

Analytic Study of Major Web Search Engines by Correlating Their Technologies and Results

Ujjwal Agarwal, Nithesh K.Nandhakumar, Faizal H and Suresh Kumar

*Lecturers, Salalah College of Technology, Oman-Post Box No-608, Postal Code-211
E-mail: ujjwal.libya@gmail.com, niteshnanda@gmail.com, faizal.h@gmail.com,
sureshkkk2003@yahoo.co.in*

Abstract

A Search engine usually refers to a web search engine, which searches for information on the public web. This research will show the comparison in terms of technology and results, as seen in different search engines. There are many different search engines in the net arena now, but the most popular three search engines are taken here, viz. Google, Yahoo and Bing. Google search engine, which is one of the prominent one, has been successful owing to its concepts of Link popularity and Page Link. Yahoo's search engine is based on the combine technology of its acquisitions and provided a service that gave pre-eminence to the web search engine. Bing, formerly known as Window Live search was launched in June 2009 and is based on Semantic Technology from Power set. To accomplish our goal, we have done literature survey for the different search engine technologies. Also, search queries pertaining to a particular domain were analyzed using the above mentioned search engines which forms the practical part of the research paper. Our study shows that the algorithms used by different search engines are strikingly different, which is the reason why the same mathematical and textual queries produced completely different search results, out of which Google is using a completely unique search engine architecture for gathering web pages, indexing them, and performing search queries over them.

Keywords: Google search engine, Semantic Technology, Page Ranking, Arithmetic Comparison, Link Popularity.

Introduction

Search engines are programs that search for specific keywords and returns a list of the documents where the keywords were found in the internet. The term Search Engine [2] refers to the process of searching files using the key words specified. The key words found are returned and collated into the user information. A *search engine* is really a general class of programs; however, the term is often used to specifically describe systems like Google, Bing and Yahoo! Search that enable users to search for documents on the World Wide Web. Typically, Web search engines work by sending out a *spider* to fetch as many documents as possible. Another program, called an *indexer*, then reads these documents and creates an index based on the words contained in each document. Each search engine uses a proprietary algorithm to create its indices such that, ideally, only meaningful results are returned for each *query*. [1] Search engines are the key to finding specific information on the vast expanse of the World Wide Web. Without sophisticated search engines, it would be virtually impossible to locate anything on the Web without knowing a specific URL. When people use the term search engine in relation to the Web, they are usually referring to the actual search forms that search through databases of HTML documents, initially gathered by a robot.

Search Engines are of four types [3]. They are

- a) Crawler based search engines.
- b) Human powered directories.
- c) Hybrid search engines.
- d) Meta search engines.

They are further classified based on -that are powered by robots (called crawlers; ants or spiders) and those that are powered by human submissions too. There is another variant which is a hybrid of the above two.

Crawler-based search engines use automated software agents (called crawlers) that visit a website, read the information on the actual site, read the site's meta tags and also follow the links that the site connects to performing indexing on all linked Web sites as well [3]. The crawler returns all that information back to a central depository, where the data is indexed. The crawler will periodically return to the sites to check for any information that has changed. The frequency with which this happens is determined by the administrators of the search engine.

Human-powered search engines rely on humans to submit information that is subsequently indexed and catalogued. Only information that is submitted is put into the index. In both cases, when you query a search engine to locate information, actually searching through the index that the search engine has created—you are not actually searching the Web. These indices are giant databases of information that is collected and stored and subsequently searched. This explains why sometimes a search on a commercial search engine, such as Yahoo! or Google, will return results that are, in fact, dead links. Since the search results are based on the index, if the index hasn't been updated since a Web page became invalid the search engine treats the page as still an active link even though it no longer is. It will remain that way until

the index is updated.

So why the same search on different search engines produce different results? Part of the answer to that question is because not all indices are going to be exactly the same. It depends on what the spiders find or what the humans submitted. But more important, not every search engine uses the same algorithm to search through the indices. The algorithm is what the search engines use to determine the relevance of the information in the index to what the user is searching for.

