

A SURVEY PAPER ON EVOLUTION AND CLASSIFICATION OF SEARCH ENGINE

Koduri Srivyshnavi¹, .R.Lakshmi Priya², T.V Madhu Meghana¹, CheerlaPriyanka¹, Beera Vivek¹

1- Student, Department of CSE, Sreenidhi Institute of Science and Technology, Hyderabad

2- Associate Professor, Department of CSE, Sreenidhi Institute of Science and Technology, Hyderabad, India.

Abstract:

Search Engine is an Information Recovery System that was designed to help in the discovery of information. It is a project based on Hypertext. An information search engine can retrieve information from websites that were put online. The very first website was <http://info.cern.ch/> and was published on August 6, 1991 for the first time. It illuminated what the world wide web wa, and how to own a browser and how to set a web server active. Based on our search and ranking of those websites results were displayed. The ranking is done by the Search Engine Optimization technique (SEO). Different types of Search Engines are classified by various kinds of searching algorithms. The paper attempts to highlight the working of Search Engine, Classification of Search Engines, Types of Searches and a brief about Search Engine Optimization.

Keywords: Search Engine, Search Query, Search Engine Optimization techniques, DNS, Directories.

1. Introduction

Search Engine is a software where information from a wide database is found. It searches the web-based on the keywords and results “Search Engine Result Pages” (SERP), with a directory of sites[6]. Every site's goal is to appear first in the SERP ranking. The Keyword of the site’s standing is most significant since more the rank, more people will see it[8]. There are many categories of Search Engines that we will deliberate in the second section. Search Engine was developed in the 1990’s; with Archie Query Form and now we are experiencing the most advanced versions of them. A comprehensive overview ofthe Evolution of Search Engine will be discussed in the third section. In section 4, the working of SE is presented. Finally we will have an overview ofthe Search Engine Optimization Technique.

2. Evolution of Search Engines

Search Engines were developed in the 1990s. The very first Search Engine was “Archie”, it examined FTP sites to generate an index of downloadable files. FTP stands for File Transfer Protocol. Due to restricted space only links were shown and content is hidden [4]. In 1991, Tim Berners Lee put up a Virtual Library (VLib). CERN Webservers hosted a directory of web servers at an early age. In 1992, Veronica searched a list of filenames that were stored in Gopher index systems. In 1993, Jughead searched similarly as veronica but only a solo server at a instance. And in the same year Matthew Grey created World Wide Web Wanderer, a bot recorded active web server and also measured the internet growth as well. The database was called “Wandex”. In 1994, Infoseek, EInet Galaxy, AliWeb, Yahoo, WebCrawler, and Lycos were developed. Infoseek webmasters could submit a page in real-time. MartijnKoster created ALIWEB and it crawled metainfo. Yahoo Search was shaped by David Filo and Jerry Yang, as a collection of man-made description with each URL. This is available still now. They have also included commercial sites on their web. The webcrawler is the first one to index whole websites. Lycos used retrieval of Rank Relevance and matching of prefixes and word proximity. In 1995 LookSmart competed with Yahoo. Six Stanford undergrads created Excite. Altavista used unconstrained bandwidth for the first time with advanced searching techniques. In 1996, Larry and Sergay began functioning on Backrub, Google’s beginning that focused on backlinks. Inktomi, a Search Engine Hotbot listed on hotwire. AskJeeves launched a natural language search engine where human editors attempted to match search queries. In 1998, the MSN search relied on

Overture, Looksmart, and Inktomi before its backlink model was proved by Google. Now Google was launched. Then later Quen, Panda, Penguin, chema.org, Google Instant were developed later on. Coming to Google Search Engine History, it is unique and its tireless improvement in the algorithm made it the most popular search engine. Google showed up on the SE scene in 1996. It also instigated judging sites by authority, by calculating how many sites referred to it and how reliable they are.

3. Types of Search Engines

There are three categories of Search Engines[1]. These are classified based on their searching procedures. They are

- a) Crawler based Search Engines.
- b) Human Powered Directories.
- c) Hybrid Search Engines.

