International Journal of Mechatronics, Electrical and Computer Technology

Dec. 2013, Special Number, pp 422-436, **ISSN**: 2305-0543

Available online at: http://www.aeuso.org

A New Page Ranking Algorithm Based On WPR_{VOL} Algorithm

Roja Javadian Kootenae [1*], Seyyed Mohsen Hashemi [2] and mehdi afzali[3]

[1]MSc student, Department of Computer Engineering, Zanjan Branch, Islamic Azad University, Zanjan, Iran

[2]Assistant Professor, Science and Research Branch, Islamic Azad University, Tehran, Iran

[3]Assistant Professor, Department of Computer Engineering, Zanjan Branch, Islamic Azad University, Zanjan, Iran

*Corresponding Author's E-mail: *r.javadiankootenae@gmail.com*

Abstract

The amount of information on the web is always growing, thus powerful search tools are needed to search for such a large collection. Search engines in this direction help users so they can find their desirable information among the massive volume of information in an easier way. But what is important in the search engines and causes a distinction between them is page ranking algorithm used in them. In this paper a new page ranking algorithm based on "Weighted Page Ranking based on Visits of Links (WPR_{VOL}) Algorithm" for search engines is being proposed which is called WPR'_{VOL} for short. The proposed algorithm considers the number of visits of first and second level in-links. The original WPR_{VOL} algorithm takes into account the number of visits of first level in-links of the pages and distributes rank scores based on the popularity of the pages whereas the proposed algorithm considers both in-links of that page (first level in-links) and in-links of the pages that point to it (second level in-links) in order to calculation of rank of the page, hence more related pages are displayed at the top of search result list. In the summary it is said that the proposed algorithm assigns higher rank to pages that both themselves and pages that point to them be important.

Keywords: search engines, page ranking, ranking algorithms, in-link, WPR_{VOL} algorithm

1-Introduction

The Web is accessible to anyone via a Web browser. The World Wide Web can be seen as a large unstructured and ubiquitous database. Finding useful information on the Web is a very difficult task. It allows people to share information globally [1]. As the amount of information on the web is increasing day by day, information retrieval from it is also getting

International Journal of Mechatronics, Electrical and Computer Technology

Dec. 2013, Special Number, pp 422-436, **ISSN**: 2305-0543

Available online at: http://www.aeuso.org

more difficult so users need specific tools to find their required information till can access to relevant information with their query more quickly and easier, which is why they use search engines for this task. Search engine is the most efficacious tool to discover useable information in World Wide Web. Without Search Engine, there are no uses of information in websites, blog, etc; because without search engine, it is almost impossible to look for one by one websites just for searching information in internet [2].

A search engine in generally is a program that searches keywords in a document or database, but on the internet this points out a program that will search keywords in the files and web pages, newsgroups, Gopher menus and FTP archives and help to users in search of the desired information among the mass of information available on the network [3]. In fact, search engine is a program that intelligently classifies the data existing on the index and categorizes them based on importance, so that search result be more in accordance with and relevant to the requested words. Engine returns results of a query to the user. Generally the order of returned results is different for various search engines.

A search engine accepts queries from users and performs search on the available indexes. After retrieving search results from the indexes, the engine is also responsible for ranking the search results according to content analysis and link analysis scores. After all the processing, the engine generates and renders the search results and submits them to the user interface to be displayed to the user [5]. In general, a search engine usually consists of three major modules [5]: crawler, indexer and ranking.

A crawler is a module aggregating data from the World Wide Web in order to make them searchable [6]. Crawlers are also called robots, spiders, worms, wanderers, walkers, and know bots. There are two policies used to traverse Web pages: breath-first policy and depth-first policy. The crawlers transmit downloaded pages to a single Store Server process, which compress the pages and store them to disk. Then the indexer process reads pages from disk. It extracts links from the pages and saves them to a different disk file [1].

Indexing is organizing contents of pages in the way that allows effective retrieval. In fact, it is a module that takes the collection of documents or data and makes searchable index from

Dec. 2013, Special Number, pp 422-436, **ISSN**: 2305-0543

Available online at: http://www.aeuso.org

them. There are different strategies for indexing such as trie and patricia, inverted file, compression and distribution [7]. The indexer takes the web pages collected by the spiders and parses them into a highly efficient index [4]. In fact, the indexer splits the page into its parameters and converts whole parameters to the numerical scale till ranking system can compares different pages parameters with each other. Then all of analyzed data in the indexer is submitted to the database in order to group, code and save them. Also before the data is saved, it is compressed using specific technique until it occupies a small volume of the database. Then ranking is done in order to sort results according to their importance of the entered query, send to interface user and display to the user [8].

