International Journal of Computer Science and Engineering (IJCSE)
ISSN(P): 2278-9960; ISSN(E): 2278-9979
Vol. 4, Issue 4, Jun - Jul 2015, 35-42

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

A COMPARATIVE ANALYSIS OF VARIOUS SEARCH ENGINE PRESENT TODAY

PRATIBHA SHARMA[1], SONAL SHUKLA[2] & BRAHMDUTT BOHRA[3]

[1,2]M.TECH Research Scholar, Department of Computer Engineering, Maharishi Arvind College of Engineering & Research Center, Sirsi, Jaipur, Rajasthan, India

[3]Associate Professor, Department of Computer Engineering, Maharishi Arvind College of engineering & Research Center, Sirsi, Jaipur, Rajasthan, India

ABSTRACT

Online searching has become a common method for obtaining information. As popularity of web increases, millions of people use search engines to discover information. But search engine users are interested only in top few result pages. SEO (Search Engine Optimization) is the art, craft and science of driving web traffic to web sites. There are various search engines like Google, Yahoo, Bing, Ask and MSN.

In this paper we will study about the process of search engine optimization, the algorithms and methods of SEO, also the categories of SEO. We are also providing the comparison between the three most popular search engines like Google, Yahoo and MSN on the basis of techniques and methods they are using for optimization of search engines.

KEYWORDS: A Comparative Analysis of Various Search Engine Present

INTRODUCTION

WWW plays a vital role in day to day life of people. A large amount of information is available on the web. A web search engine is a software system that is designed to search for information on the WWW [3]. We rely on search engines to provide us right information at right time. To satisfy users need search engine must find and filter most relevant information matching a user query and display that information to the user[2]. Search Engine Optimization (SEO) is the process of increasing website's degree of exposure to the public. It means that it helps spectators to find the website. SEO is a process in which website achieves solid, high rankings of pages of Search engines for most of the important keywords [11].

Today, there are hundreds of different search engines on the Internet, and all are having their abilities and features. There are multiple methods used by these search engines to gather data. Google, Yahoo, Ask, MSN and Bing are five among many search engines. You will find basically three forms of search engines: The ones that use robots (called crawlers; ants or spiders) and others who are powered by human submissions; and those who are a hybrid of the two. In Human-powered search engines information is submitted by humans that's subsequently indexed and catalogued. Only information that's submitted is put in to the index. In this paper we will study about comparison between three widely used search engines that are Google, Yahoo and MSN [11].

INTRODUCTION TO SEARCH ENGINE OPTIMIZATION

SEO stands for search engine optimization and it's art of improving a website (in terms of its design, structure, navigation and content) so it's more easily found within search engines like Google, Yahoo and MSN [6]. A huge number of websites are submitted everyday to the global search engines, local search engines and they are indexed. The web marketing through websites is used by almost all the organizations today. Today everyone creates website to advertise their business,

personal use etc. And the reason behind is either to earn some additional revenue or to advertise their products online. In order to achieve these business or personal objectives, either paid or unpaid organic traffic is required.

SEO is an abbreviation of the word Search Engine Optimization which means preparing Website for the search engine. The word SEO means friendship between website and Search Engine, the friendship of search Engine has rules and basics on that anyone can prepare the website for it [4].

WORKING OF SEARCH ENGINE

Search engine uses indexing software which is generally known as robots or spiders. These are programmed to find either the new documents or updated documents. Spider crawls the website, finds these new documents. It does so by following the hyperlinks of those websites which are already in the database of the search engine. Search engine then makes the index of the content and add it to their databases. Search engine ranks website by using factors and assign the weight by using their own mathematic formula. The index considers the words that are on the crawled webpage, the location where they are listed and their coding. So when the user enters the query in the search engine, the search engine first checks its own index by scanning its list, and if it finds the matching record it sends the result back to the user[11].

CATEGORIES OF SEO TECHNIQUES

- White Hat SEO techniques - White Hat SEO techniques follow some rules and guidelines suggested by the search engines. It is also called ethical because it follows certain legal guidelines and policies to achieve high ranking. Some most popular white Hat SEO techniques include Quality Content, Quality inbound links, certain tags, effective keyword use [11].
- Black Hat SEO techniques- These techniques on the other hand are unethical way of manipulation of search engine's algorithm. These techniques are not recommended as there is a risk of website getting removed from the search database index. So it does not always increases the ranking but also increases the chances of the website getting banned by the search engine. So these techniques are not approved by the search engines. Some most popular Black Hat SEO techniques include hidden text or links that are same color as the background color, keyword stuffing means loading a webpage with the keywords, Doorway Pages which takes the visitor to the page where products or services are promoted, cloaking [12] purpose is to present a different page to the visitor and the search engine, link farms which takes to the spammed sites

ORGANIC AND SPONSORED RESULTS

Almost all search engine results consist of two types, organic and sponsored. The organic or natural results are those search engine display on their own. The sponsored results typically appear at the top and side of the organic results.SEO directly affects only the organic results however it does have an indirect effect on paid search results [8].