Review Stage

Google Search Engine



Figure 1

In 1998, two Stanford University students, Larry Page and Sergey Brin, created Google, an Internet search engine. Their success was laid in the fact that they could quickly and efficiently deliver relevant search results to its users. Google was a well-developed system that could show immediate results, based on the top queries of its unique “page-rank” system [4] [5]. The page-rank system could easily read documents on the Internet, reporting back the page’s individual performances within certain criteria. Google’s page-rank system is a mathematical equation that brings together and defines three main areas of interest: content, backlinks, and outbound links, with the ultimate idea of determining a website’s “page-rank,” enabling the website to be pigeonholed into a ranking against all the millions of Internet sites in the world. The content that Google looks at, can be described as any words, information, or photos that are placed on a website to create a general topic. That is, simply put: all information included on any given site. The content then goes on to provide a base from which “keywords” can be pulled. Keywords are the most specific words derived from the real purpose or content of your website. Keywords contain the composite meaning or definition of what your website is really all about. Back-links are fairly simple; they are defined as any links pointing to your website from another website. They are called back-links because they simply go back to where they came from. The root of these links is from the original name of Google: Backrub—which literally

meant that if you rubbed my back, I'll rub yours! In essence, Google was trying to determine who was an expert in any given field by who was confirming, from other places, that the website was really an expert on the subject. Finally, Google likes to see that you are pointing to another site, or giving away back-links. Google believes that if your website is not an expert in a subject, then you might point people to the website that is an expert. The easier way to look at it is that your site is recommending someone else as a source of more important information. If you look beyond what Google technically achieves with its search engine, it has molded itself as a strong company to follow in the marketplace.

Yahoo Search Engine

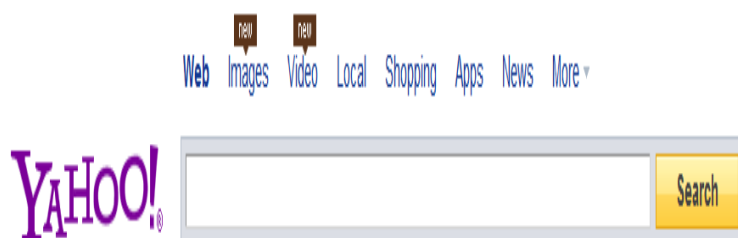


Figure: 2

Yahoo! Search is a web search directory, owned by Yahoo! Inc. and as of December 2009, was the 2nd largest search directory on the web by query volume, at 6.42%, after its competitor Google at 85.35%. Yahoo! is so difficult to gauge is because they haven't really built a search engine from the ground up like Google or MSN. Of course the Yahoo! search you see is unique unto itself, however Yahoo! has built its search on the backs of other technologies they had purchased in previous years. It was around 2002 when Yahoo! purchased search service Inktomi. Yahoo! had received their search results either from Inktomi or more recently Google. In fact, until the time they purchased Inktomi, there was speculation that Yahoo! would buy Google. It was just a few months after this that Overture (a pay-per-click advertising company) purchased Altavista – one of the first and strongest search engines out there. Then, just a few weeks after that Overture purchased Alltheweb.com from FAST.

In October 2007, Yahoo! Search was updated with a more modern appearance in line with the redesigned Yahoo! home page. In addition, Search Assist was added; which provides real-time query suggestions and related concepts as they are typed. In July 2008, Yahoo! Search announced the introduction of a new service called "Build Your Own Search Service," or BOSS. This service opens the door for developers to use Yahoo!'s system for indexing information and images and create their own custom search engine. In July 2009, Yahoo! signed a deal with Microsoft, the result of which was that Yahoo! Search would be powered by Bing. This is now in effect.

