

A Similitude Study on Weighted Pagerank and Enhanced Ratorank Algorithms Using Visit of Links of a Web Page

V.Annapoorna
Asst.Professor
Dept of IT, ANITS

M.V.Kishore
Asst.Professor
Dept of IT, ANITS

Abstract: World Wide Web is an Internet-wide distributed hypermedia information retrieval system which provides access to a large universe of documents also called as web pages. The web pages are formatted in markup language called HTML (*Hypertext Markup Language*). Each web page contains text, multimedia data as well as links to other web pages. The idea of a world of boundless information in which all items have a reference by which they can be retrieved. This means you can jump from one web page to another simply by clicking on hot spots or reference words. Web pages can be explored through their web query interfaces. To extract essential information from web, web extractors execute number of tasks based on their respective structural design. When a user enters a query to the search engine, it generally extracts a bulk number of pages in response to user's query. To extract more resourceful web pages in the result list, various page ranking techniques are applied on the query results. Most of the ranking algorithms are either link or content oriented. Which do not think about user usage trends. In this paper, an analysis of Weighted Page Rank Algorithm (WPR) based on number of Visits of Links (VOL) of a web page and Enhanced Ratio Rank (ERR) algorithm using visit of links of a web page is performed. The paper also presents the comparison between WPR using VOL and ERR algorithms. Weighted Page Rank Algorithm based on number of Visits of Links is being devised for web explorers, which works on the basis of weighted page rank algorithm and takes number of visits of both the inlinks and outlinks of the pages into account and distributes rank scores based on the popularity of web pages. Ratio Rank is improved version of Page Rank Algorithm. Ratio Rank algorithm works on the basis of linkage organization web for calculating rank value of page. The prime idea in this is to increase the weight of in links over the weight of out links, as it is more important. So, this comparison is very useful to display most precious pages on the top of the result list on the basis of user browsing performance, which reduce the search space to a large extent.

Keywords: weighted page rank, inlinks, outlinks, Ratio Rank.

INTRODUCTION

Web plays dynamic role because it contains vast data as collection of large number of WebPages and every second new pages are added, updated and deleted in web. Retrieving efficient, relevant and meaningful information from this large source of information is very challenging job. Every search engine applies an algorithm to large number of WebPages in search results which calculators rank of every WebPages and ensure that most efficient and relevant WebPages as per query made by user appear first in search results. A page ranking mechanism called

Weighted Page Rank Algorithm based on Visits of Links (VOL) is being devised for search engines, which works on the basis of weighted page rank algorithm and takes number of visits of inbound links of web pages into account.

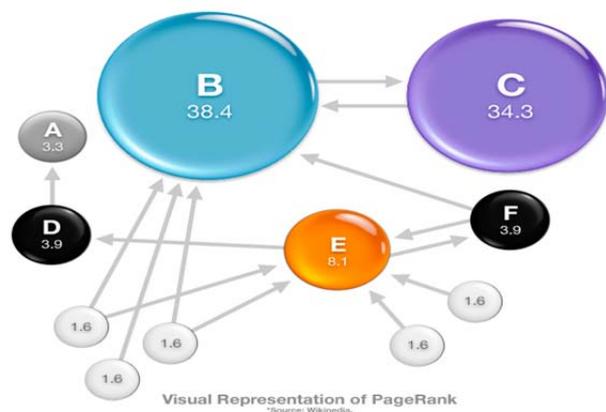


Figure: Representation of page rank

The original Weighted Page Rank algorithm (WPR) is an extension to the standard Page Rank algorithm. WPR takes into account the importance of both the inlinks and outlinks of the pages and distributes rank scores based on the popularity of the pages. The proposed algorithm is used to find more relevant information according to user's query.

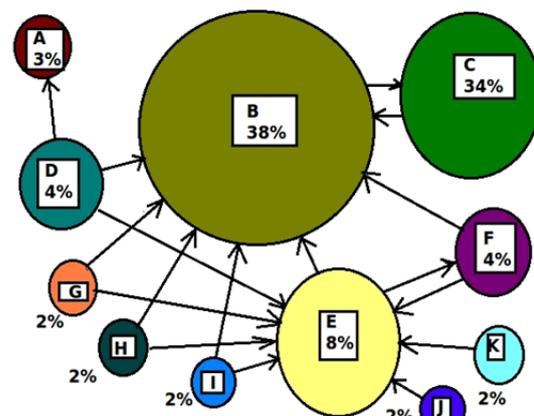


Figure: Calculation of page rank

There are a variety of algorithms developed in current web architecture which are targeted to search engine functionality; few of them are Page Rank, HITS, and

RANDOMIZE HITS etc. As most of the ranking algorithms proposed are either link or content oriented in which consideration of user usage trends are not available. Some algorithms assign larger rank values to more important (popular) pages. Each outlink page gets a value proportional to its popularity (its number of inlinks and outlinks).

