Scientific Journal of Impact Factor (SJIF): 5.71

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 5, Issue 05, May -2018

AN ENHANCED TECHNIQUE TO IMPROVE THE PAGE RANK BASED ON IN-LINKED WEIGHTAGE OF WEB GRAPH

[1]Pooja Mathur , [2] Ram Lal Yadav

[1]*MTech Scholar, Dept. Of Computer Science & Engineering*
Kautilya Institute of Technology & Engineering-Jaipur-Rajasthan-India

[2] *Associate Professor, Dept. Of Computer Science & Engineering*
Kautilya Institute of Technology & Engineering-Jaipur-Rajasthan-India

Abstract:- *The World Wide Web crops new exciting challenges for information retrieval as the volume of information on the web is rising speedily. Search engines show a vital role in information retrieval on the World Wide Web. Search engine optimization is a strategic procedure to take a web page in best search results of a search engine. Search Engine relates to the action of optimizing distinct WebPages or the whole website to make them friendlier to obtain greater ranking in the search queries. All the key search engines such as Google, Yahoo, Ask, Bing, etc. rank web-pages based on certain factors that affect its ranking. This research is based on the extension of standard Page Ranking Algorithm by using the weightage of In-Linked web pages. In this proposed research, the weight is distributed on the basis of popularity of linked web pages rather than evenly dividing to all the out linked pages. In this paper, a new enhanced page ranking algorithm known as ILW (In-Link Weightag) Page Rank Algorithm is discussed, in which the weightage of In-Linked web pages are used to calculate the rank of the web page.*

1. Introduction

Search engine optimization is a method that practices a search principle for search engine to obtain upper ranking for web-pages or websites to improve the probability of website access. Search engine is not only the essential role in the website to deliver a suitability for users, but it is also a real tool of accepting web user's behavior. Search engine optimization relays to the action of optimizing the separate web pages or the website to get higher position in the search results. When internet users search for the service related, the website weight to particular keywords which internet users hunt for online [1]. The method of optimizing search engine contains researching keywords, creating content, building links and building sure that the website is visible in the search engines.

The whole Search Engine method works with two types of optimization techniques, on-page and off-page SEO optimization. The SEO procedure starts with on-page SEO optimization. The aim of the page ranking is to make the user get the wanted result at the top of the list and the paper presents new page ranking algorithm such that it can improve the search results.

1.1. Background

Ideally everyone wants his site to be more user-friendly and to be displayed somewhere on the top pages of results. To attain a high location in search results, the particular site must be more distinguishable by a search engine crawler than simple websites. Google's history has been started from developing the first search engine algorithm, PageRank in 1997. From there on, Google has developed its algorithms in their mission to enhance the search experiences of the users. Google has developed its web searching abilities and one of the most known changes to this web search is PageRank [1]. The next big algorithm changes in Google history are Google Panda, an algorithm which focused on getting rid of low-quality websites on the search results and Google Penguin, an algorithm that focuses on reducing spammy websites. Penguin is an algorithm from Google that was targeted against web spam and it was launched in April 2012. According to Google's Principal Engineer Matt Cutts in the Official Google Webmaster Central Blog post "An additional step to reward the Penguin that it will control around 3% of search queries". Another New Google 'Hummingbird' Algorithm" (2013) that Google has introduced a new search algorithm called Hummingbird.

1.2. Motivation

Online searching has been an important part of the daily routine lives of almost people. World Wide Web browsing is easily available in handheld devices. To find the information about the latest device, the direction to go to any popular restaurant search engines become the important part of daily life routine. Beyond small requirements, search engines are

highly progressively becoming the main source guiding people to necessary information. Due to the huge number of websites, search engines become the difficult job of arranging the billions of web pages and showing only the best important desired web pages in the search engine results page (SERP) for the asked search query. As the Internet is growing continuously with the amount of websites available, it has been most challenging typical task for sites watching for a user to attain visibility.

