

Web Search Tools: Google as best player

Aviral Kumar Singhal¹, Ishani² and Niraj Singhal³

¹Student, B.Tech. (I.T.), Meerut Institute of Engineering & Technology, Meerut

²Student, M.Sc. (Mathematics with Computer Science), Jamia Millia Islamia, Delhi

³Professor, Shobhit University, Meerut

Corresponding author: aviralkumarsinghal@gmail.com

Abstract: Within the period of last few years the Internet has changed the way of doing our business or communication. The World Wide Web (or Web) is a huge source of documents available in hyperlinked form. The huge Web is poorly structured and useful information is placed with much poor quality information. So, it is not an easy task to find the right information at right time. Search engines are the tools that help users to search relevant information from the web. This paper discusses need of search engines and their components & functioning. It presents about various search engines available and their share in International market. It also presents various features about Google that makes it first choice for the web users.

Keywords: Web, Search engines, Web crawler, Index, Query Engine, Google

1. Introduction

Today, Internet [1] has become an integral part of our life. It is estimated that out of 7.2 billion population of the world, 4.2 billion uses Internet. There is an increase of 3.6 billion to 4.2 billion with growth rate of 1052% in last 18 years (2000-2018). In Latin America it is 2318% and in Asia it is 1670%. In Asia, out of 4.2 billion people 2 billion (48.7%) use Internet whereas in rest of the world 2.18 billion (51.3%) uses Internet out of 3.4 billion. Even in India .46 billion people uses Internet out of 1.35 billion. Same or higher growth is expected in future also [2].

In the last few years, the size of Web [3] has increased from few pages to billions of pages. Study reports that size of visible web is approximately 55 billion pages [4]. Today, web has become a universal source of information and but due to its huge size finding right information at the right time is a challenging task. So, it needs effective tools to search relevant information.

2. Web Search Engines

Search engines [5, 6] function as a link between the web and web documents. Without the web search engines, huge source of the information on internet, stored in web documents remain hidden for us all. A web search engine gathers information from the web documents available on the Internet, indexes the collected information and thereafter the result is stored in a huge database collection where from it can be quickly searched [7].

Search Engine may be referred as a huge database of Internet resources such as hyperlinked web pages, other web documents in different forms, newsgroups, programs, images etc. A search engine helps, to trace information on Web. A user can search for any information that he wants, using a search engine, by passing a query in the form of query or phrase (using keywords). It, then searches for relevant information in the database and returns result to the user.

2.1 Basic Components of a Web Search Engine

Three basic components of any general search engine are web crawler, database and a query engine [8]. Web search engine’s general architecture is given below in Figure 1.

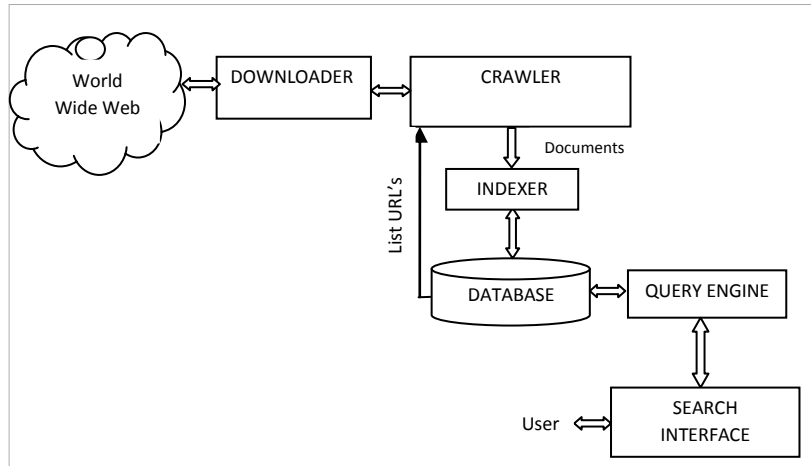


Figure 1. Architecture of a web search engine

So, Web crawler, database and the query engine are main three basic components that actually makes search engine to work. Web crawler is a program which collects the useful information from the entire web in an automated manner. The objective of web crawler is to maintain freshness of the pages in its collection as high as possible. All the information collected from the web is stored in large database. The purpose of Query Engine is to receive and get filled search requests from users. Because of the large size of the web and the fact that users put only few keywords for anything to search, the result sets are usually very large.