3.1 Crawler based Search Engines:

These entire Crawler oriented Search Engines employ a bot or spider for indexing and searching new content in a web database[5]. There are three fundamental steps this type of search engines follow. They are

- i. Crawling
- ii. Indexing
- iii. Calculating Relevancy

Crawling is the first step in these Search Engines. It uses a bot or spider for this task. It crawls all the websites sequentially as per the “URL” entered. The crawler digs through individual web pages and draws keywords from it and stores it to Search Engine’s database. Next Indexing is done to these sites. An index is used to discover the stored pages. In this process Search Engine analyzes page contents and the information is stored in Index. Coming to calculating relevancy it is ranking the sites as per their content in sites. Usually after we begin our search only top most relevant sites are displayed. This is done with the third step. Finally we display results in Search Engines. An ideal example of this is the World’s top most Search Engine “Google”. Crawler oriented Search Engines may be explained by the fig 1. shown below[1].

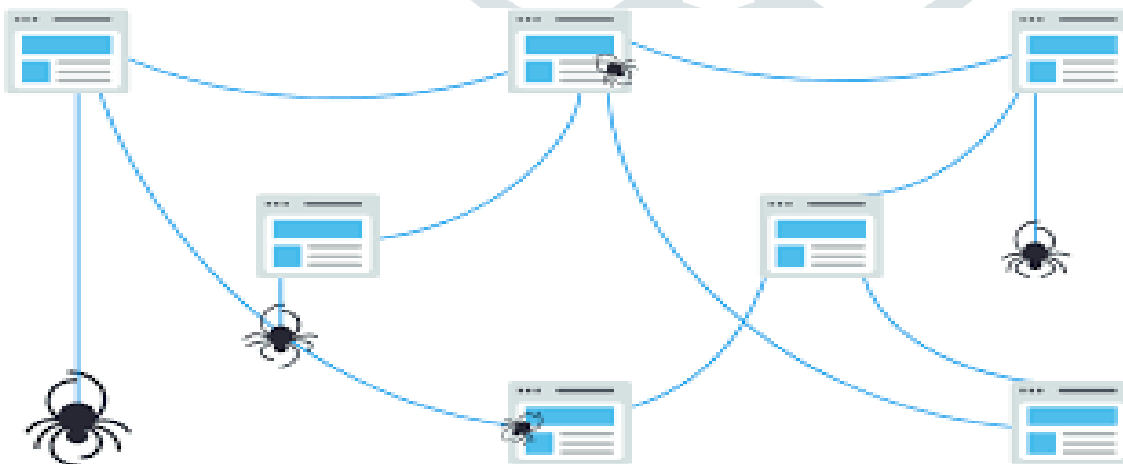


Fig 1- Crawler based Search Engine

3.2 Human-Powered Directories:

These are built by Human Selection. Here, Site Owner submits his site with a short description and submits along with category to be listed. These sites are manually revised and either added to the suitable list or excluded. The Keywords typed in the search box will be compared with the description of the site. Here only the site description is considered but not content. It depends on Human Editors in listings.

Yahoo was the best example of it. It follows the Human Powered Directory system in accomplishing its searching algorithm. The information present on the web page does not impact its ranking. It all depends on the User Description of the website. If it contains maximum keywords related to the search it is displayed as first in search results.

3.3 Hybrid Search Engines:

These Search Engines are a amalgamation of both Crawler based Search Engines and Human Powered Directories. Many Search Engines nowadays are moving towards Hybrid Search Engines.

Apart from these, Meta Search Engines can also be seen, which directs us to other sites and Semantic Search Engines, these perform search activities on a specific area are also found in the literature.

4. Working Procedure of Search Engine

Search Engine starts the search with parsing the URL we entered. It converts it into Machine Readable IP address. The process of converting URL we entered into IP address is handled by DNS which stands for Domain Name System. This can be more understandable by the following fig 2.



Fig. 2- Understanding Web

- Fig.2 clearly explains how Website, Web Hosting, Internet and Domain Name are related to each other. Websites were put online with the usage of Web Hosting. The Internet is where we find a lot of information with plenty of sites on it. Domain Name stands as a direction board that helps us in finding the website we require. Let us understand by searching “www.google.com”in Search Engine. Firstly, the DNS

system converts it into IP addresses. IP address changes because all the time we do not work on the same network.

