Web Health Analyzer for Search Engine Optimization

Manpreet Kaur, Hari Singh

***Abstract*— 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. Search Engine Optimization (SEO) relates to the activity of optimizing individual web-pages or the entire website to make them friendlier to acquire higher raking in the search results. All the major search engines such as Google, Yahoo, Ask, Bing, etc. rank web-pages based on certain factors that affect its ranking; therefore, SEO aims at generating the right types of signals on the web-pages. The optimized websites obtain better ranks, and typically get a higher number of visitors. This research is based on reviewing different optimization techniques for individual web-pages or the entire website to make them search engine friendly.This paper offers a comparative study of the previous research work regarding the techniques used in SEO and pinpoints certain gaps in the known search engine optimization techniques. Finally, we put forward some pertinent guidelines for optimizing the websites.**

***Index Terms*— Search Engine Optimization, Search Ranking, Web Crawling.**

I. INTRODUCTION

The Proposed System **"SEO Analyzer"** is known as search engine optimization analyzer. This tool is a web based system which is going to give us the analysis report of the website. The proposed system will help the website owner to know whether their website has all the necessary content that would make it search optimized. This tool will help the users take their business to a next level as this system will help the users know where their website is in the world and what scope of their business is if the website was search engine optimized. This software will give the summary of the website that can be used to make it more efficient.It is important to look across the entire broad range of search terms related to our business and build a plan for getting better visibility for the search terms that do not afford us that visibility.There are some major purposes of system to be developed.

1. Increase rankings

"Google isn't a search engine, Google is a reputation management system" Clive Thompson, Wired Magazine

SEO ANALYZER will help the general users to know about the ranking of the web page and they can they take important steps to increase the ranking of the website.

2. Increase Conversions

Manpreet Kaur, Bachelor of Engineering in IT from NCIT israna panipat and Master of Technology in Computer Science & Engineering fom NCCE israna panipat

Hari Singh, Bachelor of Engineering (with honours) and Master of Technology in Computer Science & Engineering. He has been pursuing Ph.D. in Computer Science from Thapar University, Patiala, India.

"Design is not just what it looks like and feels like. Design is how it works." – Steve Jobs

This tool is going to give us the analysis report of the website. The proposed system will help the website owner to know whether their website has all the necessary content that would make it search optimized. The proposed system will have a spider that would crawl across the website and gather information about the website's pages. Like the keywords that are there in the page. The descriptions that are in the page. And all the Meta tags that are in the page. Based on these data the tool will give the analysis report to the users about the page. The analysis report will have all the details listed and a percentage chart about where the users can see the density of the Meta tags in the page and frequency of the words.

The proposed system will also have a Suggestive tool which will generate random keywords, and description based on the keywords and description fetched by the SEO analyzer.
The Proposed system will also have a tool named as Meta Tag Creator. That will help the intended users to create Meta tag which could be used in their website for make it more search optimized.
These details can then be used by the SEO professional to make the website more search engine optimized.
The tool will also have the features like any registered user can know the page ranking of a particular page by just entering in the URL of the page.

II. LITERATURE SURVEY

[1] G. S. Bedi,(March 2014), shows the importance of some techniques of Search Engine Optimization (SEO), He described previous work done on the subject of SEO, the early beginning of SEO, its goals and ideas such as White Hat SEO, Black Hat SEO, On page SEO, Off Page SEO, their benefits.
[2] Dr S. Saravanakumar, (September 2012), implied a new methodology of Search Engine Optimization (SEO) without getting sandboxed by search engines like Google, Bing and other. In the past, the algorithm was based on the quantity of back links that a site has. This process involved in implementing safe link building techniques with link velocity as its key without compromising the on page optimization. The latest algorithmic updates are taken in to consideration and the strategy is developed to rank for a keyword. By implementing this method, any organization can take advantage of the traffic from the search engines and have a good online presence. He also explained basic guidelines recommended by all the search engines for proper indexing without sandboxing. Hence even in the future; this method will not hinder the online progress of any business.
[3] Joeran Beel, discussesed the concept of academic search engine optimization (ASEO). Based on three recently conducted studies, guidelines are provided on how to

optimize scholarly literature for academic search engines in general and for Google Scholar in particular. In addition, he briefly discussed the risk of researchers' illegitimately 'over - optimizing' their articles.

[4] Rajesh Singh (September 2013) discussed that extracting information is too much costly in terms of processor time. A distributed design approach is proposed in this paper. He discussed how to design a such type of search engine architecture in which processing of a search engine may be in distributed form by which the time or processing power may be reduced.

Search Engine Optimization

Generally people visit a website to find out information according to their need. But if they did not find right content, they became frustrated and immediately click away from site. So in order to draw their attention and bring them back maximum number of times, website is built up with proper target and quality content. It fulfills user need as well as improves rank position in search engines result list. Search engine optimization (SEO) is the process of improving the number and quality of traffic to a web site from search engines via natural listing for targeted keywords.

SEO help the web site designer to get top ranking position in search result list, attract more online visitors and finally improve the marketing capability of site.

