

**REVIEW ARTICLE**

Available Online at [www.jgrcs.info](http://www.jgrcs.info)

**A COMPREHENSIVE REVIEW ON SEARCH ENGINE OPTIMIZATION**

<sup>1</sup>Er.Tanveer Singh, <sup>2</sup>Dr.Raman Maini

<sup>1</sup>Research Scholar, Department of Computer Engineering  
University College of Engineering, Punjabi University, Patiala.

<sup>1</sup>yuvraj.singh90@yahoo.com

<sup>2</sup>Professor, Department of Computer Engineering  
University college of Engineering, Punjabi University, Patiala.

<sup>2</sup>research\_raman@yahoo.com

**Abstract:** Search Engine Optimization (SEO) has become an important part of Search engine marketing. Trends, contents, importance and working of SEO and comparison of various page ranking algorithms have been reviewed in this paper. Humanized Ranking, Mobile search which are the biggest search medium are also discussed. Due to the increase in number of web sites, web page links, there is need to properly rank the web pages from most relevant to least relevant web site links according to the need of the user. Need of the hour is a web marketing technique known as search engine optimization.

**Keywords:** search engine; page rank; optimization; SEO trends; SEM; web crawler.

**INTRODUCTION**

Search Engine Marketing (SEM) involves paid search and organic search in its strategies [4].

SEO is related to SEM that is driving the web traffic to the websites. The key components of search engine optimization are web pages [3].

Some of the search engines are Google, msn, yahoo, AltaVista, Aol.com, All the web, web crawler, free serve etc. The general procedure used by them is –

- i. Finding Websites (Discovery).
- ii. Storage of Links.
- iii. Ranking (Page Rank).
- iv. Display of the search results [3].