We mainly focus on finding more relevant information according to user's query. By Assigning more rank value to the outgoing links which is most visited by users and received higher popularity from number of **inlinks**.

Advantages:

- Works more Faster in rendering frequent pages based on user trends.
- Doesn't give more rank values based on Inbounds.
- User can't 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.

METHODOLOGY

Algorithm for weighted page rank using VOL:

The original Weighted PageRank algorithm assigns larger rank values to more important (popular) pages. Each outlink page gets a value proportional to its popularity (its number of inlinks and outlinks). The popularity from the number of inlinks and outlinks is recorded as $Win(v,u)$ and $Wout(v,u)$, respectively. Here we proposed an improved Weighted PageRank algorithm. In this algorithm we assign more rank value to the outgoing links which is most visited by users and received higher 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.

Input:

- Query to be searched is the input.

Method:

Step 1: Finding a Website: Find a website which has rich hyperlinks.

Step 2: Building a Web Map: Then generate the web map from the selected website.

Step 3: Calculate W in (v, u): Then calculate the W in (v, u) for each node present in web graph.

$$W_{(m,n)}^{in} = \frac{I_n}{\sum_{p \in R(m)} I_p}$$

Where

- $W_{(v,u)}^{in}$ is the weight of link(v,u) calculated based on the number of in links of page u and the number of in links of all reference pages of page v.
- I_n and I_p are the number of incoming links of page n and page p respectively.
- $R(m)$ denotes the reference page list of page m.

Step 4: Apply proposed formula: Now calculate the PageRank value of the pages.

$$WPR_{vol}(u) = (1 - d) + d \sum_{v \in B(u)} \frac{L_u WPR_{vol}(v) W_{(v,u)}^{in}}{TL(v)}$$

- U represents a web page,
- B(u) is the set of pages that point to u,
- d is the dampening factor.
- $WPR_{vol}(u)$ and $WPR_{vol}(v)$ are rank scores of page u and v respectively,
- L_u denotes number of visits of link which is pointing page u form v.
- $TL(v)$ denotes total number of visits of all links present on v.

Algorithm for Ratio Rank:

Ratio Rank is improved version of Page Rank Algorithm. Ratio Rank again uses link structure or web structure mining for calculating rank value of page. That uses weight of out links, weight of in links and number of visits of links by users as the parameters for calculating ratio rank of any page.

Input:

- Query to be searched is the input.

Method:

Step 1: Finding a Website: Find a website which has rich hyperlinks.

Step 2: Building a Web Map: Then generate the web map from the selected website.

Step 3: Calculate W in (v, u): Then calculate the W in (v, u) for each node present in web graph.

$$W_{(m,n)}^{in} = \frac{I_n}{\sum_{p \in R(m)} I_p}$$

Where

- $W_{(v,u)}^{in}$ is the weight of link(v,u) calculated based on the number of in links of page u and the number of in links of all reference pages of page v.
- I_n and I_p are the number of incoming links of page n and page p respectively.
- $R(m)$ denotes the reference page list of page m.

$$W_{(m,n)}^{out} = \frac{O_n}{\sum_{p \in R(m)} O_p}$$

Where

- $W_{(m,n)}^{out}$ is the weight of link(m,n) calculated based on the number of out links of page u and the number of in links of all reference pages of page v.
- O_n and O_p are the number of incoming links of page n and page p respectively.