The paper is arranged as follows: Section 2 presents the related literature survey on various linked based page ranking algorithms, Section 3 of the paper gives the proposed approach to enhance the standard page rank algorithm. Section 4 presents our proposed algorithm. Finally, Section 5 describes experimental results and Section 6 shows conclusion and future work for the proposed ILW Page Rank Algorithm.

2. Literature Survey

L. Page and S. Brin [1] proposed the Page Rank algorithm to find the importance of web pages using the link structure of the web. In this approach Brin and Page spreads the indication of simply calculating in-links equally, by normalizing by the number of links on a page. The Page Rank algorithm is specified as [1]: We assume page A has pages T1...Tn which point to it (i.e., are citations). Here the parameter d is a damping factor, which can be initialising between 0 and 1. We normally set d to 0.85. Page Ranks form a probability distribution over web pages, so the sum of all web pages' Page Ranks will be 1 [2] And the d damping factor is the probability at each page the "random surfer" will get bored and request another random page [3]. Xianchao Zhang, Hong Yu, Cong Zhang, and Xinyue Liu [4] has defined an Improved Weighted HITS algorithm shows that in the web structure, weights of links not only depend on the local information of the two pages which the link is adjacent to but also the vicinity context. Hema Dubey, Prof. B. N. Roy [5] described a new page rank algorithm which uses a normalization technique based on mean value of page ranks. This scheme reduces the time complexity of the traditional Page Rank algorithm by reducing the number of iterations to reach a convergence point.

Gyanendra Kumar, Neelam Duhan and A. K. Sharma [6] have given a new algorithm which is based on mechanism called Page Ranking based on Visits of Links(VOL) is being devised for search engines, which works on the basic ranking algorithm of Google i.e. PageRank and takes number of visits of inbound links of Web pages into account. They have given this algorithm by improving a standard page rank algorithm and added an extra parameter "number of visit of links" to standard page rank algorithm. In this manner, the score of a web page is calculated on the basis of the visits of inlinks. Ranveer Singh and Dilip Kumar Sharma [7] have given new algorithm which is based on hyper-linking structure of the pages named as RatioRank. This algorithm takes the weight of the inlinks and outlinks based on the popularity of the links in a defined ratio. As the popularity meant for number of inlinks and outlinks to that link of the page. This algorithm also includes the number of times the user visits the inlinks. This algorithm is the extension to the Weighted Page Rank [2] in which the more weights are given to the inlinks and outlinks on the basis of popularity of web pages.

Neelam Tyagi and Simple Sharma [8] have given a new algorithm named improved Weighted PageRank algorithm. In this algorithm they assign more rank value to the outgoing links which is most visited by users and received upper popularity from number of inlinks. We do not consider here the popularity of outlinks which is considered in the original algorithm. The advanced approach in the new algorithm is to determine the user's usage trends. The user's browsing behavior can be calculated by number of hits (visits) of links. They have mentioned that the original Weighted PageRank algorithm assigns larger rank values to more imperative (popular) pages. Each outlink page gets a value proportional to its popularity (its number of inlinks and outlinks) [9]. The benefit of WPR (VOL) is that this method uses link structure of pages, the popularity of inlinks and their browsing information, the top returned pages in the result list is supposed to be highly relevant to the user information needs. A link with high probability of visit contributes more towards the rank of its out linked pages. The rank value of any page by original Weighted PageRank method will be same either it is seen by user or not, because it is totally dependent upon link structure of Web graph and popularity of inlinks and outinks. While the ordering of pages using WPR (VOL) is more target-oriented [10]. In WPR(VOL), a user can not intentionally increase the rank of a page by visiting the page multiple times because the rank of the page depends on the probability of visits (not on the count of visits) on back linked pages.

M. Shamiul Amin, S. Kabir and R. Kabir [11] have given a Score Based Page Ranking (SBPR) which is based on the content mining as well as the usage information of pages. They have calculated a score for each page which is related to query using the frequencies of keywords of that query as well as the synonyms of those keywords, and also using the popularity of that particular page. In this algorithm each page becomes as core for that reason our algorithm is known as "Score Based Page Ranking Algorithm".