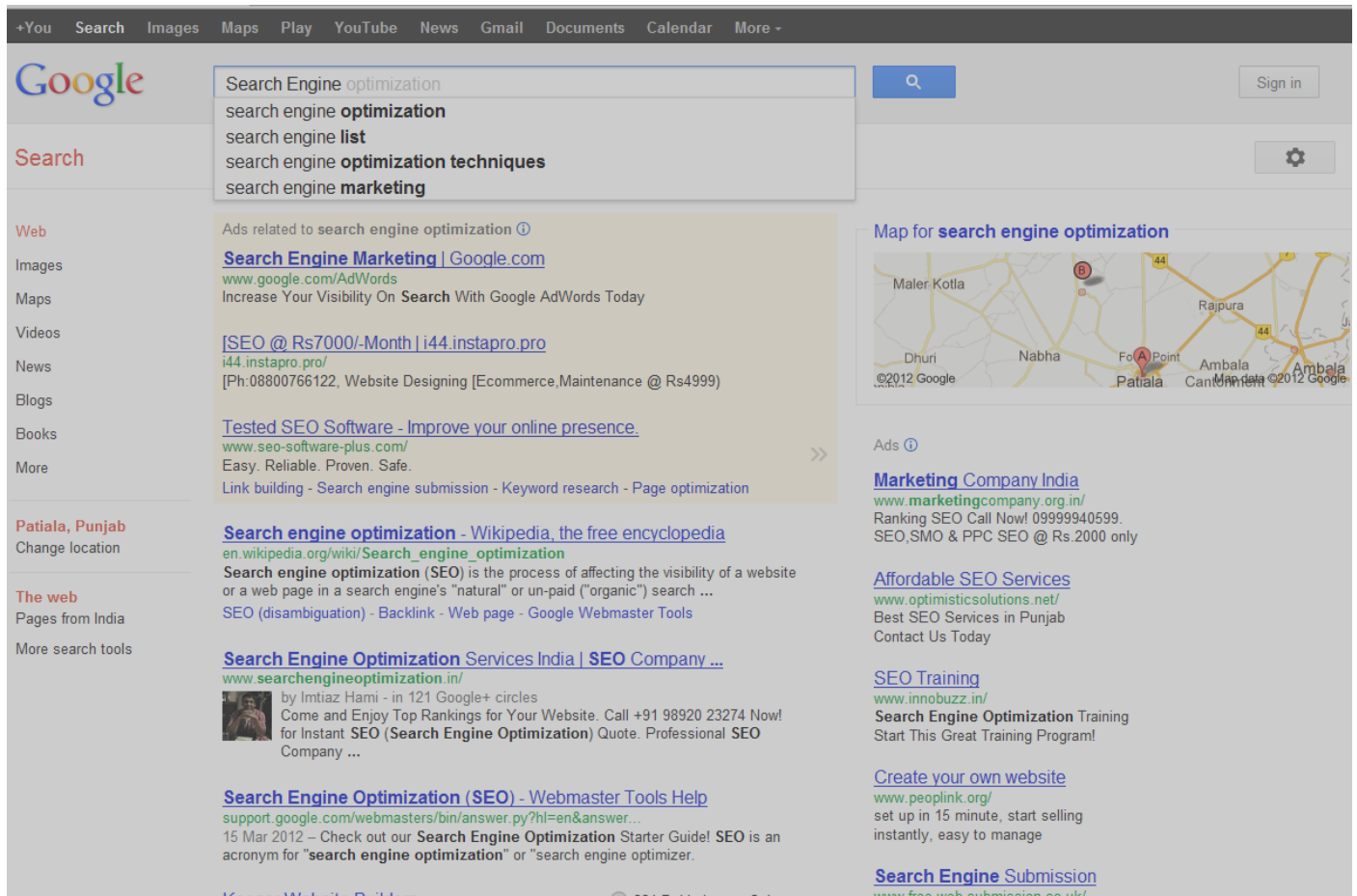


Figure 1- Organic and Paid Search [4]

The search engine optimization helps the user for retrieving better results related to their query, if users are searching through search engines. Web search engines are useful as user query is resolved by matching the results from the search engine's database [8]. Some of techniques used by SEO are Black Hat Technique, White Hat Technique etc [10].

## SEO TECHNIQUES

Two techniques used by SEO are

### **Black Hat Technique:**

### **White Hat Technique:**

### **Black Hat Technique:**

Black Hat Technique does not follow search engine guidelines. In black hat technique ranking process is fast. Hidden data is not visible to the user. This technique focuses on search engine [11]. Black hat techniques are as follows -

### **Hidden Text:**

The hidden text is not visible to the user but is readable by the search engine and identified as search spam [10].

### **Link Farms:**

Link farming is the technique for spamming the index of search engines. These do not provide traffic to websites and also have risk of banning the websites [10].

### **Keyword Stuffing:**

Whenever a web page is loaded with the keywords in the Meta tag or in content; keyword stuffing occurs. Due to repetition of words in Meta tag, a search engine no longer uses these tags. [10].

### **Scraping:**

In this technique the spammer copies the content from those websites which are popular with search engine. To copy that content to its own web site scraping technique is used [10].

### **Paid Links:**

There are paid links on web sites for increasing the link popularity and ranking of the web pages. This technique is easier for the user with little knowledge of SEO. There is less traffic on paid links [10].

### **Door way pages:**

Door way pages are fake pages designed for the search engines. For spamming the index of search engine and redirecting the users to different web pages this technique is used. Door way pages are also called as gateway pages or entry pages [10].

### **Parasite Hosting:**

It is the process of hosting the web sites on any other server without any consent just for the purpose of search engine benefit [10].

### **Cloaking:**

Cloaking is the technique to display the content of web pages differently from that of assigned content to search engines. The content delivered to search engines is not displayed to the user [10].

### **White Hat Technique:**

White Hat technique follows search engine guidelines. This technique takes much time to perform ranking process. Data is visible to the user. This technique provides better search engine ranking [11]. White hat techniques are as follows

### **Quality Content:**

The web sites containing well written content are more valuable to search engines, visitors and other users. These users can interlink their own web sites with such good content websites. Creating good content is very time consuming. But it is useful effort in the long run [10].

