Search Engine Optimization with Efficient Page Ranking Algorithm

Miss. Gayatri Vivekrao Kpase

Dr. V.M.Thakre

Abstract- **PageRank is the algorithm used by Google search to rank websites in their search engine result. PageRank was named after Larry Page. PageRank is a way of measuring the importance of website pages. This paper deals with analysis and comparison of web pageranking algorithms based on various parameter to find out their advantages and disadvantages for the ranking of the web pages. It includes use of PageRanking for SEO. How SEO techniques increase website visibility. Developing efficient model for PageRanking in which web mining approach for distributed randomized PageRank algorithm are used.**

Key Words- **hyperlink, Page Rank, search engines, Web page**

I. INTRODUCTION

Web is expanding rapidly day by day and people generally rely on search engine to search on the web. In such a scenario it is the responsibility of service provider to fulfill the need of user with providing proper, relevant and quality information against their query submitted to the search engine. It is a responsibility of them to provide accurate, required and useful information to the internet user by using the web page contents and hyperlink between the web pages.

PageRank is the algorithm used by Google search to rank websites in their search engine result. PageRank was named after Larry Page. PageRank is a way of measuring the importance of website pages. Based on search engine named Google PageRank counts the number and quality of links to a page to get the idea of calculating the importance of the website. It is hypothesis that most important websites are gets links from another websites. This paper deals with analysis and comparison of web PageRanking algorithms based on various parameters to find out their advantages and limitations for the ranking of the web pages. Studying and analyzing the different web PageRanking algorithms, a comparison is done to find out their relative strengths and limitations to find out the further scope of research in web PageRanking algorithm. This Paper aim to answer what contributes to search engine rankings? And what can web content creators and webmasters do to make their content and sites easier to find by audiences using search engines?

II. PREVIOUS WORK DONE

Author Hideaki Ishii et al [1] Assume that the pages are divided into a number of groups, based on the hosts or the domains of the pages and interact among groups, condition that all groups have only limited ratios of

outgoing links towards another groups. Author Athanasios Papagelis el al [2] had worked on the bottom-up approach, this approach has been characterized into a hybrid bottom-up search engine that produces search results based on user provided web-related data and their sharing among users. Author Dilip Kumar Sharma, et al [3] had worked to find out relative strengths and limitations of various PageRanking algorithms by comparisons and to find out the future scope of research in web PageRanking algorithm. Author John B. Killoran, et al [4] had worked to answer what contributes to search engine rankings? And what can web content creators and webmasters do to make their content and sites easier to find by audiences using search engines? Author Chen Gong, et al [5] had worked to improve adaptability of ranking and formulates the tracking process as a ranking problem.

III. EXISTING METHODOLOGY

A) Bottom up Approach

Most explanations of the user behavior while interacting with the web are based on a top-down approach. This type of approach starts from the user and incrementally builds the dynamics of the web as the result of the users' interaction with it. A bottom-up approach is proposed to study the web dynamics based on web-related data browsed, collected, tagged, and semi-organized by end users. This approach has been materialized into a hybrid bottom-up search engine that produces search results based solely on user provided web-related data and their sharing among users.

B) PageRank Algorithm

PageRank algorithm is the most commonly used algorithm for ranking the different pages. Working of the PageRank algorithm depends upon link structure of the web pages. The PageRank algorithm is based on the concepts that if a page contains important links towards it then the links of this page towards the other page are also to be supposed as important pages. In PageRank the rank score is decided on basis of the back link. If the addition of the all the ranks of the back links are large then the page then it is provided a large rank .A simplified version of PageRank is given by:

$$\sum_{v \in Bu} PR(v)/L(v) \qquad (1)$$

Where the PageRank value for a web page u is dependent on the PageRank values for each web page v out of the set Bu divided by the number L (v) of links from page v. An example of back link is shown in figure 1 below. U is the back link of V & W and V & W are the back links of X.
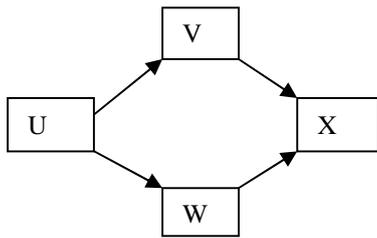
Fig 1: Illustration of back links