Bing Search Engine

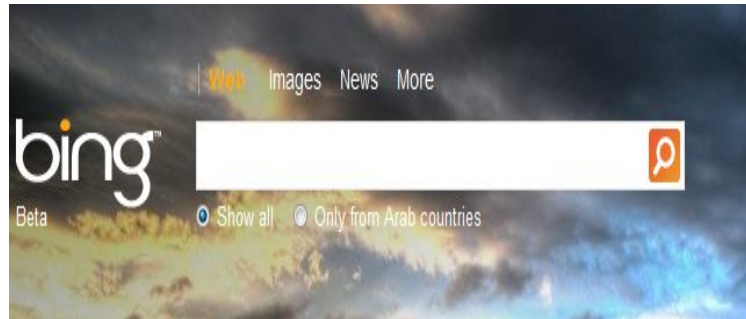


Figure: 3

Microsoft refers to Bing as a "decision engine." With nearly 240 million Web sites on the Internet as of July 2009, you can imagine how vital an effective search method can be to the success of any search engine.

What's the difference between a search engine and a decision engine? Perhaps it's philosophical. According to Microsoft, Bing is designed to minimize the amount of junk you get when you perform a search and to help simplify tasks so that you can make the most informed decision. Bing focuses on four target areas: shopping, travel, local and health. Its stated underlying goal is to simplify search. It starts with Bing's homepage, which displays a search box in the middle of a colorful image and a row of clear links on the left of the page that bring up results for video, news, shopping, images, travel or maps. Once you begin your search, Bing has an **Autosuggest** feature that recommends words based on the first few letters you type, and then lists them for you to choose if one should meet your match. **Best Match** is similar to Autosuggest -- it offers you what Bing believes is the most suitable match. In October 2011, Bing announced that it is working on new back-end search infrastructure, with the goal of delivering faster and more relevant search results for users. Known as "Tiger," the new index-serving technology has been incorporated into Bing.

Final Stage Comparison Between The



Figure: 4

In the final stage we are doing the different comparison between the Google, yahoo and Bing search engines.

Arithmetic Result Comparison

Table: 1

Search Criteria	Google	Yahoosearch	Bing
2+2=	found (4)	found (4)	no result
(2+2)/4=	found (1)	found (1)	no result
Cos 30 degrees=	found 0.866025404	no result	no result
square root of 4 =	found (2)	no result	no result
Cube root of 12=	found (2.28942)	no result	no result
log of 10 =	found (1)	no result	no result
sin60 degrees =	found (0.86602)	no result	no result

Here we analyzed the arithmetic results of some search criteria giving some pre-defined parameters. The results were given at the same time in the different search engine portals Google, Yahoo and Bing. Google and Yahoo-search gave results for simple arithmetic calculations correctly where as Bing didn't give any. Bracketed and complex calculations also yielded results in Google and Yahoo, but not in Bing.

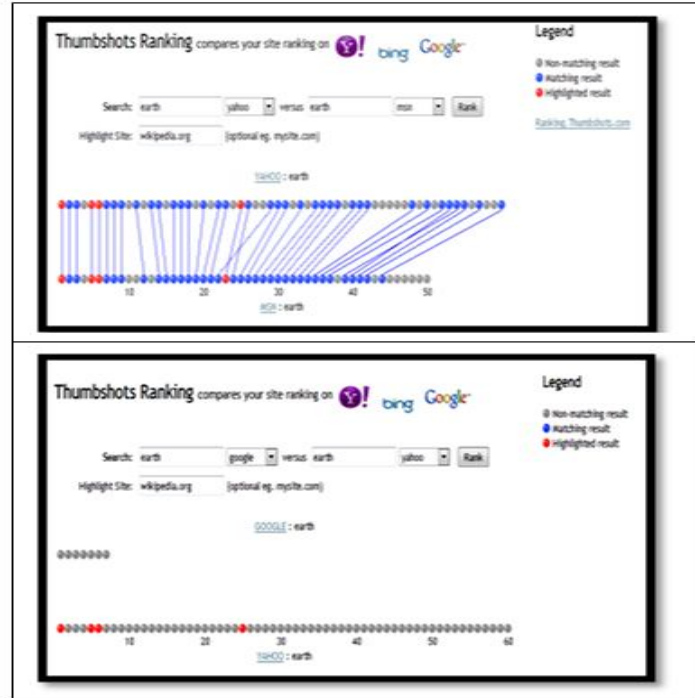
Logarithmic and trigonometric queries were answered well in Google, but not in yahoo and Bing. Google in short, excelled in simple, complex and trigonometric/log mathematical queries in comparison to other search engines.

