



An Approach To The Pagerank Based on Dwelling Time and Frequency of Visiting Links

KEYWORDS

Page Rank, Search engine, Search query, Ranking.

Mercy Paul Selvan

Research scholar, DEPARTMENT OF COMPUTER SCIENCE, FACULTY OF COMPUTING, SATHYABAMA UNIVERSITY. CHENNAI.

DR. A. CHANDRASEKAR

DEPARTMENT OF COMPUTER SCIENCE, ST. JOSEPH'S COLLEGE OF ENGINEERING. CHENNAI.

ABSTRACT

Internet is the major source of data. To retrieve the required information from the internet we will go using the search engine. Several search engines are present to retrieve the required information from the large amount of sources. The search engine should be more effective so that the time required to retrieve the data should be less. Normally the web search engine will give more web pages that are related to the search query. Many of the web pages will be of fake and some web pages may not require the user information. So we need to display the pages that contain the web pages in the first. To identify the best web pages I have proposed two approaches they are page rank and the number of visit by the user.

I. INTRODUCTION

Search engine are used to search for the web pages in the internet. The search results are displayed as list in the search engine result page. Search engine will mine the data to display the search pages. The search engine should be more optimized. The optimization of the search engine is done by the many algorithms on a web crawler. The first search engine is W3Catalog which is now inactive. The main work for the search engines is to rank the pages. The page is done by many ways such as number of visits, time spent on the page, number of clicks and etc.,

The web consists of more than billion pages so the search engine should be more effective to retrieve the more appropriated web pages required by the user. The web page also contains large amount of data. In the real world when a user enters the query in the search engine he will get the desired web pages or the user will be redirected. Normally search engines perform many functions to return the web pages need by the user. Different search engine will have different architecture. The result of the search engine will be depending on the architecture of the search engine. Normally all the search engine works in the following way Crawler, indexing and searching. Crawler stores all the information about the web pages and it is used in the web search engines. Crawler is also called as Spider. Crawler will store up-to-date information about the web page and they are used to get specific information about the web page. The web search engine index is used to collect, parse and stores the data for the future use. The most popular search engine will use full-text indexing. Indexing is used to increase the speed and efficiency of the search engine. Indexing plays an important role in the design of a search engine. Search query is a query that is given by the user to get his/her information from the World Wide Web. Once the search test is given the search engine will start searching the web pages. Once the search is over by the search engine they will display the result set. The result set may contain more web pages because the search engine will display all the pages that have the search term or the phase.

A research study shows that normally the user will search only the first two pages of the search result. A page con-

tains 10 results. So the search engine should be high efficient to provide the relevant web pages in that few pages itself. To provide this several search engines follows different techniques some of the techniques are number of visits by the user, time spent by the user or many other information.

In our approach we are going to combine two techniques so that it will provide the high search efficiency to the user. The two techniques are number of visit by the user and the time spent by the user. The result of the search engines will be given rank and the search results will be displayed using the rank. Higher rank pages will be displayed at the first and then the lower rank web pages. By using this the user will get the relevant web pages in the first few web page result.

II. RELATED WORK

The structure of web pages is exploited for search engine in [1] where as time stamp is used as a measure in [2]. In the paper Page Ranking Based on Number of Visits of Links of Web Page by Gyanendra Kumar et al., [7] have devised many algorithm for giving rank to the pages. Some of the algorithm told by them are page rank algorithm based on the pagerank score and many other factors. Based on this the search engine will display the result to the user. In [4], various page ranking algorithms are discussed. Another algorithm is weighted pagerank algorithm which is proposed by Wenpu Xing and Ali Ghorbani in Weighted PageRank Algorithm[3]. It is an extension of the pagerank method It is based on the number of linking it is having to it and others webpages having the link. Based on the this the page rank will be given to the web page. Another pagerank algorithm is pagerank based on visits of links. In this algorithm we assign more value to the visits outgoing links which is visited by the many users. Some times, query may not represent user's requirement. Therefore we need to capture user's behaviour data[6]. In [5], page rank is calculated based on user behaviour. In Methodology for Preprocessing and Evaluating the Time Spent on Web Pages by Peter I. Hofgesang[8] have proposed the search technique based on the time spent on the web page by the user. More time they spent of the page is more the useful page based on the several studies

by the information retrieval and human computer interaction. He proposed an new methodology for preprocess on time spent on web page. In [9], they used web navigational information for acquiring knowledge. In [10], to find out the intention of user query, they used feedbacks.

III. PROPOSED APPROACH

In our approach, the page rank will be based on the number of visits to the web page and the time spent on the webpage. By combining these two methods the efficiency of the search data will be more.

First the page rank should be given based on the number of visit to the web page. The page rank for this is given by the algorithm as proposed by the [7]. The algorithm is explained as follows, the page score is divide equally among the pages available and the inbound pages will bring more sore from the base page to this page. By using this algorithm we can give rank to the pages. After this we will find the rank for the web pages based on the number of time spent by the used in the web pages. By combining both these rank we get a new page rank which will be more target oriented.