Among these components, ranking has specific importance and it is referred to as the heart of a search engine. The query evaluation heavily depends on the index structure used. If a query returns a really large set of documents, the user will not be able to recognize the results relevant for his information needs. Most IR systems therefore employ some kind of ranking mechanism. This provides the user with a pre selection of results by automatically calculating the relevance of each document. Afterwards it is possible to select just the best results. Sometimes relevance information is already available from the query evaluation, for example through the membership function from the fuzzy set model [7].

So far different ranking algorithms to rank web pages have been proposed but each has limitations and problems, therefore in this paper we plan to offer a new page ranking method based on existing algorithms so that the ranking operation is executed more carefully and displays the most relevant pages at the top of search result list for the user. The purpose of this paper is to analyze some of the currently important algorithms and propose a better, faster and more accurate algorithm for ranking of the web pages in search engines. The remainder part of this paper is organized as follows:

Related work is summarized in Section 2. New proposed method based on WPRVOL algorithm is described in section 3 and is compared with previous method by providing an example its advantages is expressed. Finally a conclusion is presented in Section 4.

Dec. 2013, Special Number, pp 422-436, **ISSN**: 2305-0543
Available online at: http://www.aeuso.org

1. Related works

In the following, we will describe web pages ranking algorithm that are the background of the proposed algorithm for short:

1.1. Page Rank Algorithm (PR)

Page Rank algorithm is used by one of the most popular search engines, Google. This algorithm has proposed by Brin and Page based on hyperlink structure in 1998. Page Rank computes a ranking for every web page based on a graph of the web. The Page Rank of a page is defined recursively and depends on the number and Page Rank metric of all pages that link to it (in-links). A page that is linked by many pages with high rank receives a high rank itself. If there are no links to a web page there is no support for this specific page. A simplified version of Page Rank is given in Eq. (1) [5]:

$$PR(u) = c \sum_{v \in Bu} \frac{PR(v)}{N_v} \quad (1)$$

Where u represents a web page, B_u is the set of pages that point to u, $PR(u)$ and $PR(v)$ are rank scores of page u and v respectively, N_v indicates the number of out-links of page v, c is a factor applied for normalization [5,9,10,17]. The rank of a page is divided among its forward links evenly to contribute to the ranks of the pages they point to. The equation is recursive. But there is a problem with this simplified function. If there were two web pages that point to each other but to no other page while some other web page points to one of them, a loop will be generated during the iteration.

This loop will accumulate the rank but will never distribute any ranks. This trap formed by loops in a graph without outedges are called rank sinks. To avoid rank sinks a model of a Web surfer has to be created. This surfer simply keeps clicking on hyperlinks at random. The task is modeling the behavior that the surfer periodically gets bored and jumps to a random page [9]. Later Page Rank was customized observing that not all users follow the direct links on WWW. The modified version is given in Eq. 2 [11,12,15]:

Dec. 2013, Special Number, pp 422-436, **ISSN**: 2305-0543
Available online at: http://www.aeuso.org

$$PR(u) = (1-d) + d\sum_{i=1}^{n}\frac{PR(p_i)}{C(p_i)} \quad (2)$$

Where *d* is a dampening factor that is frequently set to 0.85. *d* can be thought of as the prospect of users' following the direct links and *(1 –d)* as the page rank distribution from non-directly linked pages. p_i are pages that point to *u*, $C(p_i)$ indicates the number of out-links of page p_i. The equation may be computed by starting with any set of ranks and iterating the computation until it converges [10], in other words, values does not change.