- ▣ The IP address for our PC is a dynamic IP
- ▣ Google uses numerous servers to handle several inward requirements.
- ▣ **Google.com IP Address Ranges**
- ▣ Google uses the following public IP address ranges:
 - ▣ 64.233.160.0 – 64.233.191.255
 - ▣ 66.102.0.0 – 66.102.15.255
 - ▣ 66.249.64.0 – 66.249.95.255
 - ▣ 72.14.192.0 – 72.14.255.255
 - ▣ 74.125.0.0 – 74.125.255.255
 - ▣ 209.85.128.0 – 209.85.255.255
 - ▣ 216.239.32.0 – 216.239.63.255

```
C:\Users\Vivek>tracert www.google.com

Tracing route to www.google.com [172.217.167.164]
over a maximum of 30 hops:

  0  0 ms  0 ms  0 ms  192.168.1.1
  1  25 ms  2 ms  3 ms  dns1 [192.168.1.1]
  2   7 ms  4 ms  4 ms  192.168.100.1
  3  38 ms 1840 ms  98 ms  dns1 [40.50.61.1]
  4   4 ms  4 ms  4 ms  103.245.198.49
  5 1801 ms  43 ms  11 ms  103.15.63.185
  6  29 ms  30 ms  33 ms  103.241.47.241
  7  154 ms  91 ms  201 ms  103.241.47.130
  8  27 ms  26 ms  26 ms  108.170.248.161
  9  29 ms  28 ms  81 ms  108.170.232.203
 10  *      27 ms  27 ms  bom12s01-in-f4.1e100.net [172.217.167.164]

Trace complete.
```

Fig.3- Route Tracing in Command Prompt

If we trace the path of request we sent to Search Engine in command prompt it shows in fig.3. All these IP addresses include IP addresses of Local Service Providers, Regional Service Providers, and National

Service Providers and finally, get connected to the Google Server in California. The IP addresses can be tracked by Keycdn's IP Location Finder.

5. Search Queries

Search Queries are the vocabulary that generally public put into search boxes to get their results[7]. In general there are three kinds of search queries.

A. Navigational Search Queries:

It is a search inquiry that is typed by the people to find a specific website or webpage. Instead of entering the "URL" of the site user can directly enter the website name. "Facebook" and "YouTube" which are Navigational Search queries are the two top searches on Google.

B. Informational Search Queries:

It is a search question that has thousands of relevant results that cover a range of related topics. These are queries that are entered to learn how to do something or to find an answer to a question. These are not like navigational queries that are looking for a certain website.

C. Transactional Search Queries:

It is a search question that is typed by the user to make a transaction otherwise to purchase something. These search queries include terms like "buy", "purchase" or "order".

6. Search Engine Optimization

Search Engine Optimization is recognized as SEO is the procedure of increasing website visibility to users[2]. It makes our site to be listed in top search results for a certain keyword. 75% of searches perform searching in Google and 67% of Google's top listed sites are clicked. This tells how important is Search Engine Optimization is to a site. While classifying SEO, it is divided into Black Hat SEO and White Hat SEO. Black Hat SEO approaches include duplicating the content, imperceptible text with stuffed content, redirecting users to another site and links from sites with non-relevant information. Whereas White Hat SEO includes relevant content, links and references with well-labeled images. Black Hat SEO is like a rich quick scheme which only emphasizes on enhancing the content of Search Engine. White Hat SEO is a way to build a sustainable online business by directing on the human viewers. We also have Gray Hat SEO which is a hybrid of both White and Black Hat SEOs.

There are two wide groups of SEO; they are on-page SEO and off-page SEO. In On-page SEO ranking aspects are decided by headlines, contents and page structure. Coming to Off-page SEO ranking is based on external factors such as social media and blogs. To do well with SEO we must focus on on-page SEO and off-page SEO's as well. The three main groups of on-page SEO are content, quality and keywords. Content must be structured and should be qualified to match with keywords. HTML contents, architecture, tags, mobile-friendliness and page speed are also the factors that determine a page Ranking. Many algorithms evolved in the optimization area and there is a much wide possibility of research in this area[4].

