A Review Paper On Web Page Ranking Algorithms

Seema Rani[1] , Upasana Garg[2]

[1]Guru Kashi University, Department of CSE,
Talwandi Sabo, Punjab, India
gupta.seema720@gmail.com

[2]Guru Kashi University, Department of CSE,
Talwandi Sabo, Punjab, India
Upasna.garg44@gmail.com

Abstract: *The World wide Web consists of millions of interconnected web pages that provide information to the user present in any part of the world. The world wide web is expanding and growing in size and the complexity of the web pages. That is why it is necessary to retrieve the best or the web pages that are more relevant in terms of information for the query entered by the user in search engine. In recent years, semantic search for relevant documents on web has been an important topic of research. Many semantic web search engines have been developed like Ontolook, Swoogle, etc that helps in searching meaningful documents presented on semantic web. To relate entities/texts/documents having same meaning, semantic similarity approach is used based on matching of the keywords which are extracted from the documents. In this paper we have a ranking scheme for the semantic web documents by finding the semantic similarity between the documents and the query which is specified by the user.*

Keywords: Page Rank, Modified HITS, Semantic Similarity, Optimization.

1. Introduction

Search engine optimization is technique that refers to the process of improving traffic to a given website by increasing the visibility of site in search engine results.

SEO can be defined as strategy used to promote website in order to have a high rank i.e. top result. SEO activity helps to increase the amount of visitors to a Web site by ranking high in the search results of a search engine. The higher a Web site rank in the results of a search, greater the chance that site will be visited by a user.

Along with the fast growth of network information, using search engines to search information has developed into an important part of one's life every day. Search engine optimization is not simple. It is more about strategy, method and structure. It can be also considered as a form of engineering. SEO services will increase website sales, traffic, and exposure of a website.

Search engines explore websites contents to gather information about a website. Therefore, there is a need to optimize a website to make it search engine friendly. This will help to bring a website in top search results. The search engine programs are called "crawlers" or "spiders".

Crawler: Crawler used for retrieving the web pages and web contents A Web crawler starts with a list of URLs to visit, called the seeds. As the crawler visits these URLs, it identifies all the hyperlinks in the page and adds them to the list of URLs

to visit, called the crawl frontier**.** Search engines crawl the Web to see what is there.

Spiders: Most Web pages contain links to other pages, a spider can start almost anywhere. As soon as it sees a link to another page, it goes off and fetches it. A spider visits Web sites and reads their pages and other information in order to create entries for a search engine index.

Techniques used in search engine optimization:

There are two types of techniques that can help to optimize the website to increase traffic.

- **On-page SEO technique:** on page optimization includes those techniques that can be done on the pages of a website. On page optimization techniques help the search engine crawlers read the website content. A readable site helps to show quality and will result in higher ranked web pages.

- **Off-page SEO technique:** Off page optimization includes those techniques that can be done off the pages to increase traffic to your website.

The need for SEO has become heightened in the past few years. If the content of your website is ignored by the search engines then they can't index your site, **if your site isn't indexed then no one will find your site** by searching for it in Google, Yahoo, MSN, etc. If no one can find your site then your hard work and money spent creating the site is in vain. So SEO is essential because a small investment can bring big reward.
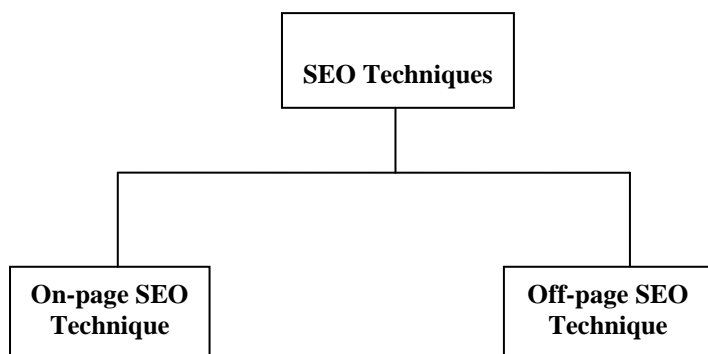
Fig 1: SEO Techniques

There are advantages of SEO:
SEO helps for increasing web-site traffic that will leads to more customer
Increase brand name or product visibility and also raise the business services
Increase in sales and popularity
Organic listings are essentially free. When you are listed at the top, you don't need to pay per click or allocate a budget for advertising.

