



Review Paper

## Search Engine Optimization: A Study

Patil Swati P.<sup>1</sup>, Pawar B.V.<sup>2</sup> and Patil Ajay S.<sup>2</sup>

<sup>1</sup>Department of Computer Science, S.S.V.P.S's Science College, Dhule, Maharashtra, INDIA

<sup>2</sup>Department of Computer Science, North Maharashtra University, Jalgaon, Maharashtra, INDIA

Available online at: [www.isca.in](http://www.isca.in)

Received 3<sup>rd</sup> November 2012, revised 24<sup>th</sup> December 2012, accepted 27<sup>th</sup> December 2012

### 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. So promoting a website in search engine result is a major task in website development. Search engine optimization (SEO) is to complete this work. But sometimes black hat SEO techniques are used which mislead the search engine and increase page ranking higher than deserved in search engine results. This paper present feature of search engine page rank algorithms, SEO techniques and black hat SEO techniques.

**Keywords:** Search Engine Optimization (SEO), Black Hat SEO, Page Rank.

### Introduction

Now a day, the enormous content of the Internet has made it difficult to find relevant information on a subject. Methods helping retrieving information have become particularly important<sup>1</sup>. So search engine becomes an integral part of everyone's life to search information. 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. If search engine fairly judge quality and relevance of every page and return high quality pages to user then "search-engine-bias" may not be a significant problem<sup>2</sup>. But unfortunately, quality of page is very subjective assumption and difficult to measure in real life. Major search engine like Google rely on page rank to measure quality of page<sup>3</sup>. Higher page rank value indicates website is very popular<sup>4</sup> In order to score a higher rank in search engine result many website promotion techniques are used by website designers. To promote website in search engine natural listing, search engine optimizers analyses the search engine results and according to that search engine optimization techniques are used by website designers. Search engine optimization (SEO) is the process which improves the volume and quality of traffic to a web site from search engines via natural search results for targeted keywords. Search engine optimizers use knowledge base. This is domain knowledge which evaluates interestingness patterns from search result<sup>5</sup> Search engine optimizers use search engine optimization techniques which follow search engine guidelines are called white hat SEO techniques. Sometimes search engine optimizers use website promotion techniques in web page development which does not follow the search engine rules and policies. Such techniques are called black hat SEO techniques.

This paper discusses the features of search engines: i. Page ranking algorithm, ii. White hat SEO techniques, iii. Black hat SEO techniques.

The goal is to provide reference for the developers of websites in their search engine optimization.

### 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<sup>6</sup>.

Suppose  $t_1, t_2, \dots, t_n$  are pages linking to page A then Page-A has its PR value as follows:

$$PR(A) = (1-d) + d \{ PR(t_1)/C(t_1) + PR(t_2)/C(t_2) + \dots + PR(t_n)/C(t_n) \} \quad (1)$$

Where  $d$  is damping coefficient, usually its value is 0.85.  $PR(t_1) \dots PR(t_n)$  is page  $t_1$  to  $t_n$  page rank value,  $C(t_i)$  means number of outgoing links page  $t_i$ .  $PR(t_i)/C(t_i)$  means page  $t_i$ 's contribution to page A's PR value<sup>2</sup>.

**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  $S_i$  of the score is computed by considering only key phrases that contain precisely  $k - i$  of the query terms.

$$S_i = \frac{\text{SUM}\{\text{key phrases } p \text{ with } k - i \text{ query terms}\}}{\text{LevelScore}(p) * \text{FullnessFactor}(p, q)} \quad (2)$$

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<sup>7</sup>:

$$\text{Expert\_Score} = 2^{32} * S_0 + 2^{16} * S_1 + S_2. \quad (3)$$

**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:

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

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<sup>8</sup>.

## Search Engine Optimization

**SEO Concept:** 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. Search engine optimizers help in building a website such that it can be found easily to search engine crawler with relevant keywords<sup>9</sup>. 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.

**White hat SEO:** White hat SEO techniques are ethical which follows search engine's rules and policies. White hat SEO

search engine ranking results in such a way that search engine don't punish the site like blocking the site from their search results. Using white hat SEO techniques search engine returns quality content. These techniques are beneficial to both users as well as search engines.

SEO include major two factors On-site optimization and Off-site optimization<sup>10</sup>.