C) HITS Algorithm

HITS algorithm ranks the web page by processing in links and out links of the web pages. Rank of the web page is decided by comparing their textual contents with a given query. This algorithm includes a web page which is named as authority if the web page is pointed by many hyper links and a web page is named as HUB if the page point to various hyperlinks. An Illustration of HUB and authority are shown in figure 2.
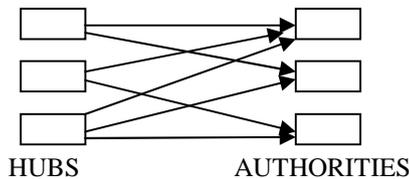

HUBS   AUTHORITIES
Fig 2: Illustration of Hub and Authorities

Table 1: Comparison of Various Algorithms

D) Weighted PageRank Algorithm

Weighted PageRank algorithm is the modification of the PageRank algorithm. WPR decides the rank score depends on the popularity of the pages by taking into consideration the importance of both the in-links and out links of the pages. This algorithm provides high value of rank to the mostly search pages and does not equally divide the rank of a page among it's out-link pages. Every out-link page is given a rank value based on its importance. Importance of a page is decided by calculating its number of in links and out links.

E) Weighted Links Rank Algorithm

This algorithm provides weight value to the link depends on three parameters length of the anchor text, tag in which the link is stored and appropriate position in the page. Aggregated results show that the results of the search engine are improved using weighted links. The length of anchor text assumes to be the good attributes in this algorithm. Relative position, which indicates the original position must not always in synchronization with logical position and it is also not result oriented. Further works in this algorithm consist of tuning of the weight factor of every term for further evolution.

IV. ANALYSIS AND DISCUSSION

Following table explains main techniques, methodology, parameters, importance, relevancy, limitations and results used.

Algorithm	PageRank	HITS	Weighted PageRank	Web PageRanking using Link Attributes
Main Technique	Web Structure Mining	Web Structure Mining, Web Content Mining	Web Structure Mining	Web Structure Mining, Web Content Mining
Methodology	This algorithm computes the score for pages at the time of indexing of the pages	It computes the hubs and authority of the relevant pages. It relevant as well as important page as the result.	Weight of web page is calculated on the basis of input and outgoing links, on the basis of weight the importance of page is decided.	It gives different weight to web links based on 3 attributes: Relative position in page, tag where link is contained, length of anchor text.
Input Parameter	Back links	Content, Back and Forward links	Back links and Forward links.	Content, Back and Forward links
Relevancy	Less	More	Less	more
Quality of Results	Medium	Less than PR	Higher than PR	Medium
Importance	High. Back links are considered.	Moderate. Hub & authorities scores are utilized.	High. The pages are sorted according to the importance	Not specifically quoted
Limitation	Results come at the time of indexing and not at the query time	Topic drift and Efficiency problem	Relevancy is ignored.	Relative position was not effective, indicating that the logical position not always matches the physical position

V. PROPOSED METHODOLOGY

How SEO techniques increase website visibility? For that Web mining technique is used which classify the web pages and internet users by taking into consideration the contents of the page and behavior of internet user. Web mining helps the internet user about the web pages to be viewed in future. Web mining is made of three branches i.e. web content mining, web structure mining and web usage mining. WCM is responsible for exploring the proper and relevant information from the contents of web. WSM is used to find out the relation between different web pages by processing the structure of web. WUM is responsible for recording the user profile and user behavior inside the log file of the web.
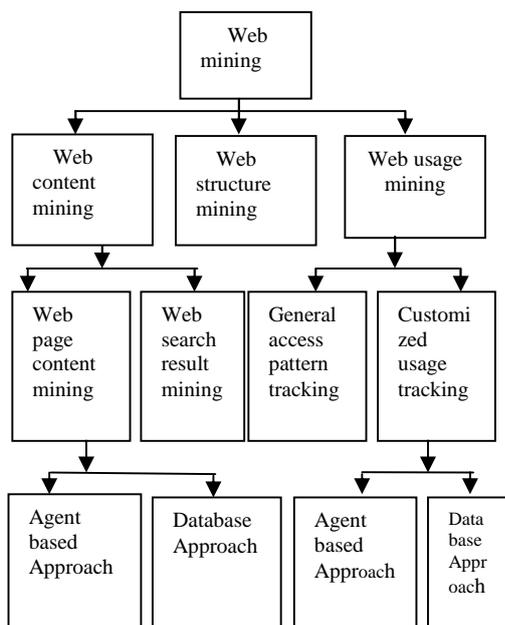

Fig 3: Classification of web mining

VI. EXPECTED RESULT

The WCM mainly works on the structure of the document and WSM explore the structure of the link inside the hyperlink between different documents and classify the pages of web.