### **Link Baiting:**

This technique involves creating the content that encourages other users to interlink their web pages with your website. This content may contain useful resources, articles or some new stories [10].

### **Internal Linking:**

Many web sites are using some form of scripts for displaying the drop down list navigation or popup menus. This type of navigation or linked pages are not crawled by the web spider of search engines, not indexed and stored in the database [10].

### **Use Structural (Semantic) mark up:**

In this technique the heading elements are essential. Headings are assigned to the content to make web site content meaningful. With headings, it is easier for the search engines and the users to find what content they are looking for [11].

### **Titles:**

Assigning titles to web pages is must. Title of a web page accurately describes the page content. Unique and accurate page titles should be created. Titles should be brief but descriptive. It must be relevant to the content of the web page [11].

### **Use of Keywords:**

Single words are not most effective; multi words must be used to make users like to visit the web site. Keywords must be used throughout the web site. Two to three keywords can be assigned to each web page [11].

### **Inbound Links:**

The links from other web pages are inbound links. There are good links and bad links. Good links are those which are ranked highly by the search engines and are relevant to the users search. Bad links are those which are banned by the search engines and are not relevant to the content of user's search [11].

### **Site Optimization:**

Changing of the content, web site structure for achieving the high web page positioning is the backbone of SEO. Creating unique, accurate titles and Meta tags, to improve the content is the key to any successful optimization effort [11].

### **Page Ranking in SEO:**

Whenever user searches anything on the web via search engine then web pages are retrieved and displayed on the user screen according to their page rank. The important web

pages or web sites are ranked higher and are normally visited first by the user [5].

There is an anchor tag with which the hyperlinks from one web page to another are given. Search Engine decides the rank of these web pages which are interlinked by the numbers assigned to these web pages. If one web page or web site is more popular to prospective customers then another web site link can be added to the popular website to get some favour of it [1]. If the web pages are interlinked then all these web pages are crawled, ranked and stored in the database respectively.

There are off page ranking factors and on page ranking factors. On Page factors are title tags, header tags, alt image tags, content, keyword frequency etc. The off page factors are link popularity, anchor text etc [8].

**SEO TRENDS**



Figure 2 – SEO Trends

**Humanized Ranking:**

According to Panda and Penguin updates SEO trend ‘humanized ranking’, search engines are much dependent upon humans for page ranking purpose. This way, humans act as the quality testers for search engines. Google is already using such factors for more real time search [9]. Second aspect is of social networks used by humans which contribute to improve ranking of web pages.

**Quality and not Quantity:**

This trend in SEO is known for the most relevant and better results in response to the user query. Quality trend of the SEO aims at freshness of the content; quality of the content must be good enough to satisfy the user’s search via the search engine [9]. Example – In Google, panda2 focuses on the quality and the freshness of the content.

**Voice Search:**

According to Panda and Penguin updates the ‘voice search’ is more useful for fast searching on the World Wide Web.

Instead of browsing on the web, humans are looking for the direct answers in response to their query [9]. Instead of writing their query in the search box they can find results relevant to their question with the help of voice search trend used in search engine optimization (SEO).

**Mobile Search:**

In SEO 2012 one can expect of huge features in the mobile search market. Mobile search can be done while talking to your phone, or using the social networking websites, or by tapping a search into your phone. It may be the biggest search medium [9].

**Title Tag and Headings:**

The title tags are very important. A website developer should be able to define the given topic or with which the website is related to in a manner that it seems to the user to be a useful link. Title link excites the user and makes the user to click on that particular link of website.

The search engines give importance to the heading tags. Heading of the topic is relevant to its description [9]. It makes some logical sense for an article to split up into different parts.