Step 4: Apply proposed formula: Now calculate the Page Rank value of the pages.

$$RR(u) = (1-d) + d \sum_{v \in B_u} \frac{(v_u * x * W_{(v,u)}^{in} + y * W_{(v,u)}^{out}) * RR(v)}{TL(v)}$$

- U represents a web page.
- $RR(u)$ and $RR(v)$ are ratio rank values of page u and v respectively
- $RR(u)$ and $RR(v)$ are Ratio Rank values of pages u and v respectively

- d is the dampening factor.
- RR (u) and RR (v) are rank scores of page u and v respectively, Lu denotes number of visits of link which is pointing page u form v.
- TL(v) denotes total number of visits of all links present on v.
- Values of x and y is always set between 0 and 1 (Where x>y)

Algorithm for Enhanced Ratio Rank:

Enhanced-Ratio Rank also consider ratio of weight of the inlinks and weight of out links and visit counts of links by users for calculation of the rank value of particular page. It checks which ratio gives the best result i.e. which ratio of inlinks weight and out links weight helps to give better relevancy of the web pages. New Enhanced algorithm is given as follows:

$$RR(u) = (1-d) + d \sum_{v \in B_i} \frac{(v_u * .7 * w_{(v,u)}^{in} + .3 * w_{(v,u)}^{out}) * RR(v)}{TL(v)}$$

It uses same parameters as Ratio Rank equation. As in equation 70 percent of the weight of in links and the 30percent of the weight of the out links is being used because this gives better result as compare to other ratios. By using all three parameters for computing the page rank value of WebPages and taking the best ratio of weight of in links and out links gives the better relevancy of web pages. But the problem of theme drift (some link may not give the search results about the query) still exists in this algorithm.

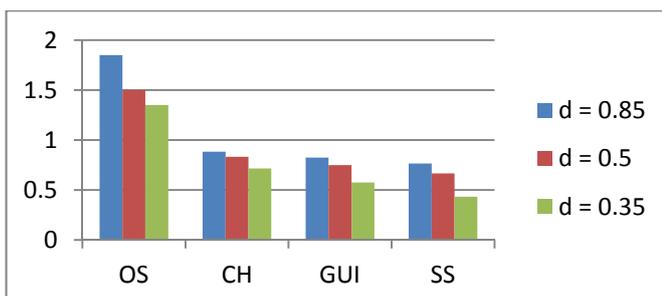
RESULT ANALYSIS

In this section we will describe and explain some results we got from the work we have been doing. Here we have taken a hyperlinked web graph and calculated Page Rank value of each page based on original weighted page rank algorithm and proposed algorithm i.e. based on number of visits of links (VOL).

Results for Weighted Page Rank Algorithm based on VOL:

Table 1: Ranks with varying values of 'd'

Keyword	Ranks		
	d=0.85	d=0.5	d=0.35
Operating Systems	1.85	1.5	1.35
Computer Hardware	0.8833	0.833	0.7166
GUI	0.825	0.75	0.575
System Software	0.766	0.6666	0.433

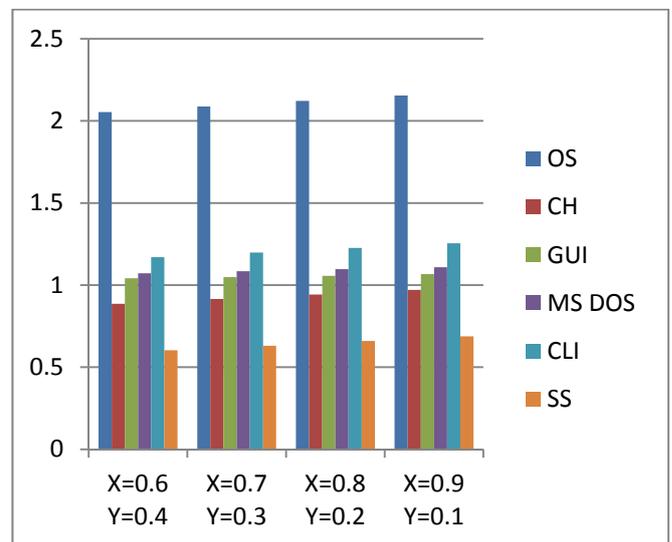
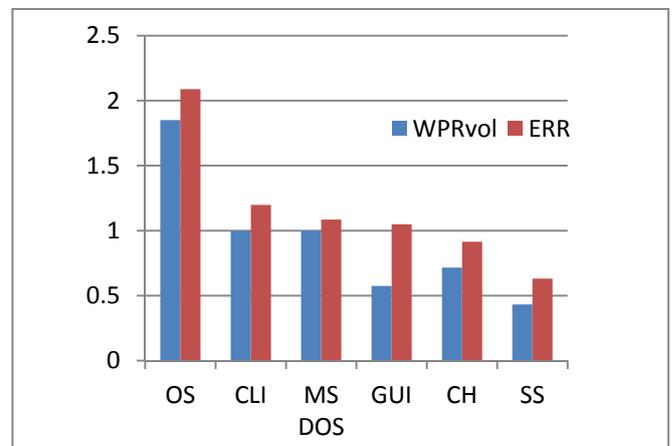


For a hyperlinked web graph and calculated Page Rank value of each page based on original weighted page rank algorithm and proposed algorithm i.e. based on enhanced ratio rank.