SEO PROCSESS

There are several segments of SEO strategy seen as optional that are actually absolutely imperative to the success of an SEO campaign Website analysis.

Competitive Analysis.

Proper Keyword Research.

Content Development.

On-page Optimization.

Off-page Optimization.

Return on Investment Analysis.

Local SEO.

Daily, weekly and monthly reports [9].

SEARCH ENGINE ALGORITHM

Page Rank (PR): Page Rank is an algorithm in which a numerical weight is assign to a webpage according to its relative importance. It uses incoming link information to assign global importance score to all pages on the web. Number of incoming links from quality sites measures the popularity of a page. It is based on quantity and quality of both inbound and outbound links. Pages which have higher rank are most important and it has chances to be listed on search engine's top result list. Page rank value is divided into levels 1-10 of which 10 represent higher PR value means that page is more popular while page rank value 1 means page is not popular. The web page which got position among first 25 top results, PR value should be 6 or above 6.

Suppose t1,t2..tn are pages linking to page A then Page-A has its PR value as follows:

PR(A) = (1-d) + d{PR(t1)/C(t1)+PR(t2)/C(t2)+...+PR(tn)/C(tn) } (1)

Where d is damping coefficient, usually its value is 0.85. PR(t1)...PR(tn) is page t1 to tn page rank value, C(ti) means number of outgoing links page ti . PR(ti)/C(ti) means page ti's contribution to page A's PR value 2.

HillTop Algorithm: When a query is given, HillTop first compute list of most relevant experts on the query topic. Then

Identify relevant links within the selected set of experts and follow them to identify target web pages. According to number and relevance of non-affiliated experts that point to them, target pages are ranked. So the score of a target page reflects the collective opinion of the best independent experts on the query topic. When expert's opinion is not available, Hilltop provides no results. Thus, Hilltop is tuned for result accuracy and not for query coverage. Hilltop is topic sensitive. It generate list of authoritative pages on topic of query. Each page is given a weight on binary scale. Value "1" represent good page on the topic and "0" indicate not relevant or not found. Thus HillTop is for result accuracy and not for query coverage.

Thus, we compute the score of an expert. Let k be the number of terms in the input query, q. The component Si of the score is computed by considering only key phrases that contain precisely $k - i$ of the query terms.

Si = SUM{key phrases p with $k - i$ query terms} LevelScore(p) * Fullness Factor (p, q) (2)

LevelScore(p) is a score assigned to the phrase by virtue of the type of phrase it is. FullnessFactor(p, q) is a measure of the number of terms in p covered by the terms in q. The score of each expert is converted to a scalar by the weighted summation of the three components 7:

Expert_Score = 232 * S0 + 216 * S1 + S2.

New algorithm (Combination of PR and HillTop): Now to give more accurate result in scientific and rational way, Google combine features of PR and HillTop to calculate ranking value of webpage. This algorithm has formula as:

{(1-d) a(RS)} * { (1-e) + b(PR*fb)} *{(1-f)+c(LS)} (4)

where a, b, c are the regulating controls of weight and d, e, f, are damping controls . RS = relevance Score. It is translation of all SEO factors.(Score based on keywords appearing in Title tag, Meta tag, Headlines, Body text, URL tag, Alt text, anchor text etc.) PR = Page Rank score. LS = Local Score. It is translation of links from expert documents [2].

In next section of the paper we will study comparison of various search engines like Google, Yahoo and MSN.

COMPARISON BETWEEN GOOGLE, YAHOO AND MSN

Google

Google Search is a web search engine owned by Google Inc. Google Search is the most-used search engine on the World Wide Web, receiving several hundred million queries each day through its various services. Google Search was originally developed by Larry Page and Sergey Brin in 1997. Google Search provides at least 22 special features beyond the original word-search capability. These include synonyms, weather forecasts, time zones, stock quotes, maps, earthquake data, movie show times, airports, home listings, and sports scores [3]. Google is designed to scale well to keep up with the growth of web. It gives exactly what we want. For fast and efficient access, its data structures are optimized. In addition to smart coding, on the back end it developed distributed computing systems around that globe that ensure fast response times[1].