1.2. Weighted Page Rank Algorithm (WPR)

Weighted Page Rank Algorithm [13] is proposed by Wenpu Xing and Ali Ghorbani in 2004. Weighted Page Rank algorithm (WPR) is the modification of the original page rank algorithm. WPR decides the rank score based 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 more popular pages and does not equally divide the rank of a page among its out-link pages. Every out-link page is given a rank value based on its popularity. Popularity of a page is decided by observing its number of in-links and out-links [17]. The popularity from the number of in-links and out-links is recorded as $W^{in}_{(v,u)}$ and $W^{out}_{(v,u)}$, respectively. $W^{in}_{(v,u)}$ given in Eq. (3) 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* [12,15].

$$W^{in}_{(v,u)} = \frac{I_u}{\sum_{p\in R_v} I_p} \quad (3)$$

Where I_u and I_p represent the number of in-links of page *u* and page *p*, respectively. R(v) denotes the reference page list of page *v*. $W^{out}_{(v,u)}$ given in Eq. (4) is the weight of *link(v, u)* calculated based on the number of out-links of page u and the number of out-links of all reference pages of page *v* [12,15].

$$W^{out}_{(v,u)} = \frac{O_u}{\sum_{p\in R_v} O_p} \quad (4)$$

International Journal of Mechatronics, Electrical and Computer Technology

Dec. 2013, Special Number, pp 422-436, **ISSN**: 2305-0543

Available online at: http://www.aeuso.org

Where O_u and O_p represent the number of out-links of page *u* and page *p*, respectively. R(v) denotes the reference page list of page *v*. Considering the importance of pages, the original Page Rank formula is modified in Eq. (5) as: [12,15]

$$WPR(u) = (1 - d) + d \sum_{v \in B_u} WPR(v)\, W_{(v,u)}^{in}\, w_{(\mathrm{v,u})}^{\mathrm{out}} \qquad (5)$$

Simulation of WPR is done using the Website of Saint Thomas University and simulation results show that WPR algorithm finds larger number of relevant pages compared to standard page rank algorithm. future work include to calculate the rank score by utilizing more than one level of reference page list and increasing the number of human user to classify the web pages [17].

2.3. Page Ranking based on Visits of links (PR_{VOL})

Gyanendra Kumar et. al. [14] proposed a new algorithm in 2011 in which they considered user's browsing behavior. As most of the ranking algorithms proposed are either link or content oriented in which consideration of user usage trends are not available. In this paper, a page ranking mechanism called Page Ranking based on Visits of Links (PR_{VOL}) is being devised for search engines, which works on the basic ranking algorithm of Google, i.e. Page Rank and takes number of visits of in-links of web pages into account.

This concept is very useful to display most valuable pages on the top of the result list on the basis of user browsing behavior, which reduce the search space to a large scale. In this paper as the author describe that in the original Page Rank algorithm, the rank score of page *p*, is evenly divided among its out-links or we can say for a page, an in-links brings rank value from base page, *p*. So, he proposed an improved Page Rank algorithm. In this algorithm they assign more rank value to that out-links which is most visited by users. In this manner a page rank value is calculated based on visits of in-links. The modified version based on VOL is given in eq. (6) [15]:

International Journal of Mechatronics, Electrical and Computer Technology

Dec. 2013, Special Number, pp 422-436, **ISSN**: 2305-0543

Available online at: http://www.aeuso.org

$$PR_{VOL}(u) = (1 - d) + d \sum_{v \in B_u} \frac{L_u \ PR_{VOL}(v)}{TL(v)} \quad (6)$$

Where *d* is a dampening factor, *u* represents a web page, B_u is the set of pages that point to *u*, $PR_{VOL}(u)$ and $PR_{VOL}(v)$ are rank scores of page *u* and *v* respectively, L_u is the number of visits of link which is pointing page *u* from *v*. *TL(v)* denotes total number of visits of all links present on *v* [15].

2.4. Weighted Page Ranking based on Visits of links algorithm (WPR_{VOL})

We have seen that the original Weighted Page Rank algorithm assigns larger rank values to more important (popular) pages. Each out-link page gets a value proportional to its popularity (its number of in-links and out-links). The popularity from the number of in-links and out-links is recorded as $W^{in}_{(v,u)}$ and $W^{out}_{(v,u)}$, respectively. N. Tyagi, S. Sharma proposed an improved Weighted Page Rank algorithm in 2012. In this algorithm they assign more rank value to the out-links which is most visited by users and received higher popularity from number of in-links. They do not consider here the popularity of out-links 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. The modified version based on WPR_{VOL} is given in eq. (7) [15]:

$$WPR_{VOL}(u) = (1 - d) + d \sum_{v \in B_u} \frac{L_u \ WPR_{VOL}(v) \ W^{in}_{(v,u)}}{TL(v)} \quad (7)$$

Where *d* is a dampening factor, *u* represents a web page, B_u is the set of pages that point to *u*, L_u is the number of visits of link which is pointing page *u* from *v*. *TL(v)* denotes total number of visits of all links present on *v* and $W^{in}_{(v,u)}$ is 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* [15].