Result for Enhanced Ratio Rank Algorithm:

Table 2: Ranks with varying values of In links(x) and out links(y)

In links (x)	Out Links (y)	Rank for each keyword					
		OS	CH	GUI	MS DOS	CLI	SS
0.6	0.4	2.054	0.886	1.042	1.072	1.17	0.603
0.7	0.3	2.088	0.915	1.049	1.085	1.19	0.631
0.8	0.2	2.122	0.943	1.056	1.097	1.23	0.66
0.9	0.1	2.155	0.971	1.067	1.109	1.25	0.688



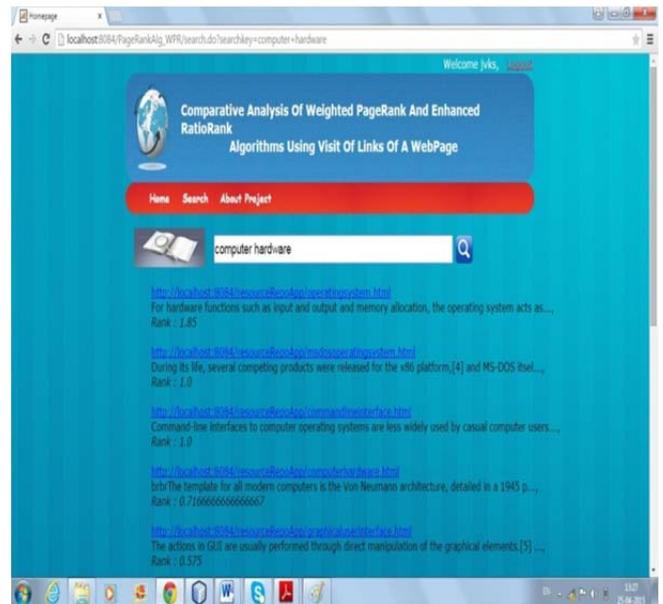
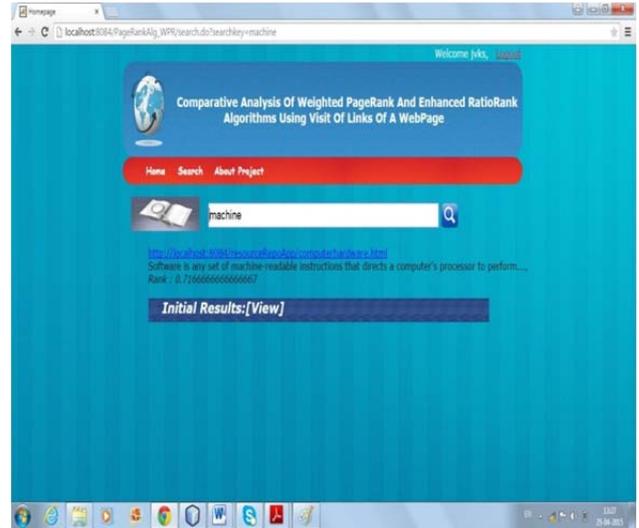
We can give a graphical representation which shows the variation between WPR (VOL) and WPR scores of different web pages of a web graph. In this we can observe that WPR (VOL) is performing better than original WPR. As the proposed algorithm calculates higher relevant score than the existing one.