Google has been in the search game a long time, it has the highest share market of Search Engine (about 81%) [7].

- Web Crawler-based service provides both comprehensive coverage of the Web along with great relevancy.
- Google is much better than the other engines at determining whether a link is an artificial link or true editorial link [8].
- Google gives much importance to Sites which add fresh content on a regular basis. This is why Google likes blogs, especially popular ones [7].
- Google prefers informational pages to commercial sites.
- A page on a site or sub domain of a site with significant age or link can rank much better than it should, even with no external citations.
- It has aggressive duplicate content filters that filter out many pages with similar content.
- Crawl depth determined not only by link quantity, but also link quality. Excessive low quality links may make site less likely to be crawled deep or even included in the index [7].
- In addition we can search for twelve different file formats, cached pages, images, news and Usenet group postings.

Yahoo

In January 1994, Yang and Filo were electrical engineering graduate students at Stanford University when they created a website named “Jerry’s guide to the World Wide Web”. In March 1994, “David and Jerry’s Guide to the World Wide Web” was renamed “Yahoo!” The “yahoo.com” domain was created on January 18, 1995. The word “yahoo” is an acronym for “Yet another Hierarchical Officious Oracle”. The term “hierarchical” described how the Yahoo! database was arranged in

layers of subcategories. The term “oracle” was intended to mean “source of truth and wisdom”, and the term “officious”, rather than being related to the word’s normal meaning, described the many office workers who would use the Yahoo! database while surfing from work. Yahoo! operates a portal that provides the latest news, entertainment, and sports information. The portal also gives users access to other Yahoo! services like Yahoo! Mail, Yahoo! Maps, Yahoo! Finance, Yahoo! Groups and Yahoo! Messenger.

Yahoo asks for many similar optimization techniques to be used—no duplicate content, solid back links, etc.—to get ranked. Experts all agree that Yahoo is headed for a downward spiral. Spending time optimizing your pages for something unpredictable might not be worth it, even if the search engine does bring in some users [5].

- It shares the second largest share market of the search engine (about 12%). Yahoo has been in the search game for many years.
- When it comes to counting back lings, Yahoo is the most accurate search engine.
- Yahoo is better than MSN but near as well as Google at Determining whether a link is artificial or natural.
- Crawl rate of the Yahoo's spiders is at least 3 times faster than Google‟s Spiders.
- Yahoo! tends to prefer commercial pages to informational pages as comparing with Google.
- At Yahoo search engine "exact matching" is given more importance than "concept matching" which makes them slightly more susceptible to spamming.
- Yahoo! gives more importance to meta keywords and description tags.

On the basis of Crawlers and database Google, Yahoo and MSN [10].

Google	Googlebot
MSN	MSNbot
Yahoo	Yahooslurp

MSN

MSN (stylized as msn, originally standing for The Microsoft Network) is a web portal and related collection of Internet services and apps for Windows and mobile devices, provided by Microsoft and launched on August 24, 1995.[3]

Along with the release of Windows 95, Microsoft debuted The Microsoft Network, a subscription-based dial-up online service that it later turned into an Internet service provider named MSN Dial-up. At the same time, the company launched a new web portal named Microsoft Internet Start and set it as the first default home page of Internet Explorer, its original web browser. In 1998, Microsoft renamed and moved this web portal to the domain name MSN.com, where it has remained since then.

MSN’s Search Engine Beta algorithm has a lot in common with Google’s and Yahoo!’s algorithms, but also includes some subtle yet significant differences. MSN Search Engine has been launched in a beta format. The Microsoft owned search engine has generated much interest and discussion among the general search using public and the search engine optimization (SEO) community. the first and probably most important aspect of the MSN Search Beta is the need for good keyword rich content. Relevance of theme and topic appears to be very important to MSN, as it is becoming for Google optimization.

- MSN has the share of 3% of the total search engine market.

- MSN Search uses its own Web database and also has separate News, Images, and Local databases.
- Its strengths include: this large unique database, its query building "Search Builder" and Boolean searching, cached copies of Web pages including date cached, and automatic local search options.
- The spider crawls only the beginning of the pages (as opposed to the other two search engine which crawl the entire content) and also the number of pages found in its index or database is extremely low.
- It is bad at determining if a link is natural or artificial in nature.
- Due to sucking at link analysis they place too much weight on the page content.
- New sites that are generally untrusted in other systems can rank quickly in MSN Search. But it also makes them more susceptible to spam.
- Another downside of this search engine is its habit of supplying the results based on geo-targeting, which makes it extremely hard to determine if the results we see are the same ones everybody sees.

