A Review Paper on Various Search Engines (Google, Yahoo, Altavista, Ask and Bing)

Manika Dutta
Research Scholar
Department of Computer Science
Himachal Pradesh University, Shimla.
(14manika@gmail.com)

Dr. K. L. Bansal
Professor
Department of Computer Science
Himachal Pradesh University Shimla.
(kishorilalbansal@yahoo.co.in)

Abstract:- Search engines are used in the web as a tool for information retrieval. As the web is a huge repository of heterogeneous and unstructured data so to filter out relevant information from unnecessary ones search engines are needed. Search engines usually consists of crawling module, page repository, indexing module, querying module and ranking module. The inter communication between these modules describes the working methodology of a search engine. This paper aims to focus on the comparative analysis of five major search engines i.e Google, Yahoo, Altavista, Ask and Bing in a tabular form based on some features. The features include search operator, search web, search images, search videos, search news, search maps, search books, advance search, change background, change search settings, display number of results, shopping, translation services, multi-language support, questions/answers, directory, advertising programs, business services, themes, case sensitive, finance, safe search, search pad, career and preferences. Google stands out as the best search engine amongst all search engines, which works on Page Rank algorithm. Page Rank is a numeric value which determines the importance of a web page by calculating the number of backlinks.

Keywords- *PageRank, Search engine.*

*****

I. Introduction

Web comprises of numerous 8 billion heterogeneous pages having unstructured data .Hence for a user to retrieve information from the web a special site is needed. Internet search engines act as the tool using which people find specific information from the Web[1]. Hence a Web search engine is an information retrieval system which is used to locate the web pages relevant to user queries. The results are being presented in the form of lists called hits in order of relevancy and one being shown at top having high priority and at bottom having low priority [2]. This information retrieval from the web has two classical performance evaluation metrics:

Precision the fraction of items retrieved by the system that are interesting to the user and recall the fraction of items of interest to the user that are retrieved by the system.

The goal of all search engines is to find and organize distributed data found on the Internet. Before search engines were developed, the Internet was a collection of File Transfer Protocol (FTP) sites in which users would navigate to find specific shared files. As the central list of web servers joining the Internet grew, and the World Wide Web became the interface of choice for accessing the Internet, the need for finding and organizing the distributed data files on FTP web servers grew. Search engines began due to this need to more easily navigate the web servers and files on the Internet[3].

II. History of Search Engine

The first search engine "Archie" was developed in 1990 by Alan Emtage, a McGill University student in Montreal. It was a collection of computer files stored on unknown FTP web sites on a computer network[3]. In 1991, Mark McCahill, a student of University of Minnesota, effectively created Gopher using hypertext paradigm which also searched for plain text references in files.

Both of these search engines did not have natural language keyword capabilities that are used in modern search engines. Hence in 1993 improving upon Gopher's text-based interface graphical Mosaic web browser was invented . Almost at the same time, Matthew Gray developed Wandex, the first search engine in the form of modern day search engines. Wandex's technology was the first to use crawling of web, indexing and searching the catalogue of indexed pages on the web. Another magnificent development in search engines came in 1994 when WebCrawler's search engine began indexing the full text of web sites instead of just web page titles. From thereafter numerous search engines were cultivated which rapidly made their mark in the market of search engines. Some of these search engines still holds an active status in the present scenario. All these search engines are briefly named as follows:

- **1990- Archie**
- **1991- Gopher,Veronica and Jughead**
- **1992- Vlib**
- **1993- Excite**
- **1994- Yahoo!, WebCrawler,Lycos,InfoSeek**
- **1995- Altavista**
- **1996- Inktomi**

- **1997-AskJeeves,Google**
- **1998- MSN Search**
- **1999- AllTheWeb**
- **2003- Overture**
- **2006-Snap, Wikia**
- **2008- Cuil, DuckDuckgo**
- **2009- Bing**
- **2010- Blekko**
- **2011-Siri**
- **2014- Cortana**

