



Search Factors used by Search Engines

Mu. Annalakshmi^{#1}, Dr. A. Padmapriya^{*2}

[#]Research Scholar, Department of Computer Science and Engineering,
Alagappa University, Karaikudi.

¹annalakshmi_mu@yahoo.co.in

^{*}Associate Professor,
Alagappa University, Karaikudi.

²mailtopadhu@yahoo.co.in

Abstract— Search Engines are of great help to users who need information from the large voluminous web. There are different search engines available globally. The basic operations of all the search engines are almost the same but they differ in the algorithms they use for searching. This brings about major differences in their result pages. This paper analyzes the search factors of some of the popular search engines.

Key words: search engines, search factors, crawling, indexing

I. INTRODUCTION

World Wide Web is a huge repository of information about various disciplines. So searching information from the web is a challenging task. A web search differs from normal information search of a text document due to the following factors[1].

1. Vast size of the web
2. Diverse nature of the web since it consists of text, images, video, audio and other multimedia content.
3. Exponential increase in size of web due to addition of new information.
4. Dynamic nature of web since the most of the web content is modified frequently.
5. Duplicate content in the web.
6. Hyperlinks to other web documents.

Goals of web search:

- i) Navigational : The goal is to find a website that the user has in mind but the URL is not known.
- ii) Informational : The need is to find information about some specified topic.

iii) Transactional : The important goal here is to perform some sort of transaction such as buying a product or downloading some information.

II. SEARCH ENGINE

Search engine[2] is a tool that helps the web users to search the required information from the voluminous data available in the web. In other words Search engine is a program that searches documents for specified keywords and return a list of matching documents.

A. Working of a search engine

Searching information with a search engine consists of the following processes.

- Web Crawling
- Indexing
- Ranking and Retrieval

Web Crawling is a process of traversing the web to collect information about the web pages using special software components called spiders or bots[5][10]. The spiders starts with a given set of URLs (base Urls) or 'seed set'[14] and traverses the links from that URL in a recursive process and he information in repositories until no more URLs are left or a threshold is reached[1].

Indexing[6] process creates a massive index of all the words of the pages it crawled, which results in a database consisting of millions of web pages. This index facilitates searching with a reduced cost. Indexing parses the web pages and extracts terms from them by removal of stop words, stemming and builds an inverted index containing a word and the list of documents in which it occurs.

Ranking and Retrieval algorithm checks the search query with the search pages that have been indexed and retrieves those pages that are more relevant to the query in the top. Each search engine uses a different ranking algorithm providing different weights for different factors because of which they provide different results for the same search query. So highly ranked pages are more relevant to the query and they are displayed in the top of the search engine result pages.

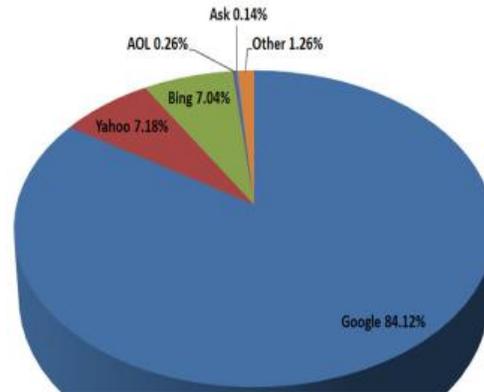


Fig. 2 Market Share of Search Engines

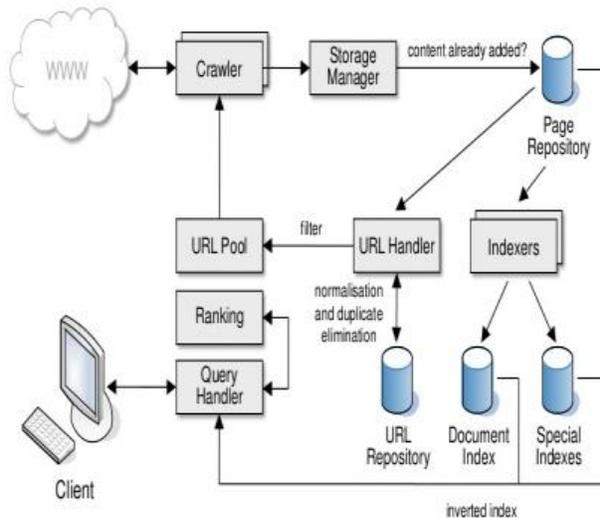


Fig. 1 Working of a search engine

III. SEARCH FACTORS OF SEARCH ENGINES

There are a number of popular search engines available. Some of them are Google, Bing, Yahoo, Altavista, Ask, AOL etc. Each of these search engines ranks the web pages according to some search factors which are listed below.

- Keywords occurring in domain.
- Keyword occurring in TITLE tag or META description tag or H1 tag.
- Fresh content is more preferred i.e. recent information is given more preference.
- Links from branded websites which are more trusted.
- Social signals from social media such as google+, facebook, twitter.
- The number of linking domains and web pages.
- Page Rank(PR)[16] of the web page.
- HTML text is given more preference than multimedia content[7] and hence flash content is not recognized.
- Site speed.
- Backlinks (links pointing to this website) are used to determine the importance, relevance and content quality of the information in the website[7][10].
- Country code Top Level Domain (.in, .uk) are ranked higher for a particular country[4] but this factor limits the site's ability to rank globally.



