper we found that the users on average use two to three keywords query for search and there is vocabulary gap between user query and keywords used in the document.

References

[1] W. Bruce Croft et.al Search Engines Information Retrieval in Practice, Pearson education 2015

[2] McBryan 94 Oliver A. McBryan. GENVL and *WWWW: Tools for Taming the Web. First International Conference on the World Wide Web*. CERN, Geneva (Switzerland), May 25-26-27 1994. http://www.cs.colorado.edu/home/mcbryan/mypapers/www94.ps

[3]Sunny lam," The Overview of Web Search Engines", February 2001

[4] J Pei, "Information retrieval and web search architecture", Lecture slides, 2017

[5] R. Baeza-Yates and B. Ribeiro-Neto. Modern Information Retrieval. Addison Wesley, New York, NY, USA, 1999

[6] Comscore report, February 2016