Time and Search Result Comparison (as on 12/04/2012)

Table: 2

Search criteria (earth)	Google	Yahoosearch	Bing
No. of Search Result	1,440,000,000	878,000,000	867,000,000
Time taken for Search	0.26 Seconds	0.63 Seconds	0.47 Seconds

For time based comparison of query results, a common search criteria was given to all the search engines, viz., google, yahoo and bing portal. "Earth" was the query that was given to the engines and the time taken was taken and tabulated as above. Along with that, the number of search queries that resulted was also taken. The results were also in favor of google search engine which gave almost twice the speed than the average of speed given by yahoo and bing. Also the number of search queries that were obtained by google was also almost 16% more than that was obtained by the average obtained by yahoo and bing.

Page ranking comparison (as on 12/04/2012)**Figure 5**

Here the page ranking comparison of Yahoo and Msn/Bing was conducted using Thumbshots ranking method and here the ranking was almost comparative. The query “Earth” gave almost exact query ranking, which reflects the fact that the search engine algorithm used by the two portals are the same. Where as, when we compared page ranking of Google and Yahoo, using Thumbshots, the results were 100% non-matching. This will be infer the fact that Google is using a totally different search engine algorithm when compared to the other two.

Conclusion

This research sheds light on the comparative analysis of different search engine technologies and their subsequent query results. Google search engine uses the concepts of Link popularity and Page Link where as Yahoo’s search engine is based on the combined technology of its acquisitions and provided a service that gave pre-eminence. Bing is based on Semantic Technology in terms of its algorithmic search. In this paper we have done comparisons and the results were analyzed to find the most effective query processing in terms of mathematical, time and page-ranking parameters-

1. **Arithmetic Comparison:** Bing is deprived of any calculator like facility, where as in Yahoo only simple arithmetic calculation is integrated. As far as

Google's mathematical queries are concerned, it can do all the parametric equivalent, scientific as well as simple arithmetic calculations.

2. **Time and Search Result Comparison:** Based on the parameters that we analyzed and the results that were obtained, Google during the experiment setup gave much faster search results when compared to Yahoo and Bing search engines
3. **Page Ranking:** Thumbshots- tool was used in the ranking method to query the hypertextual web word- "Earth" which inferred the fact that the search engine algorithm used by the yahoo and bing are the same. Comparison using the same tool for page ranking google and yahoo gave a 100% non-matching result. From this we can infer that Google is using a totally different search engine algorithm because of the 100% non matching of the query terms

Here, as discussed, the experimental bed was based on the comparative effectiveness of the query processing done by the different search engine algorithms. Google was the best pick out of them which was found to be using a totally different and unique search algorithm, as evident from the page-ranking that was obtained, as well as took the least time for executing the query along with displaying arithmetic results in the search page itself.

References

- [1] Mike Thelwall, (2002) "Methodologies for crawler based Web surveys", Internet Research, Vol. 12 Iss: 2, pp.124 – 138
- [2] Wen-Jen Yu, Shrane Koung Chou, "A Bibliometric Study of Search Engine Literature in the SSCI Database", Journal of Software, Vol5, No 12 (2010), 1317-1322, Dec
- [3] J. M. Kassim & M. Rahmany, Introduction to Semantic Search Engine", Proc. International Conference on Electrical Engineering and Informatics, 2009.
- [4] Sanjay Ghemawat, Howard Gobioff & Shun-Tak Leung, "The Google File System", Proc. The Nineteenth ACM Symposium on Operating Systems Principles, pp. 29-43, 2003.
- [5] William Yip & Dr. Liz Quiroga, "Google Page Rank Algorithm", LIS 678 Personalized Information Delivery, Oct 11, 2008.
- [6] An Intro to SEO "SEARCH ENGINE OPTIMIZATION." written by Richard Khor Available at <http://www.free-ebooks.net/ebook/Search-Engine-Theory-Optimization-And-How-Do-Search-Engines-Work>
- [7] Page Ranking tools available at: www.thumbshots.com
- [8] Types of Search Engine and its Definition available at: www.webopedia.com/term/s/seacch_engine.html