**On-site Optimization techniques:** Keywords are short descriptions. 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 formatting, keyword in meta tag, keyword in title tag, position of keywords, external link, keyword density etc., which are controlled by site itself.

**Location of keyword:** Search engine crawler check whether keyword appear in <title> tag; <header> tag; <alt> tag; <meta> tag; <body> tag, in anchor text, in URL etc.

**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 pages<sup>11</sup>.

**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 ranking<sup>12</sup>.

**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 Optimization<sup>13</sup>.

**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.

**Keyword in alt text:** Alt text or tag specifies alternative text for images. Descriptive text associated with alt tag that serve same purpose and convey same essential information the image. Alt tags are short and descriptive which reflect the body text that describes the image.

**Keyword in anchor text:** Search keyword in anchor text represents what is linking. Pages using link text based on search keywords often rank higher.

**Title Length:** The most important on-page factor is appropriate use of keyword in title tag<sup>14</sup>. Website title should be such that it

reflects the subject of website. Using title name user understands brief information within website at first glance. Title length within limit returns good result<sup>12</sup>.

**URL (Uniform Resource Locator) Length:** URL represents the address of site on internet. Search keywords are included in URL so that crawler will find it easily. Short length URL's are preferable by search engines<sup>13</sup>.

**Outgoing Link:** Webpage contain links to other related websites. Related outgoing links provide useful information to user. More number of unique outbound links improves ranking of website.

**Off-site Optimization Techniques:** 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.

**Click Popularity:** More number of clicks to the site is known as click popularity. It is also significant factor to lift website top in ranking result. If visitor clicks websites, search engine provides certain value to that site. But Search Engine keeps track of who is clicking by tracking their IP address. So owner can't click to his site hundreds of times to improve click popularity as clicks from single IP address will be consider as only once.

**Inbound Link:** High quality external links pointing to a website are called inbound links. Total number of inbound links is called as link popularity. In Google, page rank of website determine according to quantity and quality inbound links. To promote site in top ranking list, the quantity and quality of external links are still recognized as the major ranking factor<sup>15</sup>. Web page must contain more number of relevant inbound links to rank high in search engine result. Inbound links in textual form and not in graphic form like banners, advertisements and images are not preferred by search engines.

**Black hat SEO:** To develop a website is a marketing strategy which is more effective and least costly to reach many people. Promoting a website in search engine result list is one of the key of creating a profit producing web site. Sometime to get higher ranking in search engine result listing, SEO techniques are used in an unethical manner called spamming. Such black hat SEO techniques break search engine's rules and regulations and place the undeserving site on top list. Such technique not only misleads the search engine algorithms but also lowers the quality of search results and increase traffic. Also these

techniques will not have any benefit to user. Some black hat SEO techniques used by search engine optimizers are

## Content Spamming

**Invisible text:** To raise keyword density SEO optimizers insert text in website which is unrelated to that website content. Inserted text includes the words which are popular or frequently search. Such unrelated text is invisible to user but visible to search engine. Spammers add the text to page with similar background colour or with white colour, with very small font, located in an area that is define as hidden or invisible through CSS, text placed behind an image, text located so far right, far left or very below.

**Keyword Stuffing:** Spammers repeat the keywords in various HTML tags like title, meta, body, anchor etc. Also keywords are stuffed by spammers in URL. Instead of consecutively repeating the key terms they are placed in between different sentences. Also spammers dump large number of unrelated keywords so that certain page becomes relevant to many different queries.

**Link Spamming:** Search engines like Google rely on quality and quantity of sites that link to a web site to determine its ranking.

**Link Farm:** Groups of heavily interconnected pages referred to as link farms. Search engine optimizers dumps hundreds of links to different sites within different categories that are unrelated to site content<sup>9</sup>. Using this they increase link popularity by including site into link exchange program.

**Link Exchange:** Spammers make a group in which their sites point to each other. In this way link count of each in a group increases and hence link popularity increases. One site contain link to other site and that other site have link to back to site. In this way link count is increased and so link popularity increases.

**Hiding Techniques:** Spammers hide the sentences, text and links so the users are not able to see them but search engine does.

**Link hiding:** To hide a hyperlink, spammers use small image or black image and the link redirecting from that invisible image. Also spammers uses link colour same as background colour to hide link from user.