III. Features of Web based search engine:

1. **Web Indexes:** It is generated by the Web robots every time a Web search request is generated. Hence the composition of a web indexes affects the performance of the web search engine. Usually web indexes has three components namely coverage, update frequency and the positions of web pages indexed. The magnitude of these three components largely depends on the power and the hardware plus the software from which the web index has been made.
2. **Search Capability:** A Quality web search engine provides the fundamental search facilities i.e Phrase searching(a type of search using which user can search for documents containing an exact sentence or phrase) , truncation(limiting the number of digits right of the decimal point) and limiting facilities.
3. **Retrieval performance:** It states the procedure how the information is retrieved and processed in a web information system. It is evaluated on three parameters: precision, recall and response time which are quantitatively measured.
4. **Output option:** It deals with two things: one is the number of output options a web search engine offer and the other is the actual content of the output. The output content provided by a search engine usually depends upon the architectural composition of the search engine. Hence the architectural composition of a search engine impacts hugely the content quality.
5. **User effort:** It refers to the documentation and interface implies that users will only use a search engine when they will be comfortable with its interface and able to read and comprehend its documentation when consulted. Therefore, it becomes mandatory to design a user friendly interface for a search engine administrator to maintain its high value among the users in the long run.

IV. Types of Search engines:

Generally search engines are found to be of the following types[4]:

1. **Crawler based search engine**: These are the search engines that use software programs called spiders/crawlers/robots/bots .After accessing the web pages they are being categorized and analyzed and then added to the search database, whereby user can find them later on. .These search engines are constantly updated with new webpage that would be available in their database. **Examples: Google, Yahoo,AllTheWeb and Altavista.**
2. **Directories**: Directories are websites sanctioned in specific categories by human editors which place the search results in the "Directories" database. These provides more accurate and relevant search results than crawler based search engine. **Examples: Yahoo directory, Open directory and Look smart.**
3. **Hybrid search engine**: These are the search engines use both crawler based search and directory searches to obtain their result, hence employs the best features of both these search engines. **Examples: Yahoo.com,Google.com.**
4. **Speciality search engines**: They help the user search in different search areas in specific types of lists that are specialized in specific topic that would be beneficial for narrowing down the search process.**Examples: Askjeeves(** question and answer search engine),**Medhunt(** provides only medical information).
5. **Meta- search engines** : These are the search engines that combines all results from other search engines into one big list. Search results obtained can be integrated , duplicates can be eliminated and clustering by subjects within the search results can be implemented by meta search engines. Saves time as search only in one place..**Examples: Dogpile , Metacrawler.**

V. Search engine system architecture

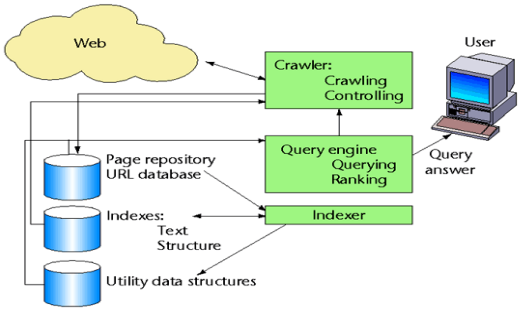


Figure 1 Architecture of search engine

Figure 1 shows the architecture of search engines which comprises of the following elements [4]. All these components are explained in detail as below.

1. Crawler
2. Page Repository
3. Indexing module
4. Querying module
5. Ranking module

The search engine works as follows using these five modules. Crawler module scans the massive and unstructured world wide web and sends the multiple spiders to fetch web pages. In return the web sends back the data in the form of web pages to the crawling module. Web pages may be presented in the form of URL or hyperlinks after they got through processing by spiders[5]. Page repository temporarily stores all the fetched web pages and later on provides them to the indexing module.