A search engine has become a necessary component of our everyday life. Enterprises use the search engine for marketing role as the aim of search engine is to improve the ability of web pages retrieval. Search engine is an approach for the retrieval of web pages related to user requests on the Internet. SEO is a process that raise the rank of web sites and web pages.

3. Proposed Approach

Page Rank algorithm is classically uses by the world famous search engine, Google. Additional search engines use Page Rank algorithm for ranking the different types of pages in the web-sites. The Page Rank algorithm is typically depends on the link structure of the several types of web pages or websites which are open on the internet. The PageRank algorithm is based on the in links and out links. Therefore, if back link are more than the page rank of particular page is also high. Page Rank grades pages based on the web structure.

Google, which among search engines is graded in the main place, uses the Page Rank algorithm. Page Rank is developed by Google and is called after Larry Page, Google's co-founder and president [1]. Page Rank is a numeric value that shows how significant a page is on the web. Google figures that when one page associates to another page, it is efficiently casting a vote for the new page. The more votes that are cast for a page, the more important the page must be.

Page Rank Formula is:

$$PR(A) = (1 - d) + d [PR (T1) / C(T1) + ... + PR (Tn) /C(Tn)]$$

Where:
PR (A): Page Rank of page A.
PR (Ti): in link Page Rank of Ti pages linked to page A.
C (Ti): Number of out Links from Page Ti.
D: Damping factor (between 0 and 1).

Damping factor d is normally set as a 0.85. So it is easy to each & every page that it is distributes 85% of its original Page Rank. To compute the page rank of every web page or web site we must to see the page rank of each page. Without information of individual page rank we are not capable to do that point to it and number of the inbound or out bound links from each of those pages.

3.1. Design Goal of a Proposed Approach

Our main goal is to improve the quality of indexing of web pages which is given by Page Rank for web search engines. Many people believed that a complete search index would make it possible to find anything easily in search engines. Our final design goal is to make a design architecture that can support to find optimum Page Rank of web pages for search engine to provide a better indexing to the web pages on the basis of popularity of web pages.

3.2. Proposed ILWPR Formula

Here, we discuss a new approach which is based on the extension of standard Page Ranking Algorithm by using the weightage of In-Linked web pages. In this proposed approach, the weight is distributed on the basis of popularity of linked web pages rather than evenly dividing to all the out linked pages.

In this paper, a new enhanced page ranking algorithm known as ILW (In-Link Weightage) Page Rank Algorithm is discussed, in which the weightage of In-Linked web pages are used to calculate the rank of the web page. We have proposed ILW Page Rank Equation as follows:

$$PR(A) = (1 - d) + d [PR (T1) / W (FT1) + ... + PR (Tn) /W (FTn)]$$

Where:
PR (A): Page Ranking of page A,
PR(Ti): Page Ranking of page Ti which points to page A
W (FTi): (PR (FTi1)/C(FTi1) + PR (FTi2)/ /C(FTi2) + ……. + PR (FTiN)/ /C(FTiN)).
FTij: Out linked page j which is linked from Page Ti.
C(FTi): Number of out-link from web page FTi.
d: Damping factor (between 0 and 1)

Features of proposed approach are:

- Weightage of In-Linked web pages are used to compute the rank of the web page.
- Weight is distributed on the base of fame of linked web pages
- Page Rank value can be calculated offline by using only web graph.
- Page Rank is created upon the linking structure of the entire web page.
- It is the query independent algorithm that allocates a value to each web document independent of query.

• It concerns with static value of a web page.

4. Algorithm

The ILW Page Rank algorithm is as follows:

1) Initially assign PAGE RANK of any value to all the web pages, let it be1.

2) Calculate new Page Rank of each web page by using previous Page Rank.