Web Search engines use Boolean expressions (AND, OR, NOT) to restrict and widen the results of a search. The steps performed by any web search engine are as follows:-

- i. A web search engine looks for the keywords in its index (stored in large databases) instead of going directly to the web to search for the keywords. It then uses specialized algorithms to search for the relevant information in its database.
- ii. On finding the related pages, then web search engine returns the relevant web pages as a result.
- iii. These retrieved web pages generally include various information regarding a web page like page title, size of text portion, first few sentences etc.
- iv. Search criteria usually differs from one web search engine to another. The information retrieved is further ranked (using specific algorithms) considering various factors such as keywords frequency, relevancy of information, links etc.
- v. To open search results, a user can click on any of it.

2.2 Search Engine Architecture

Three basic layers of a web search engine architecture are Content collection and refinement, Search core and User & application interfaces. Layered architecture of a web search engine is shown in Figure 2.

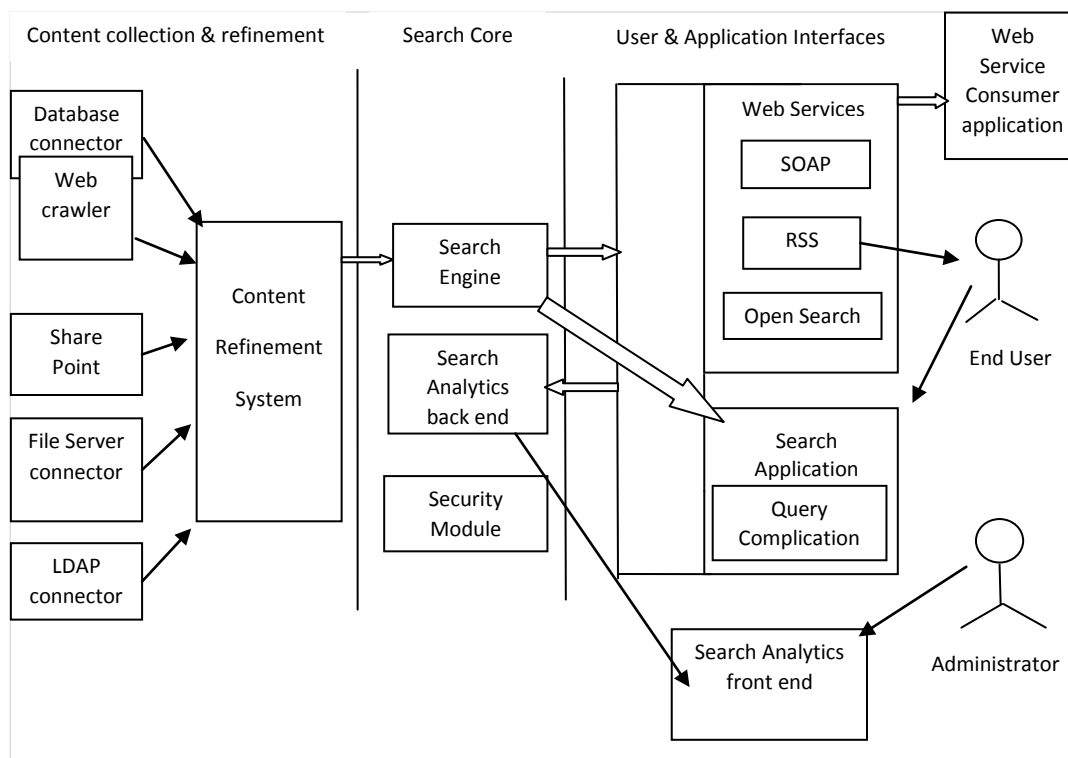


Figure 2. Layered Architecture of a Search Engine

3.0 Search Tools Available

When any individual thinks of Search Engines, the first name that comes to his mind is normally Google [9]. It is one of the most trusted brand name in the world, and it has even worked on its way into main stream. Today the people start substituting the phrase “searched online” for “Googled”.

However, Google is the prominent one but it is not the only Search Engine available in market. There are several other search engines that provide different interfaces, algorithms to search and other unique features. Many of them base their search algorithms that often attract new users.

