Search Engine Optimization with Efficient Page Ranking Algorithm

Miss. Gayatri Vivekrao Kpase

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? Author Chen Gong, et al [5] had worked to improve adaptability of ranking and formulates the tracking process as a ranking problem.

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 [7] 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 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:

\[ \sum_{v \in B_u} PR(v)/L(v) \]

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.
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.

Fig 2: Illustration of Hub and Authorities

Table 1: Comparison of Various Algorithms

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

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


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.