The algorithm will be used recursively until the values are to be stable. Authors believe that the proposed algorithm is better than WPR algorithm as the new algorithm will calculate

International Journal of Mechatronics, Electrical and Computer Technology

Dec. 2013, Special Number, pp 422-436, **ISSN**: 2305-0543

Available online at: http://www.aeuso.org

more relevant web pages than that of existing one. As in the proposed algorithm they assign more weightage to those web pages which is most visited by users, also provide more weightage to those web page which have higher popular in-links. These two considerations will make the proposed algorithm performs better than original one. The rank value of any page by original Weighted Page Rank 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 in-links and out-inks. While the ordering of pages using WPR_{VOL} is more target-oriented [15].

2. New proposed algorithm based on WPR_{VOL} Algorithm

The proposed method in this paper is improving WPR_{VOL} algorithm that call second level WPR_{VOL} algorithm and show with WPR'_{VOL}. The proposed algorithm is based on WPR_{VOL} algorithm. The idea used in proposed algorithm is that for calculation of rank score is utilized from more than one level of pointing page list (reference). In fact, in calculation of rank of a page, also rank of pointing pages to that page is considered, which means we do the calculation up to the second level.

The original WPR_{VOL} algorithm considers only number of visits of first level in-links and distributes ranking score based on it among web pages, where the proposed algorithm in this paper to calculate the rank of a page takes into account the importance of both the in-links to that page (the first level in-links) and pointing pages in-links to that page (the second level in-links), thus related pages are placed at the top of search result list. In short, it is said that proposed algorithm assigns higher rank to the pages that both themselves and pointing pages to them be important.

WPR'_{VOL} algorithm calculates rank of a page using Eq. (8):

$$WPR'_{VOL}(u) = (1-d) + d \sum_{v \in B_u} \frac{L_u \; WPR'_{VOL}(v) \; W^{in}_{(v,u)}}{TL(v)} WPR_{VOL}(v) \tag{8}$$

Where d is a dampening factor, u represents a web page, B_u is the set of pages that point to u, L_u is the number of visits of link which is pointing page u from v. $TL(v)$ denotes total

International Journal of Mechatronics, Electrical and Computer Technology

Dec. 2013, Special Number, pp 422-436, **ISSN**: 2305-0543

Available online at: http://www.aeuso.org

number of visits of all links present on v , $W^{in}_{(v,u)}$ is 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 and $WPR_{VOL}(v)$ calculate using Eq. (7) in any repeat. The advantages of WPR'$_{VOL}$ algorithm to WPR$_{VOL}$ algorithm are that this algorithm converges faster and also in the calculation of rank of pages enjoys higher accuracy and better results quality and also prevents from increasing the aberrant rank. This algorithm doesn't have rich get richer problem. This algorithm has a defect; it needs rather long computation to calculate rank of a page.

2.1. Simple example illustrating working of WPR'$_{VOL}$

To explain the working of original and proposed algorithms, consider simple graph shown in fig 1. We want calculate the rank of pages using WPR$_{VOL}$ and WPR'$_{VOL}$ algorithms.

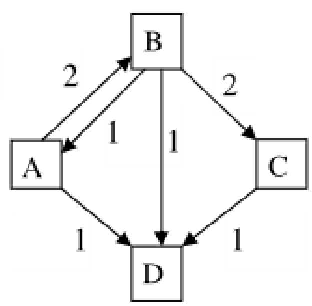

Figure 1. Hyper linked graph with link visits

At first the rank of pages is calculated using WPR$_{VOL}$ algorithm.

$$W^{in}_{(A,B)} = \frac{I_B}{I_B} = 1 \qquad W^{in}_{(B,A)} = \frac{I_A}{I_A} = 1$$

$$W^{in}_{(B,C)} = \frac{I_C}{I_A} = \frac{1}{1} = 1 \qquad W^{in}_{(A,D)} = \frac{I_D}{I_B} = \frac{3}{1} = 3$$

$$W^{in}_{(B,D)} = \frac{I_D}{I_A} = \frac{3}{1} = 3 \qquad W^{in}_{(C,D)} = \frac{I_D}{I_B} = \frac{3}{1} = 3$$