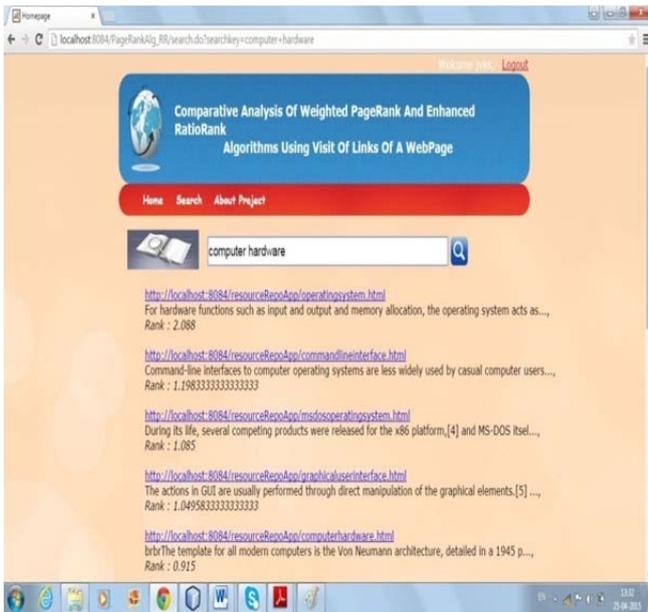
Table 3: Comparison of Weighted Page Rank based on VOL and Enhanced Ratio Rank Algorithms:

Keyword	RANKS	
	WPR vol	Enhanced RR
OS	1.85	2.088
CLI	1.0	1.198
MS DOS	1.0	1.085
GUI	0.575	1.049
CH	0.7166	0.915
SS	0.433	0.631

SAMPLE SCREENS:

The following are the sample screens of Weighted Page Rank Algorithm based on VOL and Enhanced-Ratio Rank algorithms displaying page rank values of more relevant web pages.





CONCLUSION

The weighted page rank algorithm is discussed which is more target-oriented. This weighted page rank algorithm calculates Page Rank value of web pages based on the visits of incoming links as well as the popularity of inlinks of a web page and takes into account ratio rank factors. It is not only consider link structure it includes the users focus on a particular page, but the main problem in this concept is calculation of visits of a links for that we have given a simple concept to monitor and count the hits or visits. User generally spends a lot of time in surfing through the search results to find the relevant pages. This analysis provides more efficient results than original WPR. The ordering of pages in this way increases the relevancy of pages and thereof provides the user with quality search results.

FUTURE SCOPE

Algorithms which use link structure has mainly many challenges like emphasis on old pages, theme drift, page cheating. This work can be further be enhanced by basing it on content rather than purely link structure or web structure. The page ranking algorithm which use web structure mining doesn't care about user's query, Only link structure of WebPages are considered in calculation of page rank value of WebPages. On the other side the page ranking algorithm which uses web content mining take user's query into account and doesn't care about link structure of WebPages for calculating page rank values of the WebPages.

REFERENCES

- [1] Neelam Tyagi, 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
- [2] Dilip Kumar Sharma, A. K. Sharma, "A Comparative Analysis of Web Page Ranking Algorithms" Dilip Kumar Sharma et al. / (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 08, 2010, 2670-2676
- [3] N. Duhan, A. K. Sharma and Bhatia K. K., "Page Ranking Algorithms: A Survey, Proceedings of the IEEE International Conference on Advance Computing, 2009, 978-1-4244-1888-6.
- [4] S. Brin, and Page L., "The Anatomy of a Large Scale Hypertextual Web Search Engine", Computer Network and ISDN Systems, Vol. 30, Issue 1-7, pp. 107-117, 1998.
- [5] Larry Page, and Sergey Brin, Rajeev Motwani, Terry Winograd, "The PageRank Citation Ranking: Bring Order to the Web", Technical report in Stanford U, 1998.
- [6] R. Cooley, B. Mobasher, and Srivastava, J., "Web Mining: Information and pattern discovery on the World Wide Web". In proceedings of the 9th IEEE International Conference on tools with Artificial Intelligence (ICTAI' 97).Newposrt Beach,CA 1997.
- [7] J. Kleinberg, "Hubs, Authorities and Communities", *ACM Computing Surveys*, 31(4), 1999.
- [8] Gyanendra Kumar, Neelam Duahn, and Sharma A. K., "Page Ranking Based on Number of Visits of Web Pages", International Conference on Computer & Communication Technology (ICCCCT)-2011, 978-1-4577-1385-9.
- [9] R. Kosala, and H. Blockeel, "Web Mining Research: A Survey", SIGKDD Explorations, Newsletter of the ACM Special Interest Group on Knowledge Discovery and Data Mining Vol. 2, No. 1 pp 1-15, 2000.
- [10] Wenpu Xing and Ghorbani Ali, "Weighted PageRank Algorithm", Proceedings of the Second Annual Conference on Communication Networks and Services Research (CNSR '04), IEEE, 2004.