**Personalized Search:**

The SEO trend ‘Personalized Search’ has been updated by post Panda and Penguin. More changes in rankings can be expected from the personalized search. Example - Using social networking, more changes in ranking and personalized search have occurred. [9].

**Social Media:**

In 2012 the search engine optimizations became more of the social media optimizers (SMO’s) [9]. In the social media instead of just focusing on the page factors of the search engine optimization, it also focuses on the social integration.

**Link Building Techniques:**

The authority of the web pages is measured by the web pages stored in search engines’ database and further through the links from these web pages. Example - Search engine Google uses Page Rank algorithm to rank the web pages and decide the position of the web pages which increases their importance [9].

**CONTENTS OF SEO**

The search engine optimization focuses on search engine marketing that helps website rank higher in the natural results or paid search [3]. A search engine contains the crawler or spider that downloads the web pages. For the storage of downloaded and processed pages, database is used. The result engine is used to extract the web pages from the database. The web server is responsible for the interaction between the user and search engine components [8].

The results are retrieved and displayed to the user by sorting from most relevant to least relevant web sites [8].

**IMPORTANCE OF SEO**

The Search Engine Optimization is much important for making the web sites popular by bringing the relevant websites in notice of the prospective customers [8].

SEO is the most effective website marketing tool. Example - Google is the search engine used to make searching easier for the user and maintaining the websites rank higher by displaying the links of most relevant web pages at the top of the list. Usually normal user accesses the web pages that are shown at the top of the list. The Google search result position is shown in a figure 3.

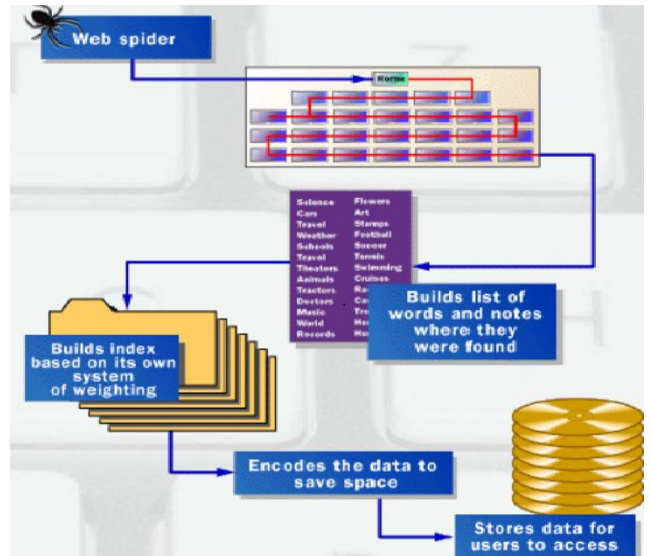


Figure 4 - Working of SEO [2]

The result engine extracts the results from the database and displays it to the user as most relevant web pages to the user query.

When user types something in the search engine box the search engine processes the user request by matching the user query with the results stored in the database. The results are stored in the database in the form of web pages. These web pages are ranked on the basis of content of the web page, relevant keywords used in the web pages, the frequency of occurring of keywords in the web page [2]. If the title, description, content of the web page is more relevant and important then web pages are listed at the top. Description of website is given as the summary. The title or description of the web sites should appear to user as the useful link [8]. The users normally visit or attempt to click on the web pages that are given at the top.

The web pages are ranked on basis of the numbers assigned to these web pages. The web pages are stored in the database and retrieved with help of result engine. Results are displayed on the user screen in the form of a list from the most relevant to least relevant web sites[3].



Figure 3 - Google search result position chart [9]

**WORKING OF SEO**

The web spider crawl the web pages and builds a list of words location then indexes them and stores the data in database for user access [2].

**COMPARISON OF PAGE RANKING ALGORITHMS**

Table 1(a) – Comparison of Page Ranking Algorithms [6]-[7].