To find the page rank based on the time spent on the web page we will use the algorithm proposed by [8]. By using this two algorithm we can find the page rank by combining both the algorithm. The survey says that the time spent on the web page is the important factor for giving the page rank.

In this approach we will add the page scores obtained from the above two algorithms. But the main factor for the more important web page is the time spent on the web page. The page score obtained by the time spent on the web page algorithm is multiplied by two. It is used to increase the probability of the time spent. The page score obtained from the number of visits is added as such so that this probability is little lesser than the time spent on the web page. Consider that the page score of a particular web page obtained from number of visits is given by A. The page score obtained from the time spent on the web page is given by B. Then the page score obtained from our proposed approach is C. Then the C is given by $C = A+2B$. Now based on the page score C the search engine will display the page result.

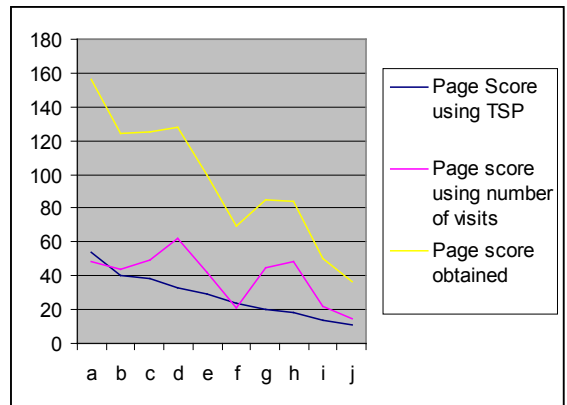
The page score should not be depend only on one attribute. The web page may contain an useful information but the time spent on it will be very less or the number of visits can be as it is an new web page. So this proposed approach we have used two attributes to calculate the page rank. This approach has given more important to the time spent on the web page as this is the main attribute to find the page rank. Consider an example. The following table

consist of ten web pages with page score obtained using the algorithms

Web page	Page Score using TSP	Page score using number of visits	Page score obtained
a	54	48	156
b	40	44	124
c	38	49	125
d	33	62	128
e	29	42	100
f	24	21	69
g	20	45	85

h	18	48	84
i	14	22	50
j	11	15	37

Table 1: Page Score of various web pages



From the above chart we can see that the web page d is fourth in the time spent on the web page and first in the number of visits. By using our proposed algorithm the new page rank for this web page is second. From the chart we can see that the change in the page rank. The proposed method will be more effective in providing the user with more optimized web page result.

IV. CONCLUSION & FUTURE WORK

In this we have introduced an approach to change the page rank to a web site by combining two algorithms time spent on the web page and the number of visits to the web page. This approach shows that the search will be more optimized to the user because it uses two attribute to find the page rank. In future, we could also calculate page rank by the other factors like 'sharing', 'like', 'book marked'.

REFERENCE

[1] S. Brin and L. Page. The Anatomy of a Large-Scale Hyper textual Web Search Engine. In Proceedings of the 7th International World Wide Web Conference, pages 107–117, 1998. | [2] Shiguang Ju, Zheng Wang, Xia Lv-Improvement of Page Ranking Algorithm Based on Timestamp and Link. In 2008 International Symposiums on Information Processing. | [3] Wenpu Xing, Ali Ghorbani. Weighted PageRank Algorithm[C], Proceedings of the Second Annual Conference on Communication Networks and Services Research, IEEE, 2004. | [4] Neelam Duhan, A. K. Sharma, Komal Kumar Bhatia, Page Ranking Algorithms: A Survey Advance Computing Conference, 2009. IACC 2009 IEEE International. | [5] Mercy Paul Selvan et al., Web page importance based on User's Browsing Behaviour – | A Novel Approach , International Conference on Mathematical Sciences, Elsevier 2014. | [6] Zhiuan chen et al. A learning Approach to SQL Query Results Ranking Using Skyline and User's Current Navigational Behaviour, IEEE transactions on Knowledge and Data Engineering, vol.25, No.12 December 2013. | [7] Gyanendra Kumar1, Neelam Duhan2, A. K. Sharma , Page Ranking Based on Number of Visits of Links of Web Page, ICCCT pages 11-14, 2011 | [8] Peter I. Hofgesang, Methodology for Preprocessing and Evaluating the Time Spent on Web Pages, IEEE – 2006. | [9] C. Shahabi, A. M. Zarkesh, J. Adibi, and V. Shah. Knowledge discovery from users web-page navigation. In RIDE, page 20, Birmingham, England, 1997. | [10] Cheng Lu et al. A New Algorithm for inferring User Search Goals with Feedback Sessions. IEEE Transactions on Knowledge and Data Engineering , vol 25, no.3 March 2013. |