Indexing module strips off the contents from these web pages. Usually the extracting key elements here are title tag, description tag and internal links using the search engine optimization techniques. Hence the summaries of these web pages are being created which are further broken down to chunks of indexes in the form of content, video and image. The above three modules described and the interactions between them function independent of the query. The execution query part depends upon the last two modules plus the search engine. As a result, the indexes from indexing module are then sent to the querying module. When the user types the query in the search box of a search engine for example, Google, it is immediately sent to the querying module for further action.

The querying module breaks it into a language which the specific search engine understands. After that the module shows the vast amount of results in the form of web pages. In the end querying module sends further these results to the ranking module which ranks the web pages using page popularity and their rank , thereby reducing the result set and providing them according to user needs. Finally ranking modules sends the desired result set to the user as per requested[6].

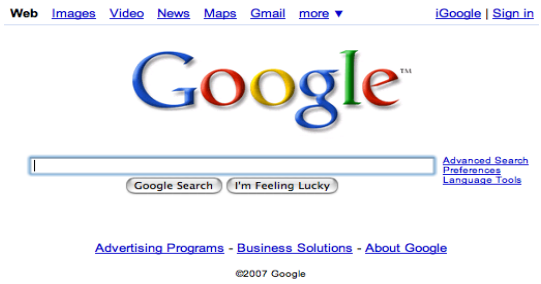
VI. Search Engines

1. Google: It Runs on a unique combination of hardware and software which began in January 1996 as a research project. Larry page and Sergey Brin laid its foundation at Stanford university in 1997[7]. The purpose of inventing Google's is to organize the haphazard information present on the web in a logical manner and make it globally accessible and useful . The life span of a Google query normally lasts less than half a second, yet involves a lot of different steps that gets completed before displaying the results to the end user. The heart of Google is PageRank algorithm for ranking web pages[8]. The architecture of Google constitutes of the following components[9]:

(a) Googlebot a web crawler that finds and fetches web pages.

(b) Indexer which sorts every word on every page and stores the resulting index of words in a huge database.

(c) Query processor which compares search query to the index and recommends the most relevant documents.



Following are the unique features provided by Google search engine:

- ✓ **Time taken:** It tells the time taken by the user for searching information on the web.
- ✓ **Books:** Searching and downloading number of books belonging to different categories.
- ✓ **Calendar:** Way to organize daily, weekly and monthly schedule and share events with friends.
- ✓ **Blogs:** Where a user discusses his views and ideas regarding a particular topic.
- ✓ **Scholar:** Search engine which allows academic researchers globally to search for research papers and articles published in various journals .
- ✓ **Photos:** User share number of photos with friends.
- ✓ **Discussion:** Create online documents and share them through various social medias.
- ✓ **Groups:** Create discussion groups to share information with friends having common interests.
- ✓ **Recipes:** Finding best recipes from top websites and best cookery shows around the world.
- ✓ **Google Chrome:** A faster, safer and easier browser built for users.
- ✓ **Picasa:** For finding, editing and sharing your pictures from PC.
- ✓ **Alerts:** Send email to the user based on their previous search.
- ✓ **iGoogle:** Personalized web page that enables the user to add links, photos, audio,video,letters etc.
- ✓ **Google Earth:** Allows a person to freely go to any device or location and view almost anything on earth.
- ✓ **I'm feeling lucky:** Provides the user with the first web page for the search query.
- ✓ **Realtime:** Provides recent current updates about burning topics from around the world.

2. Yahoo: David Filo and Jerry Yang two electrical engineering graduates of Stanford University created a website and named it as " David and Jerry's guide to the World Wide Web", at California in Jan. 1994. Later on, in April 1994, " David and Jerry's Guide to World Wide Web", was renamed as "Yahoo!"which stands for " Yet Another Hierarchical officious Oracle". Hence the website

called "yahoo.com" was created on January 18,1995[10]. The term "hierarchical" describes how the Yahoo! database is arranged in layers of categories and subcategories. The term "oracle" means "source of truth and wisdom", and the term "officious", unlikely describes the many office workers who would use the Yahoo! database while surfing from work.Yahoo provides numerous web communication services such as Yahoo! Messenger and Yahoo! Mail, Yahoo! Maps, Yahoo Finance and Yahoo! Groups. Apart from this Yahoo also offers social networking services such as My Web, Yahoo Personals and Flicker[11]. Its page could be easily accessed , is clear and has user friendly links. Yahoo's content main focus is on financial news. Unique features provided by Yahoo! are as follows:

- ✓ **Sports**: Provides news related to the latest happening in the sports world.
- ✓ **Local**: Yahoo! helps to find local news of multiple business listings.

3. Altavista: AltaVista is an early popular search engine invented in the 1995.Apart from doing full-text searches ,it also searches graphic images and discloses the linked structure between different Web pages. AltaVista's search robot is known as Scotter, which looks at and collect data from million Web pages every single day. Ni2,the indexer of Altavista indexes approximately one gigabyte of data per hour. It lost ground to Google and was purchased by Yahoo ! in 2003. Consequently now the domain of this search engine works under the guidance of Yahoo!. Few unique features of AltaVista includes

- ✓ **Travel**: To arrange trip plans and reserves hotels and resorts.
- ✓ **Yellow page**: Directory listing business names, addresses, mobile numbers etc.
- ✓ **People finder**: People can search for person's address and phone number.
- ✓ **Text only search**: Search the web with a faster graphics free vision.
- ✓ **Web master search**: Find site's pages in the AltaVista index.

4. Ask: Ask.com which is originally called as Ask Jeeves, is a web search engine which works on the principle of question answer.It was invented in 1996 by Garrett Gruener and David Warthen in Berkeley,California. Ask.com had received criticism for its browser toolbar, for behaving like malware due as it gets bundled with other software, hence becoming utterly difficult to uninstall .The purpose behind Ask Jeeves was to provide answers to the end users to get answers regarding daily life events in natural language, as well as by traditional keyword searching. The current Ask.com still supports traditional search facilities with maths, dictionary, and conversion questions. Unique feature provided by Ask.com is:

- ✓ **Encylopedia**: A type of unique reference work that contains articles on various topics ranging from arts to science.

5. Bing: Bing is one of the most recent search engine prevailing in the market of search engines. It is operated by Microsoft who launched it on May 28, 2009. As the search results obtained always fulfills the user needs which are required from the query. Hence Microsoft calls it a "Decision Engine,". When searching on Bing, apart from providing relevant search results, the search engine also showcases a list of related search items on the left-hand side of the search engine results page (SERP).Access to see recent search history is also provided by means of a quick link. Bing uses technology from a company called Powerset, which now Microsoft has acquired. It provides a variety of search services such as web, video, image and map search products. It uses the ASP.NET programming language and follows the "Metro" design language of Microsoft. It's unique feature is:

- ✓ **Community**: Share and learn resources with other people who use Bing.

All these five search engines Google, Yahoo ,Altavista, Ask and Bing provides some basic features which are present on the home page of each of them. Table 1.1 shows provides the tabular comparison of these five search engines based on the various features such as search operator, search web, search images, search videos, search news, search maps, search books, advance search, change background, change search settings, display number of results, shopping, translation services, multi-language support, questions/ answers, directory, advertising programs, business services, themes, case sensitive, finance, safe search, search pad, career and preferences.

Table 1 Comparative Analysis of Features present in Five Search Engines.