S. No.	Parameters	Algorithms		
		Page Rank Algorithm	Weighted Page Rank algorithm	HITS Algorithm
1.	Main Web Mining technique	This algorithm uses web structure mining (WSM) technique.	It also uses web structure mining (WSM) technique.	It uses both web structure mining (WSM) and web content mining (WCM).
2.	Working Methodology	The scores are computed at index time and the results are arranged with respect to pages important for the user.	According to page importance the results are arranged and the scores are computed at index time.	This algorithm computes the scores of the pages which are highly relevant to the query of the user.
3.	Input Parameters of algorithm	Back links are used as input parameters in this algorithm.	Forward links and back links are used as input parameters.	Content, Forward links and back links are used.
4.	Complexity of algorithm	The complexity of this algorithm is $O(\log N)$ .	Weighted page rank algorithm complexity is $<O(\log N)$ .	Complexity of HITS algorithm is $<O(\log N)$ .
5.	Weakness	This algorithm is query independent.	It is also query independent.	There is a problem of efficiency and topic drift.
6.	Search Engine	Page rank algorithm is implemented in Google search engine.	This algorithm is implemented in Research Model.	IBM prototype (Clever) uses HITS algorithm.

7.	<b>Relevance</b>	It is less relevant because it ranks the web pages at indexed time.	It is also less relevant as the calculation of web pages weight is done at the indexed time.	It is more relevant as it considers content in its input parameters.
8.	<b>Formulae</b>	$PR(A) = (1-d) + d(PR(T1) / C(T1) + \dots + PR(Tn) / C(Tn))$ Where d- Damping Factor can be set between 0 and 1. PR(A) – Page rank of page A. PR(T1) – Page rank of pages T1 which link to page A. C(T1) – Number of outbound links on page T1.	$W^{in}_{(m,n)} = \frac{In}{\sum I_p}$ $W^{out}_{(m,n)} = \frac{On}{\sum O_p}$ Where In, Ip – Number of incoming links of page n and p. R(m) – Reference page list of page m. W <sup>out</sup> <sub>(m,n)</sub> - Weight of link (m, n). On, Op- Number of outgoing links of page n and p. $WPR(n) = (1-d) + d \sum_{m \in B(n)} W^{in}_{(m,n)} W^{out}_{(m,n)}$	$H_p = \sum_{q \in I(p)} A_q$ $A_p = \sum_{q \in B(p)} H_q$ Where Hp – Hub weight Ap – Authority weight Ip, Bp – Set of reference and referrer pages of page p.
9.	<b>Results Quality</b>	Quality of results produced by this algorithm is medium.	High quality results are produced than page rank algorithm.	Low quality results are produced than page rank algorithm.
10.	<b>Importance of algorithm</b>	Much Important as back links are considered as input parameters.	Ordering or Sorting of pages is done according to the importance.	Hub and authorities scores are used, so importance is moderate.

Cont... Table 1(b) – Comparison of Page Ranking Algorithms [6]-[7].