International Journal of Mechatronics, Electrical and Computer Technology

Dec. 2013, Special Number, pp 422-436, **ISSN**: 2305-0543

Available online at: http://www.aeuso.org

$$\frac{L_B}{\mathrm{TL(A)}} = \frac{2}{3} \qquad \frac{L_D}{\mathrm{TL(A)}} = \frac{1}{3}$$

$$\frac{L_A}{\mathrm{TL(B)}} = \frac{1}{4} \qquad \frac{L_C}{\mathrm{TL(B)}} = \frac{2}{4} = \frac{1}{2}$$

$$\frac{L_D}{\mathrm{TL(B)}} = \frac{1}{4} \qquad \frac{L_D}{\mathrm{TL(C)}} = \frac{1}{1} = 1$$

$$WPR_{VOL}(A) = (1-d) + d(\frac{L_A \ WPR_{VOL}(B) \ W^{in}_{(B,A)}}{TL(B)})$$

$$WPR_{VOL}(B) = (1-d) + d(\frac{L_B \ WPR_{VOL}(A) \ W^{in}_{(A,B)}}{TL(A)})$$

$$WPR_{VOL}(C) = (1-d) + d(\frac{L_C \ WPR_{VOL}(B) \ W^{in}_{(B,C)}}{TL(B)})$$

$$WPR_{VOL}(D) = (1-d) + d(\frac{L_D \ WPR_{VOL}(A) \ W^{in}_{(A,D)}}{TL(A)} + \frac{L_D \ WPR_{VOL}(B) \ W^{in}_{(B,D)}}{TL(B)} + \frac{L_D \ WPR_{VOL}(C) \ W^{in}_{(C,D)}}{TL(C)})$$

Then calculate all these values iteratively until values become stable. Table 1 is formed using above calculation (whit $\varepsilon = 10^{-4}$, d = 0.85):

International Journal of Mechatronics, Electrical and Computer Technology

Dec. 2013, Special Number, pp 422-436, **ISSN**: 2305-0543

Available online at: http://www.aeuso.org

Table 1. Shows WPR_{VOL} values

WPR_{VOL} (A)	WPR_{VOL} (B)	WPR_{VOL} (C)	WPR_{VOL} (D)
0.3625	0.716666666	0.575	4.1875
0.30229166	0.355416666	0.454583333	2.38125
0.225526041	0.312298610	0.301052083	1.792713539
0.218275954	0.277798089	0.286551909	1.314207810
0.209032093	0.273689707	0.268064187	1.243338211
0.208159062	0.268451519	0.266318125	1.185718144
0.207045947	0.267956801	0.264091895	1.177184265
0.206940820	0.267326036	0.263881640	1.170245848
0.206806782	0.267266464	0.263613565	1.169218227
0.206794123	0.267190509	0.263588247	1.168382726
0.206777983	0.267183336	0.263555966	1.168258984
0.206776458	0.267174190	0.263552917	1.168158376
0.206774515	0.267173326	0.263549030	1.168143474

Now the rank of pages is calculated using WPR'_{VOL} algorithm.

$$WPR'_{VOL}(A) = (1-d) + d\left(\frac{L_A\ WPR'_{VOL}(B)\ W^{in}_{(B,A)}}{TL(B)} \times WPR_{VOL}(B)\right) \Rightarrow$$

$$WPR'_{VOL}(A) = (1-d) + d[(\frac{L_A\ WPR'_{VOL}(B)\ W^{in}_{(B,A)}}{TL(B)}) \times ((1-d) + d\,\frac{L_B\ WPR_{VOL}(A)\ W^{in}_{(A,B)}}{TL(A)})]$$

$$WPR'_{VOL}(B) = (1-d) + d\left(\frac{L_B\ WPR'_{VOL}(A)\ W^{in}_{(A,B)}}{TL(A)} \times WPR_{VOL}(A)\right) \Rightarrow$$