Several search engines available today are as follows:-

Search Engine	Description
Google	Original name BackRub. It is the most popular search engine used today globally.
Bing	Launched by Microsoft in 2009. It is one of the latest web-based search engines that also delivers Yahoo’s results.
Ask	It was launched in 1996, originally known as Ask jeeves. It includes support for dictionary, match, conversation question etc.
AltaVista	It was launched by Digital Equipment Corporation (DEC) in 1995. Since 2002, AltaVista is powered by Yahoo technology.
AOL. Search	It is powered by Google.
LYCOS	According to media Matrix it is 13 th largest online property, and in top five internet portals.
Yahoo Search	Second largest Search Engine on the web. It is powered by Bing since 2011.
Alexa	A subsidiary of amazon, and used to provide website traffic information.

Market share of various popular search tools is shown in Figure 3.

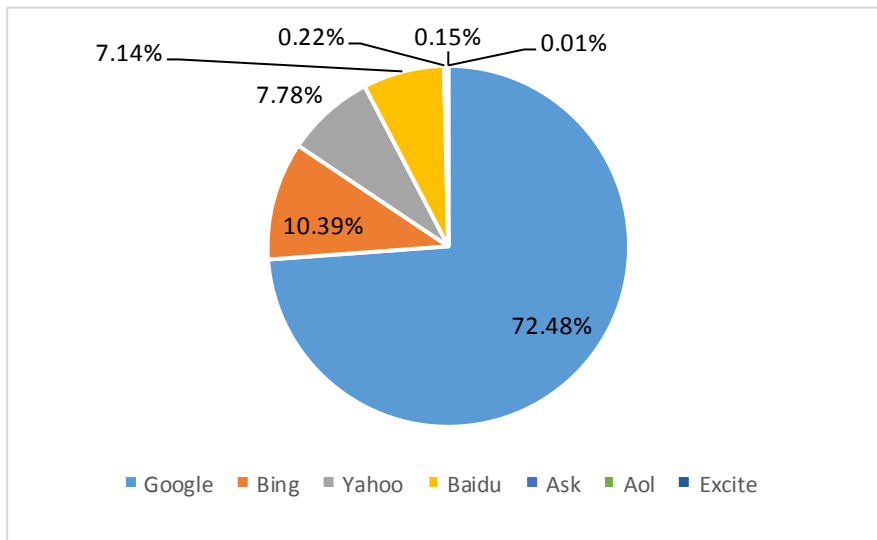


Figure 3. Market share of various popular Search Engines

3.1 Google is different

Google Search Engine Result Page (SERP) changes regularly.

That means different search results for the same query are seen by different people. Various reasons that show that why Google is unlike other players are [10]:-

i. Personalized results

Google personalizes the search based on user’s history and behavior also, to give him results that are most likely fit for them. The data includes:-

- **Search history** -Which is based on previous searches using Google.
- **Visited websites** -History from the web browser’s history.
- **Whether you’re logged or not to Google account** –When a user logged in Google account, it lets Google to watch and learn what websites he surfs, time he spends on and how he interact with them.
- **Engagement** – If a user buys a product form a particular shopping website, shares anything on social media, comment a blog post or play a video on YouTube, i.e. his each activity is used by Google to come up with a personalized SERP.

ii. Location/IP address

User’s current location also leads to personalized search. Google uses user’s IP address to identify his location. One can disable this option for not to use the location by web browser? But it won’t resolve the issue.

iii. Device used

The mobile SERP consists of mobile-friendly websites if applicable. The usage of Smartphones has brought local Search Engine Optimization (SEO) to a new level of importance.

iv. Google data centers synchronization

For quick delivery of results, Google has come up with several data centers across the globe. Data among these data centers are need to be synchronized regularly with significant delays. Based on the location of which is the

search query submitted, Google search engine fetches the data from the closest data center. Which means that sometimes when the data may have not been synchronized, the results for same keywords may be different.

v. Search settings

We all do various searches in Google i.e. websites, images, maps, videos, and couple of others. Search settings have visible impact on the results which are not reflected in keyword ranking tools. Google results may differ for the same query on the search type.

vi. Time

Update of websites is also an important factor. SERP updates the rank and checks by rank trackers daily. It is also necessary to compare results during matching time period. Because of other mentioned factors one may not probably get the same rankings. Also, SERP can change multiple times a day for extremely time-sensitive keywords (breaking news, global affairs, etc.)

vii. Other factors

Some other factors causing the differences are as follows:-

- **Typos**-A small mistake in the text made when it is typed.
- **Writing in English (US/UK) (or any other languages)** – Google automatically corrects user's search queries in almost all languages but it can happen, a user may add keywords with mistakes too.
- **AdWords change the appearance of SERP**- These don't have anything to do with results and are situated on the top (or bottom) of the SERP.