SEO technique entails two major processes: on-site optimization and off-site optimization.

a. ***On-site Optimization*:** On-site optimization pertains to the practices which are used during the development of a website. On-site optimization focuses on factors affecting the website to do with its hosting, structure, accessibility and content. Users enter keywords to search information on search engines. Keyword represents the relationship between search term and several billion of web pages. On-site optimization includes website design elements such as keyword in meta tag, keyword in title tag, external link, keyword density etc., which are controlled by site itself.

Title tag: Title is the biggest ranking factor. Most search engine use the website's title tag as main factor of sites listing in search result pages11.

Keyword density: Density of Keyword means frequency of keyword present on web page compare to total number of words on the page. Frequency of keyword in title tag and frequency of keyword in body tag should be strong optimization factor. Density of keyword should be within 2% -8% for improving website ranking12.

Keyword in URL: Keyword included in URL, The website will be found more easily by search engine crawlers if keyword included in URL. Search engine pays priorities to different

domain name suffixes like edu or gov. Also shorter length URL is preferred in Search Engine Optimization13.

Keyword in Meta tag: The meta description tag contains description of page that is informative and reflects the content of web page. The website will be indexed if related keywords are found in meta description tag.

Off-site Optimization: Off-site optimization relates to the practices through which the website and its contents are propagated/proliferated over the Internet to increase its traffic which in turn helps enhancing Google ranking of the website. Off-site optimization revolves around the links that point to the site from other web pages. These links back to the site are called back links. Site with most back links in most cases will come out on top. Offsite optimization includes the following techniques:

Link Reputation: Web pages and websites with more number of back links improve ranking in search engine result. But it is important that the quality of external links is also very important. External links must have good reputation, relevant or similar content. Also have key phrases similar to search term.

TECHNIQUEs

PRE EXISTING GOOGLE SEARCH TECHNIQUES

Google Penguin

Google Penguin is a codename for a Google algorithm update that was first announced on April 24, 2012. The update is aimed at decreasing search engine rankings of websites that violate Google's Webmaster Guidelines by using now declared black-hat SEO techniques involved in increasing artificially the ranking of a webpage by manipulating the number of links pointing to the page. Such tactics are commonly described as link schemes. According to Google's John Mueller, Google has announced all updates to the Penguin filter to the public.

Effect on search results

By Google's estimates, Penguin affects approximately 3.1% of search queries in English, about 3% of queries in languages like German, Chinese, and Arabic, and an even bigger percentage of them in "highly spammed" languages. On May 25, 2012, Google unveiled another Penguin update, called Penguin 1.1. This update, according to Matt Cutts, was supposed to affect less than one-tenth of a percent of English searches. The guiding principle for the update was to penalize websites using manipulative techniques to achieve high rankings. The purpose per Google was to catch excessive spammers. Allegedly, few websites lost search rankings on Google for specific keywords during the Panda and Penguin rollouts.[according to whom?] Google specifically mentions that doorway pages, which are only built to attract search engine traffic, are against their webmaster guidelines.

Google Panda

How the Google Panda Algorithm Works??

Google Panda is a change to Google's search results ranking algorithm that was first released in February 2011. The change aimed to lower the rank of "low-quality sites" or "thin sites", and return higher-quality sites near the top of the search results. CNET reported a surge in the rankings of news websites and social networking sites, and a drop in rankings for sites containing large amounts of advertising. This change reportedly affected the rankings of almost 12 percent of all search results. Soon after the Panda rollout, many websites, including Google's webmaster forum, became filled with complaints of scrapers/copyright infringers getting better rankings than sites with original content. At one point, Google

publicly asked for data points to help detect scrapers better. Google's Panda has received several updates since the original rollout in February 2011, and the effect went global in April 2011. To help affected publishers, Google provided an advisory on its blog, thus giving some direction for self-evaluation of a website's quality. Google has provided a list of 23 bullet points on its blog answering the question of "What counts as a high-quality site?" that is supposed to help webmasters "step into Google's mindset".(https://en.wikipedia.org)

III. CONCLUSION

The research here is for a tool which is known as search engine optimization analyzer. This tool is going to give us the analysis report of the website. The proposed system will help the website owner to know whether their website has all the necessary content that would make it search optimized. An important aspect of Search Engine Optimization is making your website easy for both users and search engine robots to understand. Although search engines have become increasingly sophisticated, in many ways they still can't see and understand a web page the same way a human does. SEO Analyzer helps the engines figure out what each page is about, and how it may be useful for users.SEO Analyzer will help these pages to be more search engine optimized and hence it will be benefit for increasing the traffic on the website and hence the business will also increase.