International Journal of Mechatronics, Electrical and Computer Technology

Dec. 2013, Special Number, pp 422-436, **ISSN**: 2305-0543

Available online at: http://www.aeuso.org

$$WPR'_{VOL}(B) = (1-d) + d[(\frac{L_B\ WPR'_{VOL}(A)\ W^{in}_{(A,B)}}{TL(A)}) \times ((1-d) + d\frac{L_A\ WPR_{VOL}(B)\ W^{in}_{(B,A)}}{TL(B)})]$$

$$WPR'_{VOL}(C) = (1-d) + d\left(\frac{L_C\ WPR'_{VOL}(B)\ W^{in}_{(A,C)}}{TL(B)} \times WPR_{VOL}(B)\right) \quad \Rightarrow$$

$$WPR'_{VOL}(C) = (1-d) + d[(\frac{L_C\ WPR'_{VOL}(B)\ W^{in}_{(B,C)}}{TL(B)}) \times ((1-d) + d\frac{L_B\ WPR_{VOL}(A)\ W^{in}_{(A,B)}}{TL(A)})]$$

$$WPR'_{VOL}(D) = (1-d) + d\left(\begin{array}{c}\frac{L_D\ WPR'_{VOL}(A)\ W^{in}_{(A,D)}}{TL(A)} \times WPR_{VOL}(A) + \\ \frac{L_D\ WPR'_{VOL}(B)\ W^{in}_{(B,D)}}{TL(B)} \times WPR_{VOL}(B) + \\ \frac{L_D\ WPR'_{VOL}(C)\ W^{in}_{(C,D)}}{TL(C)} \times WPR_{VOL}(C)\end{array}\right) \quad \Rightarrow$$

$$WPR'_{VOL}(D) = (1-d) + d[$$

$$\left(\frac{L_D\ WPR'_{VOL}(A)\ W^{in}_{(A,D)}}{TL(A)}\right) \times \left((1-d) + d\frac{L_A\ WPR_{VOL}(B)\ W^{in}_{(B,A)}}{TL(B)}\right)$$

$$+\left(\frac{L_D\ WPR'_{VOL}(B)\ W^{in}_{(B,D)}}{TL(B)}\right) \times \left((1-d) + d\frac{L_B\ WPR_{VOL}(A)\ W^{in}_{(A,B)}}{TL(A)}\right)$$

$$+(\frac{L_D\ WPR'_{VOL}(C)\ W^{in}_{(C,D)}}{TL(C)}) \times \left((1-d) + d\frac{L_C\ WPR_{VOL}(B)\ W^{in}_{(B,C)}}{TL(B)}\right)]$$

With placement and repeat the algorithm Table 2 is reached.

Dec. 2013, Special Number, pp 422-436, **ISSN**: 2305-0543

Available online at: http://www.aeuso.org

Table 2. Shows WPR'$_{VOL}$ values

WPR'$_{VOL}$ (A)	WPR'$_{VOL}$ (B)	WPR'$_{VOL}$ (C)	WPR'$_{VOL}$ (D)
0.302291666	0.355416666	0.454583333	2.38125
0.174266412	0.188632297	0.198532824	0.629723493
0.159971024	0.168771014	0.166942048	0.324594513
0.158630642	0.166848603	0.167261284	0.297251013
0.158505403	0.166670708	0.167010806	0.294744262
0.158493821	0.166654151	0.166987642	0.294511382
0.158492745	0.166652618	0.166985490	0.294489814
0.158492645	0.166652476	0.166985290	0.294487812

Table 3 shows comparison between rank values of WPR'$_{VOL}$ and WPR$_{VOL}$ algorithms for graph in Fig. (1).

Table 3. Comparison between rank values of WPR'$_{VOL}$ and WPR$_{VOL}$ algorithms

	(A)	(B)	(C)	(D)
WPR$_{VOL}$	0.2067	0.2671	0.2635	1.1681
WPR'$_{VOL}$	0.1584	0.1666	0.1669	0.2944

As the results show the order of pages rank in the graph in Figure 1 for WPR$_{VOL}$ algorithm is A<C<B<D and for WPR'$_{VOL}$ algorithm is A<B<C<D. Here we find out that our proposed algorithm (WPR'$_{VOL}$) can do better and also converges faster and will set more relevant web pages at top of the result list. As we see in this example, it is true that in-links of B and C pages are equal but with consideration second level in-links we conclude that C page more visited by users than B page thus shall C page has higher rank score. Then we realized that our proposed algorithm can rank the pages better than previous algorithms. Because of the calculation of page rank assign the higher scores to pages that were actually visited by users

Dec. 2013, Special Number, pp 422-436, **ISSN**: 2305-0543
Available online at: http://www.aeuso.org

more due to considering the popularity of both the in-links to themselves and pointing pages in-links to them. These considerations will make the proposed algorithm performs better and also has more accuracy in page ranking.