There are a number of purposes [11] that Google is able to lead and excel over all the other competitor search engines. Google is better than several others in many aspects. If Google had been better in just some of these areas, for Google it would have been very difficult to surpass them, but being best of all made it an easy choice for users to switch and to use it.

- i. **Speed:** Google is very fast than its competitors (almost ten times) as far as returning search results is considered. End users had never imagined such a fast rate at which results are returned.
- ii. **Much deeper index:** Google crawls a large number of webpages than its competitors for its database collection that leads to a richer set of results. Google has shown that they are bigger and better, on top of the search results to convince users.
- iii. **Relevance of results:** The results returned by Google are more relevant than others. The PageRank algorithm used by Google is an important part of this. There are many other important factors that went into their ranking algorithm as well.
- iv. **Simplicity of interface:** Many other search engines have very messy interfaces with advertisements and other distractions. On the other hand, Google provides a minimalistic and clean interface where only the thing that the user focused on is web search.
- v. **Query-specific snippets:** Google has been the first web search engine that displays snippets on results that show a user what part of the result matched the query. This is very useful for a user in figuring out which results are relevant and which are not by looking at the search results page.
- vi. **Single-minded focus on search and continued interaction on it over time:** This is one of the most important reasons in the list. Since beginning, the main product of Google has been web search. It was its only product for the first few years of Google's business. This focused view of Google on a single area is like a wonder for many companies when it comes to product development. The day after it became the dominant search engine for the web, Google didn't wait and rest on its successes and slow down product development on search. Even, Google doubled its efforts and invested heavily in research for further improving.

4. Conclusion

This paper has discussed about usage of Internet and its worldwide growth. It also presents world wide web and ways for finding information from it. The design of a search engine along with its components, are also presented. It is concluded that crawling the complete web is not an easy task. The crawlers have to be tricky enough and reliable to perform the crawling process efficiently and reliably. We also discussed that a crawler has to deal with billions of web pages. It has presented various search tools in use today and their presence at global level. It has also discussed several features of Google that makes it a favorite search tool for web users.

References

- [1]. Barry M. Leiner, Vinton G. Cerf, David D. Clark, Robert E. Kahn, Leonard, Kleinrock, Daniel C. Lynch, Jon Postel Larry G. Roberts, Stephen Wolff, "A Brief History of the Internet", www.isoc.org/internet/history. (Accessed on 19 August 2018)
- [2]. www.internetworldstats.com (Accessed on 19 August 2018).
- [3]. Alexandros Ntoulas, Junghoo Cho, Christopher Olston, "What's New on the Web? The Evolution of the Web from a Search Engine Perspective", Proceeding of the World Wide Web Conference (WWW), May 2004.
- [4]. www.worldwidewebsite.com (Accessed on 19 August 2018).
- [5]. Franklin, Curt, "how internet Search Engines Work", www.howstuffworks.com. (Accessed on 19 August 2018).
- [6]. www.wikipedia.org/wiki/web-search-engine. (Accessed on 19 August 2018).
- [7]. Grossan. B, "Search Engines: What they are, how they work, and practical suggestions for getting the most out of them", Feb97, webreference.com.
- [8]. Vinit Kumar, Niraj Singhal, Ashutosh Dixit, A. K. Sharma, "A Novel Architecture of Preparation Oriented Web Search based on decision theory", Indian Joul of Science and Technology, Vol.8, No. 7, pp.635-641, Apr 2015.
- [9]. Sergey Brin, Lawrence Page, "The anatomy of a large-scale hyper textual Web search engine", Proceedings of the Seventh International World Wide Web Conference, pages 107-117, April 1998.
- [10]. <https://mangools.com/blog/reasons-why-google-search-results-are-different-from-rank-trackers/> (Accessed on 19 August 2018)
- [11]. <https://medium.com/@nikhilbd/how-did-google-surpass-all-the-other-search-engines-8a9fddc68631> (Accessed on 19 August 2018)