3) Apply following formula to calculate Page Rank in Step 2:

PR (A) = (1 - d) + d [PR (T1) / W (FT1) +.....+ PR (Tn) / W(FTn)]

Where:
PR (A): Page Rank of page A,

PR (Ti): in link Page Rank of Ti pages linked to page A.

PR(Ti): Page Ranking of page Ti which points to page A
W (FTi): (PR (FTi1)/C(FTi1) + PR (FTi2)/ /C(FTi2) + ……. + PR (FTiN)/ /C(FTiN))

FTij: Out linked page j which is linked from Page Ti.

C(FTi): Number of out-link from web page FTi

d: Damping factor which can be set between 0 and 1, but it is usually set to 0.85.

4) Compare newly calculated Page Rank of each web page to its previous Page Rank of same web page.

5) If there is any difference between previous Page Rank and newly calculated Page Rank of any web page, than assign newly calculated Page Rank to each web page and go to step 2.

6) Finally assign latest calculated Page Rank to each web pages

7) Sort the web pages on the basis of Page Rank of high to low.

8) Display the web pages according to step 7 and show the web page having the highest Page Rank.

So, we have seen that this newly developed Algorithm gives rank to separate web pages and does not give rank to whole web sites. The Page Rank of page Ti which point to page A does not affect the Page Rank of page A evenly. In this proposed Page Rank algorithm, the Page Rank of a page Ti is always weighted by the Page Rank of outbound pages FTij on page Ti. The result of this is that an extra inbound link for page A will continuously increase page A's Rank. Lastly, the totality of the weighted Page Ranks of all pages Ti is multiplied with a damping factor d which can be set between 0 and 1. Usually d is set as 0.85.

5. Experimental and Result Analysis

In the section of this paper, we have summarized the results that we have found from our experimental setup. As we have mentioned in our goal and expected outcome that this Research Work will provide enhancement over Standard Page Ranking for indexing the web pages of entire web site. Here, we have taken live web site www.indianrail.gov.in to find PageRank for each web pages by using proposed ILWPR approach and comparison of the Page Rank and Indexing is done with the results what we have found by using SPR approach. In this section we will show and clarify some results what are taken from the work we have done. Basically we will try to justify experimentally what we already explained in the algorithm in previous sections. Here we have taken an example of a web page http://www.indianrail.gov.in.

URL	URL SNo
http://indianrail.gov.in/#	12
http://indianrail.gov.in/tatkal_Scheme.html	14
http://indianrail.gov.in/about_Concert.html	10
http://indianrail.gov.in/hindex.html	9
http://indianrail.gov.in/dont_Know_Station_C...	8
http://indianrail.gov.in/booking_Location.html	5
http://indianrail.gov.in/copyright.html	13
http://www.cris.org.in/	23
http://indianrail.gov.in/stn_code.html	16
http://indianrail.gov.in/fare_Enq.html	15
http://indianrail.gov.in/disclaimer.html	4
http://www.indianrail.gov.in	36
http://indianrail.gov.in/sitemap.html	19
http://www.irctc.co.in	29
http://www.indianrailways.gov.in/railwayboard/	22
http://indianrail.gov.in/international_Tourist.html	24
http://indianrail.gov.in/MobileTicketing.html	17
http://indianrail.gov.in/faq.html	25
http://indianrail.gov.in/know_Station_Code.html	26
http://www.indianrailways.gov.in/railwayboar...	6
http://indianrail.gov.in/pnr_Enq.html	27
http://indianrail.gov.in/inet_curbkg_Enq.html	31
http://indianrail.gov.in/train_Schedule.html	32
http://indianrail.gov.in/other_Rly_Sites.html	33
http://indianrail.gov.in/sms_Service.html	35

Figure 1: URL of Web Pages with Serial No.