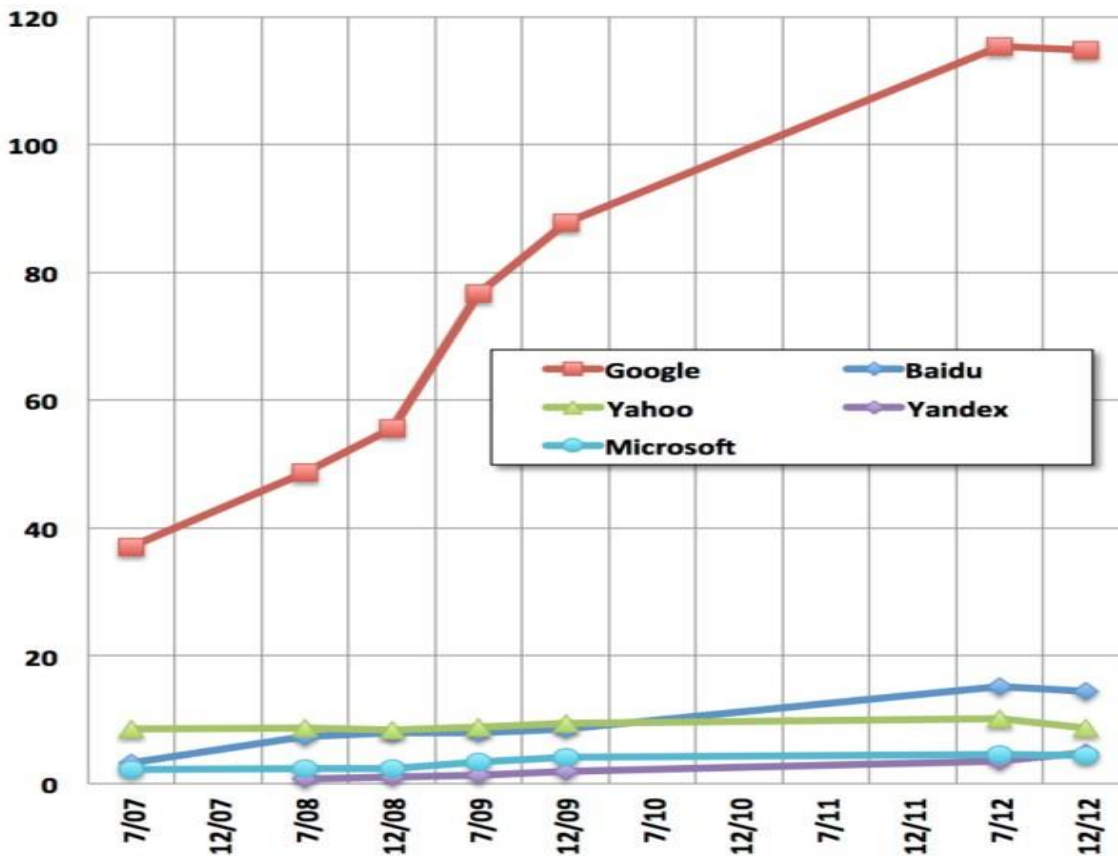


Fig 4- Usage of different Search Engines

7. Conclusion

The above figure explains the usage of Search Engines[3]; we can observe that Google is top among all the Search Engines. It stands top among all Search Engines because of its Search Engine optimization technique. In this paper Search Engines are detailed from their evolution to working along with classification. This makes readers familiar with the history of Search Engines and brings engrossment to research Search Engine Optimization Techniques.

References

- [1] <https://moz.com/beginners-guide-to-see/how-search-engine-operate>
- [2] Prashant Ankalkoti, “ Survey on search engine optimization tools and techniques”, Imperial Journal of Interdisciplinary Research, Vol.3, Issue 5, pp. 40-43.
- [3] <https://searchengineland.com/google-worlds-most-popular-search-engine-148089>
- [4] Patil Swati, Pawar B.V, Patil Ajay.S, “ Search Engine Optimization: A Study “, Research Journal of Computer and Information Technology Sciences, Vol.1(1), pp.10-13, 2013
- [5] https://en.wikipedia.org/wiki/web_search_engine
- [6] Ding,W & Marchionini, G, “ Comparative study of web search service performance:, Proceedings of 59th ASIS Annual Meeting, pp.136-142
- [7] Jansen, B,J, D.L,Booth and A. Spink, “Determining the Informational, Navigational and Transactional Intent of Web Queries”, Information Processing and Management, Vol.44, pp. 1251-66
- [8] Henzinger, M, “Search Technologies for the Internet”, Science, Vol.317(5837), pp.468-71.