Features	Google	Yahoo	Altavista	Ask	Bing
Website	www.google.com	www.yahoo.com	www.altavista.com	www.ask.com	www.bing.com
Search operator	AND,OR,NOT	AND,OR	AND,OR,NOT	AND,OR	AND,OR,NOT
Search Web	Yes	Yes	Yes	Yes	Yes
Search Images	Yes	Yes	Yes	Yes	Yes
Search Videos	Yes	Yes	No	Yes	Yes
Search News	Yes	Yes	Yes	Yes	Yes
Search Maps	Yes	No	No	No	Yes

Search Books	Yes	No	No	No	No
Advance Search	Yes	Yes	Yes	Yes	Yes
Change Background	Yes	No	No	Yes	Yes
Change Search Settings	Yes	Yes	Yes	Yes	Yes
Display no. Of results	Yes	Yes	Yes	No	Yes
Shopping	No	Yes	Yes	No	
Translation Services	Yes	No	Yes	No	Yes
Multi-language Support	Yes	No	No	No	No
Questions/Answers	No	Yes	Yes	Yes	No
Directory	Yes	Yes	Yes	No	No
Advertising Programs	Yes	Yes		No	No
Business Solution/Services	Yes	No	Yes	No	No
Themes	No	No	Yes	No	No
Case Sensitive	No	No	No	No	No
Finance	Yes	Yes	No	No	No
Safe Search	Yes	Yes	Yes	Yes	Yes
Search Pad	No	Yes	Yes	No	No
Careers	No	Yes	No	Yes	No
Preferences	Yes	Yes	Yes	Yes	Yes

Table 1 shows the feature wise comparison of search engines namely Google, Yahoo, Altavista , Ask and Bing. The first feature used is search operator which means the operators which are used internally by the search engine for retrieving information. While Google, Altavista and Bing uses the operators AND, OR, NOT, Yahoo and Ask uses the operators AND, OR. Search web, images, news features search for information, images and news from the web respectively. All the five search engines Google,Yahoo, Altavista, Ask and Bing searches the web, images and news. Except Altavista all the other four search engines searches the videos. Search maps enables the users to search for direction from one location to another. Only Google and bing provides the facility for searching maps. Search books feature allows the users to preview millions of books from libraries and publishers worldwide. Except Google other four search engines doesn't have the facility to search for books. Advance search allows the user with advance option to write specific query and obtain precise results. Google, Yahoo, Altavista, Ask and Bing all allow the user to go for advance search. By using the feature of the change background user can customize the web page, hence changing the back ground settings as per convenience. Google, Ask and Bing changes the background while Yahoo and Altavista doesn't. Using change search setting user can change the search setting as per requirement. Google, Yahoo, Altavista, Ask and Bing all these search engines changes the search settings. Display number of results show cases the number of results fetched by using the search engine. Except Ask all the other four search engines displays the number of results. The feature Shopping provides the user facility of buying online products by accessing various e- commerce websites. Only through Yahoo and Altavista a user can do shopping. Translation services allows the users to translate text and web pages from one language to another language. Except Yahoo and Ask all the other three search engines translate the present text to multiple languages. The feature present multi-language support ensures whether the search engine supports multiple languages or not. Here only Google provides the multi-language support, rest do not. Answers provides good facility where people ask and answer questions on any topic of their choice and can share facts, opinions and personal experiences within a community. Except Google and Bing , Yahoo , Altavista and Ask provides the answers to users questions. Directory enables the facility to search the web, organized by topic or category. Except Ask and Bing a user can search for information topic wise on Google, Yahoo and Altavista. Advertising programs provides the facility to search engines for promoting and helping user's business. Only Google and Yahoo helps promote the user business while Altavista, Ask and Bing do not. Business services provides the user facility to promote and help user's business. Except Google and Altavista the other three search engines do not provide the business services. With themes, users can change the theme according to own choice. Except Altavista, with the other four search engines user cann't change themes. Case sensitive checks whether the search engine presently at use is case sensitive or not. Google, Yahoo, Altavista, Ask and Bing are not case sensitive. Finance gives the user information regarding stock market. Only Google and Yahoo gives the information regarding stock market while Altavista, Ask