CONCLUSION

Depending on the algorithm used, the ranking algorithm provides a definite score to resultant web pages. A search engine should use web PageRanking techniques based on the specific needs of the users. After going through analysis of algorithms for ranking of web pages compare with the various parameters like methodology, input parameters, relevancy of results and importance of the results, it is concluded that existing techniques have limitations especially in terms of accuracy of results, time response, and importance of the results. An efficient web PageRanking algorithm should overcome these challenges efficiently with global standards of web technology.

FUTURE SCOPE

Devise a distance metric to compare the URLs of a user with the URLs of all other users. When a specific user passes a query we could bound the users that participate on the results based on that metric. Plan to test the effectiveness of reward schemes. In the context of collaborative services a reward scheme offers incentives for end users to contribute. Next in this work provide a generic merging process of Users Rank and PageRank. A reasonable future research direction is to evaluate its effect in the search results of a current search engine.

REFERENCES

[1]Hideaki Ishii, Roberto Tempo, and Er-Wei Bai "A Web Aggregation Approach for Distributed Randomized PageRank Algorithms" IEEE Transactions on automatic control, VOL. 57, NO. 11, pp. 2703-2717 November 2013

[2]Athanasios Papagelis and Christos Zaroliagis "A Collaborative Decentralized Approach to Web Search" IEEE transactions on systems, man, and cybernetics—part a: systems and humans, VOL. 42, NO. 5, pp.1271-1290, September 2012

[3]Dilip Kumar Sharma, A.K.Sharma "A comparative analysis of web PageRanking algorithms" CSI Journal International journal on computer science and engineering vol.02, pp. 2670-2676, August 2010

[4]John B. Killoran "How to Use Search Engine Optimization Techniques to Increase Website Visibility" IEEE transaction on professional communication VOL. 56, NO. 1 pp. 50-66, March 2013

[5]Chen Gong, Keren Fu, Artur Loza, Qiang Wu, Jia Liu, and Jie Yang "PageRank Tracker: From Ranking to Tracking" IEEE Transaction on cybernetics VOL. 44, ISSUE NO. 6, p.p 882-893, June 2014

[6] Wenxiao Zhao, Han-Fu Chen, and Hai-Tao Fang "Convergence of Distributed Randomized PageRank Algorithms" IEEE Transactions on automatic control, VOL. 58, NO. 12, p.p 3255-3259, December 2013

[7] Erick Gomez-Nieto, Frizzi San Roman, Paulo Pagliosa, Wallace Casaca, Elias S. Helou, Maria Cristina F. de Oliveira, Luis Gustavo Nonato "Similarity Preserving Snippet-Based Visualization of Web Search Results" IEEE transactions on visualization and computer graphics, VOL. 20, NO. 3, p.p 457-470, March 2014

[8] Giorgos Kollias, Efstratios Gallopoulos, and Ananth Grama "Surfing the Network for Ranking by Multidamping" IEEE Transactions on knowledge and data engineering, VOL. 26, NO. 9, p.p 2323-2336 ,September 2014 [9] Ziyu Guan, Gengxin Miao, Russell McLoughlin, Xifeng Yan and Deng Cai, Member, "Co-Occurrence-Based Diffusion for Expert Search on the Web" IEEE Transactions on knowledge and data engineering, VOL. 25, NO. 5, p.p1001-1014, May 2013

AUTHOR'S PROFILE

She had completed B.E degree in Information Technology, from J.D.I.E.T,Yavatmal And perusing M.E degree in Computer Science and Information Technology, from department of computer science S.G.B. Amravati University, Amravati. She had published paper in International journal of Electronics Communication and Computer Engineering Volume 3, Issue 1, ISSN 2249 –071X on page number 131 on Feature Extraction Techniques for Voice Operated PC Application.