LITERATURE REVIEW:

Poonam Chahal, Manjeet Singh, Suresh Kumar, "Ranking of Web Documents using Semantic Similarity"
In this paper we have proposed a ranking scheme for the semantic web documents by finding the semantic similarity between the documents and the query which is specified by the user. The novel approach proposed in this paper not only relies on the syntactic structure of the document but also considers the semantic structure of the document and the query. The approach used here includes the lexical as well as the conceptual matching. The combined use of conceptual, linguistic and ontology based matching has significantly improved the performance of the proposed ranking scheme. We explore all relevant relations between the keywords exploring the user's intention and then calculate the fraction of these relations on each web page to determine their relevance with respect to the query provided by the user. We have found that this semantic similarity based ranking scheme gives much better results than those by the
prevailing methods.

Parveen Rani and Er. Sukhpreet Singh, "An Offline SEO (Search Engine Optimization) Based Algorithm to Calculate Web Page Rank According to Different Parameters"
This paper describes the new algorithm for calculating web page rank according to different parameters. The proposed algorithm called M-HITS (Modified HITS) is a new version of HITS algorithm. It is developed by extending the properties of HITS algorithm. Author present new algorithm in which six parameters are used to evaluate rank for web page. Future work can be done by using some AI techniques in addition to these proposed techniques to improve the rank of web pages.

Gyanendra Kumar, Neelam Duhan and A. K. Sharma "Page Ranking Based on Number of Visits of Links of Web Page"
In this paper author presented a modified page ranking algorithms which is more target oriented than original page rank. The modified algorithm calculates page rank value or importance of web pages based on the visits of incoming links on a page. The paper presented a novel page ranking algorithm called VOL that provides more relevant results than original Page Rank. As a result, Author proved that VOL is far dynamic than original Page Rank algorithm and also observed that the page which has more visits of incoming links is carrying more rank value than less visited pages. The paper also presents a method to find link-visit counts of Web pages and a comparison between VOL with the Page Rank algorithm.

Ashish Jain, Rajeev Sharma, Gireesh Dixit and Varsha Tomar "Page Ranking Algorithms in Web Mining, Limitations of Existing methods and a New Method for Indexing Web Pages"
This paper proposed a new method called Intelligent Search Method (ISM). Author developed new method to index the web pages using an intelligent search strategy in which meaning of the search query is interpreted and then indexed the web pages based on the interpretation. This Paper also described the limitations of existing methods and discussed the different algorithms used for link analysis like Page Rank (PR), Weighted Page Rank (WPR), Hyperlink-Induced Topic Search (HITS) and CLEVER algorithm. The new method can be integrated with any of the Page Ranking Algorithms to produce better and relevant search results.

N. Batra, A. Kumar, Dr. Dheerendra Singh and Dr. R.N. Rajotia "Content Based Hidden web Ranking Algorithm (CHWRA)"
This paper proposed the algorithm called CONTENT BASED HIDDEN WEB RANKING ALGORITHM (CHWRA). The Proposed ranking algorithm consists of four different attributes. This method tries to cover all the aspects which directly or indirectly affect the popularity of the web page. This method is an effort to generate an ordered Hidden web searched result set. The ranking algorithm (CHWRA) gave the desired result. Future work can be done based on "Access Time length" parameter. This parameter can be used as "Length of the time spent on a web page" i.e. user's access time-length on the web page.

2. Existing Techniques :

Following are some most commonly used search engine optimization techniques.
Page Rank Algorithm: Page Ranking Algorithm is Surgey Brin and Larry Page. Page ranking algorithm is based on link structure of web. Page ranking algorithm is used by the popular search engines to rank the web page. It basically works on link structure on web pages. This algorithm counts the no. of links in web page and estimate the importance of web site. According to Page Rank Algorithm, The web site will be more important if it have more no. of links. The algorithm assigns a numerical weight to each element of World Wide Web so it referred to as page rank of E and it is denoted by PR (E).

HITS Algorithm: HITS Algorithm stands for Hyperlink-Induced Topic Search Algorithm. This algorithm is developed

by Jon Kleinberg. It is also called Link Analysis algorithm. HITS algorithm is used to rank the web page focusing on *Hubs and Authority*. When user issue some search query HITS algorithm expands list of web pages returned by search engines in response to user search query. Hits algorithm is query dependent. The famous Twitter web site uses the HITS style algorithm.

HUB: Hub represent the page that point to the authorities.
Authority: Authority represent as a source of valuable information.

Steps of HITS Algorithm:
Step 1: Enter the adjacency matrix of the web pages.

Step 2: Then enter the frequency of different types of parameters (Bold, Italic, Keyword, and No. of Unique Click).

Step 3: Compute the Hubs and Authorities for each web page.