REFERENCES

[1]Hartzer, B. (2014). search-engine-ranking-problems. Retrieved October tuesday, 2014, from https://www.billhartzer.com: https://www.billhartzer.com/search-engine-ranking-problems/

[2]Search, T. (2011, September Monday). why-do-you-need-seo. Retrieved November Thursday, 2014, from http://thinksearch.co.uk: http://thinksearch.co.uk/seo-consultancy/why-do-you-need-seo/

[3]Waterfall model,14. (n.d.). Retrieved September 2014, 28, from http://istqbexamcertification.com/what-is-waterfall-model-advantages-disadvantages-and-when-to-use-it/

[4]martinbauer. (2005). Successful Web Development Methodologies . Retrieved September 20, 2010, from www.martinbauer.com: http://www.martinbauer.com/Articles/Successful-Web-Development-Methodologies

[5]minds, O. (2008, Aug Saturday). Projects.pdf. Retrieved November Tuesday, 2014, from www.optimalminds.com: www.optimalminds.com/projects/J2EE%20Projects.pdf

[6]Search, T. (2011, September Monday). why-do-you-need-seo. Retrieved November Thursday, 2014, from http://thinksearch.co.uk: http://thinksearch.co.uk/seo-consultancy/why-do-you-need-seo/

[7]Springshare. (2012, February). literaturereview03.htm. Retrieved November sat, 2014, from http://www.library.ncat.edu: http://www.library.ncat.edu/ref/guides/literaturereview03.htm

[7]Waterfall model,14. (n.d.). Retrieved September 2014, 28, from http://istqbexamcertification.com/what-is-waterfall-model-advantages-disadvantages-and-when-to-use-it/

[8]Wu di,Luan Tian,Bai Yan,Wei Liyuan,Li Yanhui,"Study on SEO monitoring system based on keywords &links ",ICCSIT,2010 3rd IEEE International Conference, Vol 5, Pages:450-453.

[9]Fuxue wang, coll. Of Econ &Manage,.Chongqing Univ.of Posts & Telecommun.,Chongqing,China;Yi Li;Yiwen Zhang;"An empericial study on the search engine optimization technique and its Conference,Pages:2767-2770

[10]Chengling Zhao,Dept.of Inf.Technol.,Huazhong Normal Univ.,Wuhan,China;Jiaojiao Lu;Fengfeng Duan;"Application and Research of SEO in the development of Web2.0 Site" Knowledge Acquisition and Modeling,2009,KAM'09.Vol 1, Pages 236-238

[11]What is Search Engine Optimization http://en.wikipedia.org/wiki/Search_engine_optimi zation

[12] Jöran Beel and Bela Gipp. Google Scholar's Ranking Algorithm: The Impact of Citation Counts (An Empirical Study). In André Flory and Martine Collard, editors, Proceedings of the 3rd IEEE International Conference on Research Challenges in Information Science (RCIS'09), pages 439–446, Fez (Morocco), April 2009. IEEE. doi: 10.1109/RCIS.2009.5089308. ISBN 978-1-4244-2865-6. Available on http://www.sciplore.org.

[13] Jöran Beel and Bela Gipp. Google Scholar's Ranking Algorithm: An Introductory Overview. In Birger Larsen and Jacqueline Leta, editors, Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09), volume 1, pages 230–241, Rio de Janeiro (Brazil), July 2009. International Society for Scientometrics and Informetrics. ISSN 2175-1935. Available on http://www.sciplore.org.

[14] Jöran Beel and Bela Gipp. Google Scholar's Ranking Algorithm: The Impact of Articles' Age (An Empirical Study). In Shahram Latifi, editor, Proceedings of the 6th International Conference on Information Technology: New Generations (ITNG'09), pages 160–164, Las Vegas (USA), April 2009. IEEE. doi: 10.1109/ITNG.2009.317. ISBN 978-1424437702. Available on http://www.sciplore.org.

[15] Google. Google's Search Engine Optimization Starter Guide. PDF, November 2008. URL http://www.google.com/-webmasters/docs/search-engine-optimization-starter-guide.pdf.

[16] Albert Bifet and Carlos Castillo. An Analysis of Factors Used in Search Engine Ranking. In Proceedings of the 14th International World Wide Web Conference (WWW2005), First International Workshop on Adversarial Information Retrieval on the Web (AIRWeB'05), 2005. http://airweb.cse.lehigh.edu/2005/bifet.pdf.

[17] Michael P. Evans. Analysing Google rankings through search engine optimization data. Internet Research, 17 (1): 21–37, 2007. doi: 10.1108/10662240710730470.

[18] Jin Zhang and Alexandra Dimitroff. The impact of metadata implementation on webpage visibility in search engine results (Part II). Cross-Language Information Retrieval, 41 (3): 691–715, May 2005.

[19] Harold Davis. Search Engine Optimization. O'Reilly, 2006.

[20] Jennifer Grappone and Gradiva Couzin. Search Engine Optimization: An Hour a Day. John Wiley and Sons, 2nd edition, 2008

Manpreet Kaur received her Bachelor of Engineering in IT from NCIT israna panipat and Master of Technology in Computer Science & Engineering fom NCCE israna panipat.Her areas of interest are networking and cloud computing.

Hari Singh received his Bachelor of Engineering (with honours) and Master of Technology in Computer Science & Engineering. He has been pursuing Ph.D. in Computer Science from Thapar University, Patiala, India. His areas of interest are Distributed Computing, Cloud Computing, Programming and logic development. Hari Singh is an Associate Professor in the Department of Computer Science & Engineering, N.C. College of Engineering, Israna, Panipat (Haryana). He has 13 years of teaching experience in Computer Science & Engineering. He is a member of IEEE, ISTE, and CSI.