**URL Redirection:** URL redirection means URL forwarding. Spammers hide the spam pages by redirecting the browser to another URL as soon as page is loaded. So in search engine index spam page is return but through redirection the target page is return to user.

**Doorway pages:** Doorway pages are primarily designed for search engines and not for human beings. Users who access doorway pages are redirected to fake scanning or video

streaming pages that then lead to different malware binaries. Before user reach the target page, series of redirection takes place which hide the actual URLs.

**Cloaking:** It is a technique by which spam web servers returns different web page to user and different web page to crawler. When user enters a query sites normal page is returned but when search engine request for same URL the page that has been created for search engine is returned and normal page is hidden from search engine. So such websites may lead to user to some other domain. Spammers implement cloaking with scripts that are not read by search engines<sup>16</sup>.

**Content Scraping:** Spammers copy the content from high ranking websites and paste that content in their website to boost their site's ranking position in search engine result. It is a violation of copyright law.

## Conclusion

This paper studies page ranking algorithms, search engine optimization techniques, black SEO techniques. Website ranking in search result is strongly depends on how SEO is implemented. White hat SEO techniques return quality content. These techniques give slow results but for long time. These are beneficial to both users as well as search engines. Black hat SEO techniques provide quick results but for short time and if search engine find out the unethical activities of the site, then site can also penalized. The goal of paper is provide awareness and stimulate further research in this area.

## Acknowledgements

This work was supported by a grant from UGC, WRO, Pune under Minor Research Project scheme (No. 47-948/09).

## References

1. Belsare S. and Patil S., Study and Evaluation of user's behaviour in e-commerce Using Data, *Research Journal of Recent Sciences*, 1(ISC-2011), 375-387 (2012)
2. Cho J. And Roy S., Impact of search engines on page popularity, Proc. 13<sup>th</sup> International conference on World Wide Web, 20-29 (2004)
3. Page L., Brin S., Motwani R. and Winograd T., The PageRank Citation ranking: Bringing order to the web, Technical Report, Standford Info. Lab (1999)
4. Knezeric B. Vidas-Bubanja M., Search Engine Marketing As Key Factor For Generating Quality Online Visitors, MIPRO, Proc. 33<sup>rd</sup> International Convention, 1193-1196 (2010)
5. Raorane A.A. and Kulkarni R.V., Association Rule – Extracting Knowledge Using Market Basket Analysis *Research Journal of Recent Sciences*, 1(2), 19-27 (2012)
6. Feifei X. and Guangnian Z., Design and implementation of a Java-based search engine algorithm analysis system, Proc. 4<sup>th</sup> International Conference on Computer Science and Education, 1040-1043 (2009)
7. Bharat K. and Mihailo G.A., Hilltop: A Search Engine: Based on Expert Documents, Technical Report, University of Toronto (1999)
8. Yunfeng M., A Study on Tactics for Corporate Website Development Aiming at Search Engine Optimization, Second International Workshop on Education Technology and Computer Science, 3, 673-675 (2010)
9. Somani A. and Suman U., Counter Measures against Evolving Search Engine Spamming Techniques Proc. of International Conference on Network and Computer Science, 6, 214-217 (2011)
10. Shi J., Cao Y. and Zhao X., Research on SEO Strategies of University Journal Websites, Proc. 2<sup>nd</sup> International Conference on ICISE, 3060-3063, (2010)
11. Gyongyi Z. and Garcia-Molina H., Web Spam Taxonomy, Proc. 1<sup>st</sup> International Workshop on Adversarial Information Retrieval on the Web, 12, (2005)
12. Wang F., Li Y. and Zhang Y., An Emphirical study on the Search Engine Optimization Technique and Its Outcomes" Proc. 2<sup>nd</sup> International Conference on AIMSEC, 2767-2770 (2011)
13. Zhu V., Wu G. and Yunfeg M., Research and Analysis of Search Engine Optimization Factors Based on Reverse Engineering, Proc.3rd International Conference on Multimedia Information Networking and Security, 225-228 (2011)
14. Kent P., Search Engine Optimization for Dummies, Wiley Publishing, 2, 67-68 (2003)
15. Kumar R. and Saini S., A Study on SEO Monitoring System Based on Corporate Website Development, *International Journal of Comp. Sci., Engg. and Infor. Tech.*, 1(2), (2011)
16. Wikipedia <http://en.wikipedia.org/wiki/Cloaking> (2012)