URL SNo	WILPR Score	SPR Score
12	0.7035975456	1.95384195761133
14	0.374274416117003	0.670255806607209
10	0.34780465056	0.61704882174066
9	0.34580758944	0.628216239158077
8	0.34530832416	0.662043944800282
5	0.34518350784	0.660626824296009
13	0.34323545629563	0.590329547000475
23	0.327946953417224	0.559756969781881
16	0.326007672836394	0.554732836180065
15	0.32576691491037	0.552745722812576
4	0.31323052992	1.0837173186365
36	0.312009820063482	0.436768330968671
19	0.311456674480214	0.517302941652984
29	0.303470004572571	0.501908943312603
22	0.302315733514014	0.49899457431816
24	0.301987403242333	0.494305718605473
17	0.299295572657147	0.491269708296794
25	0.295130209902057	0.475930335868475
26	0.290093425853311	0.464969871157401
6	0.279942105479714	0.446795397150629
27	0.279820705113885	0.441686194584961
31	0.267613692048139	0.408524597328114
32	0.262774569942887	0.398329278053653
33	0.251867141485893	0.373826490988612
35	0.238028538173957	0.341834149870829

Figure 2: ILWPR and SPR Score

In the Figure 1, we show the URL and assigned unique serial no of each web page which is associated to the root web page http://www.indianrail.gov.in. In figure 2, we show the page ranking of each web page according to the Standard PageRank Algorithm as well as ILW Page Rank Algorithm and it is sorted in the order of ILWPR Algorithm. In figure 3, the calculated ILWPR and SPR indexing is shown in the sorted order of ILWPR Ranking. Experimental setup assigns High Rank to root web page https://www.indianrail.gov.in.

Our implementation gives the In-link and out-link of each associated web page of the root web page and ILWPR and SPR Rank Scored are calculated on the basis of linked structure of all the web pages of website. URL serial number is assigned to all the associated web links in ascending order starting from 1 at the start of Page Rank calculation. We compare the results of both the algorithms and we have seen that some web pages has been enhanced their Page Ranking with compare to Standard Page Ranking Algorithm.

URL SNo	WILPR Index	SPR Index
12	36	36
14	35	32
10	34	28
9	33	29
8	32	31
5	31	30
13	30	27
23	29	26
16	28	25
15	27	24
4	26	35
36	25	13
19	24	23
29	23	22
22	22	21
24	21	20
17	20	19
25	19	18
26	18	17
6	17	15
27	16	14
31	15	12
32	14	11
33	13	10
35	12	9

Figure 3: ILWPR and SPR Indexing

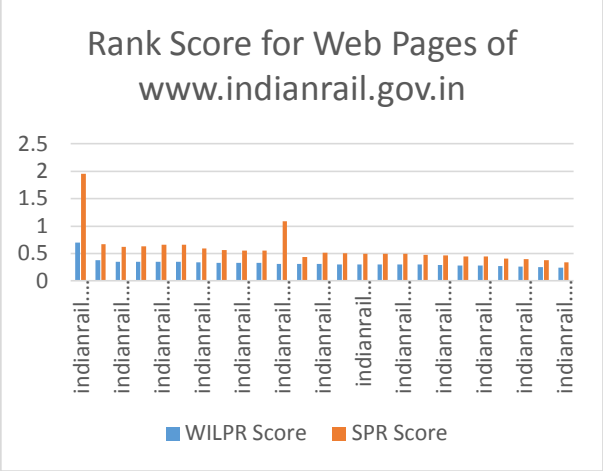

Figure 4: Result Analysis of ILWPR with SPR

We have seen that counting of two or more than two web pages having the same Page Rank score, is more in SPR results while it is lesser in the case of ILWPR Results. In other words, the distinct Page Rank scores are better in ILWPR results with compare to SPR results. Finally, we compare the results of both the algorithms in figure 4. We have seen that some pages having more in-linked web page weightage get more rank by using proposed ILW Page Rank Algorithm with compare to the Standard Page Rank Algorithm.