Conclusion

Search engines typically provide a search interface based on text keywords: a user enters some descriptive keywords after which web pages containing these keywords are returned [16]. This is done so that the crawler traverses the Web and collects the documents. Each document is then processed by the indexer, which inserts the index terms into the index. The user enters a query through the interface, which is then processed by the query engine. The engine uses the index to retrieve the documents and applies a ranking mechanism. The (ordered) list of results is then presented to the user through the interface [7]. Ranking component is the most important part of search engine. There are many algorithms for ranking of pages but each one has disadvantages, thus we have always been trying to improve the algorithms. The proposed method in this paper is improving WPR_{VOL} algorithm that is called the second level WPR_{VOL} algorithm and displays it as WPR'_{VOL} for short. The idea used in the proposed algorithm is that for calculation the rank score is utilized from more than one level of pointing page list (reference). In fact, in calculation of the rank of a page, also ranks of pointing pages to that page is considered, which means we are doing calculation up to the second level. The consideration of two levels of in-links causes our calculation to be more accurate and also related pages are placed at the top of search result list. WPR'_{VOL} algorithm gives higher rank scores to pages that are important both themselves and pointing pages to them, namely they have been visited more by users.

Reference

[1] S. Lam, The Overview of Web Search Engines University of Waterloo, Canada, 2001.

[2] DING Jian-li, YANG Bo; A New Model of Search Engine based on Cloud Computing, International Journal of Digital Content Technology and its Applications. June 2011, Vol. 5, Num. 6.

[3] http://www.20tools.com/search/?etc=WhatsSearchRobots

[4] J. D. King, Search Engine Content Analysis Queensland University of Technology, Brisbane, Australia, phd_thesis, 2008

[5] S. Rahimi, B. Gupta, K. Adya. A Novel Page Ranking Algorithm for Search Engines Using Implicit Feedback Engineering Letters, Advance online publication, Southern Illinois University, USA, 2006.

[6] Knut Magne Risvik, scaling internet search engine: methods and analysis phd thesis, 2003

[7] T. B. Hubel, The Holumbus Framework Creating fast, flexible and highly customizable search engines with Haskell Fhwedel University of applied sciences, Department of Computer Science, Master's Thesis, 2008.

[8] http://www.irpdf.com/book-5604.html

[9] F. Hertel, PageRank – Bringing Order to the Web Seminar Report, School of Computer Science, Institute of Knowledge and Language Engineering, Otto-von-Guericke-Universit¨at Magdeburg, 2004.

[10] S. Brin, R. Motwani, L. Page, T.Winograd, The PageRank Citation Ranking: Bringing Order to the Web, Technical Report, 1998.

[11] S. Brin, and L. Page, The Anatomy of a Large Scale Hypertextual Web Search Engine Computer Network and ISDN Systems, 1998, Vol. 30, Issue 1-7, pp. 107-117.

[12] A. Kumar Singh, R. Kumar P, A Comparative Study of Page Ranking Algorithms for Information Retrieval International Journal of Electrical and Computer Engineering 4:7 2009.

[13] W. Xing and Ali Ghorbani, Weighted PageRank Algorithm Proc. of the Second Annual Conference on Communication Networks and Services Research (CNSR '04), IEEE, 2004.

[14] G. Kumar, N. Duahn, A. K Sharma, Page Ranking Based on Number of Visits of Web Pages International Conference on Computer & Communication Technology (ICCCT)-2011, 978-1-4577-1385-9.

[15] N. Tyagi, S. Sharma, Weighted Page Rank Algorithm Based on Number of Visits of Links of Web Page International Journal of Soft Computing and Engineering (IJSCE), July 2012, Vol. 2, Issue. 3.

[16] P. Min, A 3D Model Search Engine Princeton University in Candidacy, phd thesis, 2004.

[17] D.k. Sharma, A.K. Sharma, A Comparative Analysis of Web Page Ranking Algorithms International Journal on Computer Science and Engineering (IJCSE), 2010, Vol. 02, 2670-2676.