Table 1: Comparative Analysis of Popular Search Engine Optimization Techniques

Search Optimization Techniques	Google	Yahoo	MSN
Keyword optimization	Google Adwords, Google Adsense, DoubleClick,	keyword density, H1 tags	Descriptive page titles and page keyword help you to get a good rank in MSN.
Link optimization	PageRank is a weighted measure of link popularity Factors affecting are: Rate of link acquisition, Link age , link source quality, anchor text diversity, deep link ratio etc.	Yahoo! places more importance on the number or quantity of backlinks	Attracted by new links
Website content	Keywords in the url. Keywords in the title. Sites which add fresh content on a regular basis will be rewarded. This is also known as “buzz”.	Meta keywords Description tags, Yahoo’s search engine, focuses on on-page optimization	Less emphasis on link popularity and more on page content
Website optimization	Biased toward informational websites and web pages	“Exact Matching” is more important than “concept matching”	New sites that are generally not trusted (according to other search engines) can get a good rank in MSN Search
Social media optimization	Youtube, Blogger, Gmail, Chrome, Google+, Orkut	Prefer commercial pages, Yahoo Messenger Del.icio.us, a social bookmarking site, Yahoo! Answers	Paid search and ad campaigns are there in msn. Microsoft’s paid search product is Ad Center It uses Online Commercial Intention Detection Tool
Tools used for SEO	Google Sitemaps Ad Words Keyword Tool Ad Words, Traffic Estimator, Google Suggest, Google Trends, Google Sets, Google Zeitgeist, Google related ites, Google related word search	Overture Keyword Selector Tool Overture View Bids Tool Yahoo! Site Explorer Yahoo! Mindset Yahoo! Advanced Search Page Yahoo! Buzz	Keyword Search Funnel Tool Demographic Prediction Tool Online Commercial Intention Detection Tool Search Result Clustering Tool

CONCLUSIONS

Today the amount of information available on the Web is growing rapidly. Search Engine technology had to scale according to the growth of the Web. The present study estimated the precision of Google, Yahoo and Bing. The results of the study also showed that the precision of Google was high as compared to Yahoo and Bing and Yahoo has better precision than Bing. It was observed that Google, Yahoo and Bing showed diversity in their search capabilities, user interface and also in the quality of information. However these two search engines retrieved comparatively more relevant sites or links as compared to irrelevant sites. Google utilized the Web graph or link structure of the Web to become one of the most comprehensive and reliable search engines.

REFERENCES

1. **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*
2. **Patil Swati P., Pawar B. V. and Patil Ajay;**" Search Engine Optimization: A Study" *Research Journal of* Computer and Information Technology Sciences Vol. 1(1), 10-13, February (2013).
3. **http://www.computerscijournal.org/vol7no1/a-comparative**-study-of-two-major-search-engines-google-and-yahoo/].
4. **Deepti Gaur, Preety Dagar ITM University ,Gurganv**, India" SEO Techniques: New Dimensions for Popular Search Engines "International Journal on Advanced Computer Theory and Engineering (IJACTE) Volume -3, Issue -4, 2014.
5. http://www.marketingprofs.com/opinions/2012/23655/seo-for-bing-and-yahoo-is-it-really-worth-the-effort#ixzz3YCkmXDq2.
6. http://www.resource-media.org/wp-content/uploads/2012/07/SEO_Guide.pdf.
7. http://seolutions.net/Search-Engine-Differences.php.
8. http://softnik.com/download/seobasics.pdf.
9. **Geetanjali Tyagi, Megha Sharma, Kumar Kaushik ;**" Using Search Engine Optimization Technique Increasing Website Traffic and Online Visibility" International Journal of Advanced Research in Computer Science and Software Engineering Volume 5, Issue 1, January 2015.
10. **Vinit Kumar Gunjan, Pooja, Monika Kumari,Dr Amit Kumar,Dr (col.) Allam appa rao ; "**Search engine optimization with Google" IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 3, January 2012.
11. **Gurpreet Singh Bedi, Ms. Ashmita Singh** ; "Analysis of search engine optimization" International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue 3, March 2014.
12. **Antriksha Soni, Ugrasen Suman,** "Counter Measures against Evoloving Search Engine Spamming Techniques," in IEEE, 2011. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5942084