Step 4: Normalize all these values for every web page and then compute the partial rank for every web page.

Step 5: And then add weights of the parameters to the calculated partial rank.

Step 6: Sort the web pages positions according to the calculated ranks corresponding to both Hub Values and Authority Values of web pages.

Step 7: Exit.

Semantic Similarity: Semantic similarity concept used in many fields. When the data or documents have same meaning then semantic similarity approach is used based on identical keywords which are extracted from different documents.

Steps of semantic similarity Algorithm:

Step1: Firstly construct a Text-List (by links).

Step 2: Then acquire query as a text: a String.

Step 3: For each Text in Text-List do:
(a) Create Text-Vector-Space.
(b) Create Domain-Dictionary of words.
(c) Using Statistical-Model () and Domain-Dictionary,
Compute relevance-value of Text corresponding to user Query.
(d) Make Domain-Ontology of the Text.
(e) Compute Domain-Similarity of Text value with
Domain- Ontology.
(f) Verify the maximum of Domain-Similarity value and relevance-value and call it Relevance-Score.
Step 4: Then Go to step 3 until there are no text left in the Text-List or else no more text is to be considered.
Step 5: Organize the links according to the decreasing order of relevance-score and assign the rank to them.
Step 6: And then finally display the contents according to their ranks.

Conclusion and Future Scope

In this paper we present the review on various techniques to rank a web page. Algorithms of the corresponding techniques are also discussed in their corresponding steps. From the existing techniques it can be concluded that algorithm should be based on AI that can rank the web pages in a particular domain based on user requirement or user search history. Hence this type of algorithm is need to be developed.

References

[1] P. Chahal, M. Singh and S. Kumar "Ranking of Web Documents using Semantic Similarity" Information Systems and Computer Networks (ISCON), 2013 International Conference on 2013 IEEE, DOI 10.1109/ICISCON.2013.6524191 Page(s): 145 - 150

[2] G. Kumar, N. Duhan and A. K. Sharma "Page Ranking Based on Number of Visits of Links of Web Page" Computer and Communication Technology (ICCCT), 2011 2nd International Conference on 2011 IEEE, DOI 10.1109/ICCCT.2011.6075206 Page(s): 11 – 14

[3] P. Rani and Er. S. Singh, "An Offline SEO (Search Engine Optimization) Based Algorithm to Calculate Web Page Rank According to Different Parameters" INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY Vol. 9, No 1 J u l y 1 5, 2 0 1 3

[4] A. Jain, R. Sharma, G. Dixit and V. Tomar, "Page Ranking Algorithms in Web Mining, Limitations of Existing methods and a New Method for Indexing Web Pages" Communication Systems and Network Technologies (CSNT), 2013 International Conference on, 2013 IEEE, DOI 10.1109/CSNT.2013.137 Page(s): 640 – 645

[5] N. Batra, A. Kumar, Dr. D. Singh and Dr. R.N. Rajotia "Content Based Hidden web Ranking Algorithm(CHWRA)" Advance Computing Conference (IACC), 2014 IEEE International, 2014 IEEE, DOI 10.1109/IAdCC.2014.6779390 Page(s): 586 - 589

[6] H. Dubey, Prof. B. N. Roy "An Improved Page Rank Algorithm based on Optimized Normalization Technique" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (5), 2011, Pages(s): 2183-2188

[7] R. Kumar and S. Saini "A Study on SEO Monitoring System Based on Corporate Website Development" International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.1, No.2, June 2011

[8] K. ur Rehman and M. N. Ahmed Khan " The Foremost Guidelines for Achieving Higher Ranking in Search Results through Search Engine Optimization" International Journal of Advanced Science and Technology Vol. 52, March, 2013

[9] N. V. Pardakhe, Prof. R. R. Keole "Analysis of Various Web Page Ranking Algorithms in Web Structure Mining" International Journal of Advanced Research in Computer and Communication Engineering Vol.2, Issue 12, December 2013

[10] M. Cui, S. Hu "Search Engine Optimization Research for Website Promotion" Information Technology, Computer Engineering and Management Sciences (ICM), 2011

International Conference on Vol.4, 2011 IEEE, DOI 10.1109/ICM.2011.308 Page(s): 100 - 103

Author Profile

Seema Rani received the B.Tech degree in Computer Engineering from Punjab technical University Jalandhar in 2012. Currently She is pursuing M.Tech degree in Computer Science from Guru Kashi University, Talwandi Sabo, Bathinda (Punjab). Her research interests include Internet Technology And Web Mining.