6. Conclusion and Future Scope

In this paper we presented a modified page ranking algorithms which is more target oriented than original standard page rank algorithm. This modified algorithm calculates page rank value or importance of web pages based on the extension of standard Page Ranking Algorithm by using the weightage of In-Linked web pages. In this proposed research, the weight is distributed on the basis of popularity of linked web pages rather than evenly dividing to all the out linked pages. In this paper, a new enhanced page ranking algorithm known as ILW (In-Link Weightage) Page Rank Algorithm is discussed, in which the weightage of In-Linked web pages are used to calculate the rank of the web page. In this paper page ranking algorithm, for webpage ranking based on different parameters, along with the proposed ILW Page Ranking Algorithm, in the parameters hyperlinking structure is used. The proposed algorithm will consider the inlinks, outlinks and the weightage of in-linked web pages. The algorithm will give results according to the weightage of in-linked web pages

In the future, the researchers can plan to explore more on the page rank algorithm based on damping factor to enhance the performance of the proposed scheme.

7. References

[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] Wenpu Xing, Ali Ghorbani, "Weighted PageRank Algorithm", Proceedings of the Second Annual Conference on Communication Networks and Services Research, IEEE, 2004.
[3] Ying Liu, Yongjing Lin, "Supervised HITS Algorithm for MEDLENE Citation Ranking, 1- Proceedings of IEEE, 4244-1509-8/07, IEEE, 2007.
[4] Xianchao Zhang, Hong Yu, Cong Zhang, and Xinyue Liu, "An Improved Weighted HITS Algorithm Based on Similarity and Popularity", Second International Multi Symposium on Computer and Computational Sciences, IEEE, 2007.
[5] Hema Dubey, Prof. B. N. Roy, "An Improved Page Rank Algorithm based on Optimized Normalization Technique", International Journal of Computer Science and Information Technologies, Vol. 2 (5), 2011, pp. 2183-2188.
[6] Gyanendra Kumar, Neelam Duhan and A. K. Sharma, "Page Ranking Based on Number of Visits of Links of Web Page", International Conference on Computer & Communication Technology (ICCCT)-2011.
[7] Ranveer Singh and Dilip Kumar Sharma, "RatioRank: Enhancing the Impact of Inlinks and Outlinks", 3rd IEEE International Advance Computing Conference (IACC), 2013.
[8] Neelam Tyagi and Simple Sharma, "Weighted Page Rank Algorithm Based on Number of Visits of Links of Web Page", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3, July 2012.
[9] Ranveer Singh and Dilip Kumar Sharma, "Enhanced-RATIORANK: Enhancing Impact of Inlinks and Outlinks", Proceedings of 2013 IEEE International Conference on Information and Communication Technologies, ICT 2013.
[10] Zaved Akhtar and Saoud Sarwar, "Page Ranking Algorithm Based on Counts of Link Hits (PRCLH): An Implementation", International Journal of Computer Science and Information Technologies, Vol. 5 (3), 2014, pp. 2707-2712.
[11] M. Shamiul Amin, S. Kabir and R. Kabir, "A Score based Web Page Ranking Algorithm", International Journal of Computer Applications (0975 – 8887) Volume 110 – No. 12, January 2015.
[12] Z. Gyongyi, H. Garcia-Molina, and J. Pedersen, "Combating web spam with TrustRank", In Proceedings of the 30th International Conference on Very Large Databases (VLDB), 2004.
[13] R. Barrett, P. Maglio, D. Kellem, "How to personalize the Web", Proc. Conf. on Human Factors in Computing Systems, 1997.
[14] T. Berners-Lee, R. Cailliau, A. Luotonen, H.F. Nielsen, A. Secret. "The World-Wide Web", Communications of the ACM, 37(1994), pp. 76–82.
[15] K. Bharat, A. Broder, M.R. Henzinger, P. Kumar, S. Venkatasubramanian, "Connectivity Server: Fast Access to Linkage Information on the Web", Proc. 7th Intl. World Wide Web Conf., 1998.