and Bing do not. Safe search allows the user to filter out explicit adult- oriented content from results. Google, Yahoo, Altavista, Ask and Bing all these search engines efface out the adult oriented results from the search results. Search pad keeps track of the websites one chooses from the search results and to make notes on them. Except Yahoo and Altavista , Google, Ask and Bing do not keep track of the websites a user visit. With careers feature user can browse various jobs according to his choice. Yahoo and Ask allows the user to search for jobs while Google, Altavista and Bing do not. Preferences allows the user to search for the required information according to priority. Google, Yahoo, Altavista, Ask and Bing all allows the user to search for information according to the priority of the user. Hence from the table 1 it can be concluded that Google emerges as the best search engine among the other four search engines. Also Yahoo and Bing secures the 2nd and 3rd positions in terms of providing search features to the users. Google, the best search engine works on the Page Rank algorithm which decides the importance of a web page by counting the number of votes cast to a particular web page. Hence this Page Rank algorithm assumes that a web page is important if it is being pointed to by another web pages by means of backlinks.

VII. Conclusion and Future scope

Today web search engines are used for retrieving relevant query based information within few seconds by the user. Search engines have the modules for crawling, indexing, ranking, querying and a page repository for temporary storage of web pages. This paper has attempted to provide a review of various search engines namely Google, Yahoo, Altavista ,Ask and Bing. Also the tabular comparison of these search engines on various features has been given. The features used are search operator, search web, search images, search videos, search news, search maps, search books, advance search, change background, change search settings, display no. Of results, shopping, translation services, multi-language support, answers, directory, advertising programs, business services, themes, case sensitive, finance, safe search , search pad, careers and preferences. Consequently Google emerges as the best search engine. The heart of Google is PageRank algorithm. Page Rank is Google's way of deciding a page's importance. Google calculates a page's importance from the votes cast for it. Importance of each vote is taken into account when a page's Page Rank is calculated. More vote means more importance. Hence Page Rank is a numeric value that represents the importance of a page present on the web. In future Page Rank can be calculated in terms of link distance and a modified procedure can be given for calculating the PageRank.

Refrences

[1] Kuyoro Shade O, Okolie Samuel O, Kame Richmond U and Awodile Olodele, " *Trends in Web based Search Engine" ,* Journal of emerging trends in Computing and Information Sciences, Vol. 3, No. 6, June 2012.

[2] Trilok Gupta and Archana Sharma , " *Anatomy of Web search engines" ,*International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 4, April 2014.

[3] The…/Short_History_ of _Early_Search_Engines.aspx.

[4] Sayush Gupta , Krishan Dev Mishra and Prerna Singh, " *Effective Searching Policies for Web Crawler" ,* International Journal of Modern Engineering Research , Vol. 3, Issue 5, Sept-Oct 2013.

[5] R. Rubini , Dr. R. Manicka Chezina ," *An Analysis on Search Engines: Techniques and Tools",* International Journal of Advanced Research in Computer and Communication engineering, Vol 3, Issue 9, Sep 2014.

[6] Monica Peshave , Kamyar Dezhgosla, " How Search engine work and a Web Crawler Application".

[7] Kaushal Kumar, Abhaya, Fungayi Donewell Mukoko, " *PageRank Algorithm and its variations : A Survey Report",* IOSR Journal of Computer Engineering, Vol 14, Issue 1, Sept-Oct 2013.

[8] M. Vengateshwaran and E.V.R.M Kalaimani, " *Web Mining Research Direction and Open Source Tools" ,* International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, Issue 7, July 2014.

[9] Krishan Kant Lavania, Sapna Jain, Madhur Kumar Gupta and Nicy Sharma, " *Google: A Case Study(Web Searching and Crawling),* International Journal of Computer Theory and Engineering, Vol. 5, No. 2, April 2013.

[10] Sushma Malik, "A Comparative study of two major Search Engine: Google and Yahoo" , Oriental Journal of Computer Science and Technology.

[11] Pratibha Sharma, Sonal Shukla and Brahm Dutt Bohra, " *A Comparative Analysis of Various Search Engines Present Today" ,* International Journal of Computer Science and Engineering, Vol. 4, Issue 4, June-July 2015.