Bing:

Bing is a search engine launched by Microsoft. It has its roots from MSN and Windows Live which are the earlier search engine versions of Microsoft. Bing provides a wide range of search services, including web, video, image and map search products. The factors considered by Bing are

- ❖ Keywords in domain names.
- ❖ Inbound anchor texts.
- ❖ Age of the web page (Bing favours older websites)
- ❖ Authority of sites.
- ❖ Flash content (is given more preference.)
- ❖ Exact keywords occurring in web pages.
- ❖ Position of the keywords in title.
- ❖ Backlinks.
- ❖ Good quality content.
- ❖ Anchor text.
- ❖ Click Through Rates (CTR)

Yahoo :

The crawler used by Yahoo is called as Yahoo!Slurp. Initially Yahoo was considered as a directory of listings which later provides searching capabilities also. The major ranking factors of Yahoo are

- Presence of keywords in title is more important.
- Keywords in description of the web page in the Yahoo web directory.
- Click popularity – The more the users click on the website from Yahoo Search Engine Results page the more it would be ranked higher.
- Website in Yahoo directory listings.
- Repetition of keywords
- Link anchor text.

Ask:

Ask also known as “Ask Jeeves” allows the users to provide their keywords in the form of simple queries. It uses natural language to answer such queries. Initially used human editors for ranking but later it started using the services of Teoma, a search engine which it acquired later. The factors taken into account by Ask are

- Presence of keywords in title, meta tag
- Keywords in bold, strong, italics, emphasis
- Repetition of keywords in text.
- Presence of keyword related terms
- Number and type of inbound hyperlinks from top ranking relevant and quality websites

- Good page content

AOL :

AOL is the acronym for American OnLine, which uses a dmoz directory for its listings. The search results are displayed in an accessible way by clearly marking sponsored results and web results. The results are almost similar to that of Google, but with a little difference. The search factors considered are

- Click rate of users on the first page of search results
- Keywords in title, meta tags.
- Backlinks
- Anchor text

Altavista :

Altavista means “a view from above”. It was the most popular search engine during its introduction in 1995. With the advent of Google, its popularity reduced. It ranks web search results based on the following factors [13].

- Search keyword occurs in the HTML title or first few lines of the text or in the META tags of the web page.
- The frequency of occurrence of the query word or phrase.
- Whether all the words in the search string are present in the document. A web page having all the words in the query is ranked higher than that containing fewer words.
- Whether multiple query words or phrases are found close to each other in the web page.
- The query words occur in anchor text with hyperlinks to the current pages.
- Long pages rich in meaningful content.

IV. OBSERVATIONS

These popular websites have some common factors such as presence of keywords in title, meta tags, headings and anchor tags, good quality content, links from trusted sites, frequency and position of occurrence of keywords. Some of the varying factors include recognition of flash content (Google does not recognize flash unlike Bing), branded websites (Google favours brands), page rank (important to Google but not in case of Bing), domain age (Google favours fresh content while Bing and Yahoo prefer older content). But the important point is the weights they assign to each of these common and varying factors which determine the relevance of the web pages and their ranking in SERPs (Search Engine Result Pages).



REFERENCES

- [1] G.K. Gupta, "Introduction to Data Mining with case studies", Prentice Hall of India
- [2] www.webopedia.com
- [3] <http://positiononly.com/blog/seo/google-ranking-factors>
- [4] <http://backlinko.com/google-ranking-factors>
- [5] <http://www.makeuseof.com/tag/how-do-search-engines-work-makeuseof-explains/>
- [6] <http://www.totallycommunications.com/latest/search-engine-basics-crawling-indexing-ranking/>
- [7] <http://www.optimizationtheory.com/difference-in-google-and-bings-algorithm/>
- [8] <https://www.quora.com/What-is-the-difference-between-search-methods-and-algorithms-used-by-Google-Yahoo-and-Bing-search-engines-How-would-you-explain-it-in-a-simple-way>
- [9] <http://www.irkawebpromotions.com/search-engines/yahoo/>
- [10] <http://www.webconfs.com/seo-tutorial/>
- [11] http://www.thehistoryofseo.com/The-Industry/Short_History_of_Early_Search_Engines.aspx
- [12] <http://www.seo-theory.com/2007/04/22/seo-algorithm-roundup-aol-ask-google-live-yahoo/>
- [13] <http://www.upenn.edu/computing/web/webdev/meta/rank.html>
- [14] Mike Grehan,(2002) "How Search Engines Work", available at http://cms.searchenginewatch.com/digital_assets/3859/how-search-engines-work-mike-grehan.pdf
- [15] Khalil ur Rehman and Muhammad Naem Ahmed Khan, "The Foremost Guidelines for Achieving High Ranking in Search Results through Search Engine Optimization", International Journal of Advanced Science and Technology Vol. 52, March, 2013 pp 101-110.
- [16] Sergey Brin and Lawrence Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine", available at infolab.stanford.edu.