S. No.	Parameters	Algorithms	
		Distance Rank Algorithm	Dirichlet Rank Algorithm
1.	<b>Main Web Mining Technique</b>	This algorithm uses web structure mining technique.	It also uses web structure mining (WSM) technique.
2.	<b>Working Methodology</b>	The minimum average distance between the pages is calculated to compute the scores.	Working of this algorithm is same as page rank but the transition probabilities are computed using Bayesian estimation.
3.	<b>Input Parameters of Algorithm</b>	Back links are the input parameters of distance rank algorithm.	The back links are input parameters.
4.	<b>Complexity of Algorithm</b>	Complexity of distance rank algorithm is O (log N).	Dirichlet algorithm's complexity is O (log N).
5.	<b>Weakness</b>	This algorithm needs to work along with the page rank.	Dirichlet rank algorithm needs to work along with page rank.
6.	<b>Search Engine</b>	Distance rank algorithm is implemented in Research Model.	Implementation of dirichlet rank algorithm done in Research Model.
7.	<b>Relevance</b>	It is more relevant as it calculates the minimum average distance.	It is more relevant than page rank algorithm as it calculates the transition probabilities using Bayesian estimation.
8.	<b>Formulae</b>	$d_b = \frac{\sum_{a=1}^v d_{ab}}{V}$ Average distance for page b $d_b = d_a + \log(O(a))$ Distance Rank of Page b $d_b = \min(d_a + \log(O(a)), a \square \square b)$ Where d <sub>b</sub> - Average distance for page b V – Number of web pages d <sub>ab</sub> - Distance between pages a and b.	$\omega(n) = \frac{\mu}{n + \mu}$ where $0 \leq n \leq \infty$ n - number of out links μ - dirichlet parameter d <sub>o</sub> , d <sub>s</sub> - Dirichlet rank scores $do(T) = \frac{\sigma + \tau}{N}$ $ds(T) = \left( \frac{1 + \frac{k}{\mu^2 + (K+1)\mu}}{\frac{\sigma + K + \mu + 1}{\mu + 1} \frac{\tau}{N}} \right)$
9.	<b>Results Quality</b>	Better quality results are produced than page rank algorithm.	More effective results are produced using this algorithm.
10.	<b>Importance of algorithm</b>	Much important as back links act as the input parameters.	Stability is the important factor to make dirichlet rank algorithm more reliable.

**CONCLUSION**

Search engine optimization (SEO) has evolved from search engine marketing (SEM). Without the help of search engine (to search anything) the user will have to remember almost

all URLs exactly. It is very difficult and searching will become for the user. To accomplish this task and to provide the best results to the prospective customers (users) the search engine optimization is done. SEO uses page ranking algorithms for providing best relevant results to the users for every search query. Page ranking is done by a powerful web

marketing technique known as Search Engine Optimization (SEO). SEO uses techniques such as black hat technique, white hat technique. White Hat technique is more effective as it follows the search engine guidelines. Data is visible while using white hat technique as it focuses both on search engine and the searcher both.

SEO is the effective website marketing tool. To make web sites popular among the users, the web site can be synchronized with the social networking web sites and by interlinking your web site with popular web sites. The rank of those web sites will rise and will be shown at the top of the list when user makes a search. The complexity of page ranking algorithms are  $O(\log N)$ ,  $<O(\log N)$ . Ranking is based on various factors -on page factors and off page factors. On page factors are title tags, header tags, alt image tags, content, hyperlinks, and keyword frequency. The off page factors are link popularity and anchor tags. SEO is a technique used as a part of SEM. Website without SEO is just like a shop without salesman.

## REFERENCES

[1] Hair Christopher, *Search Engine Optimization*, University of Teesside school of computing middles borough tees valley TS13BA, April 2008.

- [2] Buha Yuriy, *Search Engine optimization*, CMP 220 section1, professor Kelly, October 2010.
- [3] Davis Horald, Search Engine optimization, *Building Traffic and making money with SEO*, Oreilly Media, 2006.
- [4] Google, *Search Engine Optimization Starter Guide*, Google trademark, 2010.
- [5] Callen Brad, *Search Engine Optimization Made Easy*, Bryxen Softwares.
- [6] Kumar Singh Ashutosh, Kumar P Ravi, 'A Comparative Study of Page Ranking Algorithms for Information Retrieval', 2009.
- [7] Sharma Kumar Dilip, 'Comparison of various web page ranking algorithms', UCSE, International journal on computer science and engineering, 2010.
- [8] Introduction to search engine optimization 'Getting started with SEO to achieve business goals', Hubspot.

## Web References

- [9] <http://visual.ly/important-seo-trends-2012-2013>.
- [10] <http://cognitiveseo.com/blog/229/black-hat-vs-white-hat-seo-infographic/>
- [11] <http://www.pushon.co.uk/articles/top-5-white-hat-and-black-hat-search-optimisation-techniques/>