International Journal of Information Science and Management

Structural Investigation of Websites' Ranking Systems in view of Information Retrieval

J. Mehrad, Ph.D.
President of Regional Information Center for Science and Technology, I. R. of Iran
email: dean@srlst.com

A. Shemrani, M.S.
Research Department of System Design and Operations, Regional Information Center for Science and Technology
Corresponding Author: shemirani@srlst.com

Abstract

For years, search engines have been considered as one of the most frequently used information seeking tools through the web. Efficiency, ease of use and search quality are the main factors for giving precedence over search engines. Search quality is assessed based on the concept of page rank which is applied to scoring web documents. The structures designed for linkage among the pages of a site as well as the linkage structure among the sites inside a web are known as the most effective factor on page rank. This paper is to describe the concept of ranking web pages and investigate the structure of inter-linkage of pages.

Keywords: Ranking, Page Rank, Information Retrieval, Search Engines.

Introduction

In the process of ranking web pages, amount of the relevance among the content of web pages and search subject is the first thing investigated by search engines. Then based on priority of importance, the retrieved pages could be ranked by the rank page factor. Analyzing the current algorithms within the search engines, the page rank factor and those other factors causing increase and/or decrease in the rank of pages will be described concerning ranking of pages. This article, first, offers an introductory on how to measure the page rank and then, shows how the factors such as the inbound and/or outbound links of a page as well as the number of whole links could affect the websites to be positioned at the top or the low ranks based on the results retrieved by search engines. Therefore, those websites with more unreal and useless links will be placed at the low levels of ranking. We are to investigate the effective factors on making the optimum use of the internal capacity of websites. Such investigations will contribute to better identifying the websites relevant to the search subjects. This paper could help websites' managers, software companies and website designers in designing information retrieval systems and doing research on search engines.

Method

What is Page Rank?

Page rank means to allocate a numeric degree to the web pages the amount of which demonstrates the importance and priority of a page among all retrieved pages. Inter-pages link increases the value of the page receiving the link. The more the value of a page, the more its importance is (Brin & Page, 1998). The weight of the page also refers to the importance of the value itself. Google as a search engine offers the definition of page ranking as "the way of appointing the significance of a page within the web" (Brin & Page, 1998). It should be noted that some of the links among all the links inside a page are not considered in the ranking process. Such links could cause negative influence on the significance of a page and reduce its value in ranking (Mehrad & Binesh, 2007). Although there is no possibility to control others on making external links to our website, we are able to steer the links made by our site to other websites. It is worth mentioning that making links to the websites with low ranks could bring our website bad rankings from search engines. In other words, paying attention to the rank of the goal website should be considered as a must in making links (Jeh & Widom, 2002).

How to Calculate the Page Rank?

All the links entering the page, whether outside or inside the site, have to be taken into consideration in calculating the page rank. Brin and Page (1998) offered, for the first time, the algorithm for achieving the page rank. The following formula is used to calculate the page rank:

$$PR(A) = (1- d) + d(PR(t1)/C(t2)) + \ldots + PR(tn)/C(tn)$$

The current formula is the first one documented by the search engines for ranking pages in searches. Here, t1-tn represents those pages have made links to the page A, C stands for the number of external links entering each page, and d is a moderator factor being usually assumed equal to the fixed amount of 0.85. Each page increases the rank of the target page receiving the link while this linkage does not affect the rank of the page making the link. The total of increase in ranking caused by any page for other pages is a bit less than the rank of the link-maker page; that is 85% of the rank of the link-maker page. This amount will be equally shared among all the pages receiving the link.

Therefore, it could be stated that receiving links from a page with the page rank of 4 and enjoyment of 5 links made to other pages is more beneficial than receiving links from a page with the page rank of 8 and embodying 100 links made to other pages. The rank of the page making link to your page is a matter of great importance but we cannot ignore the importance of the number of outbound links of this page. The more the

number of outbound links of the page, the less our share will be from the page rank of the link-maker page. It should be noted that making link to other pages does not cause decrease in the ranking of the link-maker page. This linkage is not a transmission of page rank from one page to another (Dwork, Kumar, Naor & Sivakumar, 2001). This is as same as voting to someone in which the status of the voter is a matter of great importance.

In order to calculate the ranking of a page, its current rank –if available– will be totally ignored and a new calculation will be carried out in which the page rank is specified only based on the views of other pages about the page receiving the links. That is so because the number of links entering the page may change after the last round of calculating the rank.

Computing rank based on the above formula for one time dose not give the reliable value for ranking. As an example, you can suppose a link made from page A to page B and, conversely, a link from page B to page A while there is no link from outside to these pages. Under such condition, the rank of pages A and B will be computed through the following stages:

Step 1: To calculate the page rank of A based on the number of inbound links.

As links enter page A, this page gains a new value of ranking. The links made from B to A are used in calculating the rank of A while because of accepting link from A, the new rank of B should also be computed. Therefore, computing the new rank of A is done based on inaccurate data and consequently the result cannot be reliable.

Step 2: To calculate the rank of page B based on the number of inbound links.

As links enter page B, this page gets a new value in ranking but this value cannot be useful because the inaccurate number of the links coming from page A has been used in computing the page rank of B.

It is not possible to calculate the page rank of A without knowing the page rank of B and vice versa. The only way here is to do calculations by new quantities the result of which is to achieve a page rank more accurate than the previous method. As the data used in the formula lack of adequate accuracy, the result will also be incorrect. To solve such a problem, the computations should be repeated for many times with new quantities which eventually give more exact results. At least, 40 to 50 times of calculations are necessary to achieve the results with expected accuracy. The large number of pages and the necessity to repeat the calculations are the reasons the search engines keep the information up to date at long intervals.

There will be given a numeric example to investigate the issue. Suppose a site including three pages of A, B and C. The page A has made links to B and C. The page B has made link to C and the page C has made link to A (Figure 1). In order to have easy calculations, the moderator factor will be considered equal to 0.5. The initial quantity of

page rank will be assumed as 1 and it is emphasized that the initial value of rankings has no effect on the final result of calculations; "this first quantity may cause affects on the preliminary stages of computations but such effectiveness will no longer stay after repetition of calculations" (Efactory, 2003). According to the formula, the page ranks of A, B and C will be calculated as following:

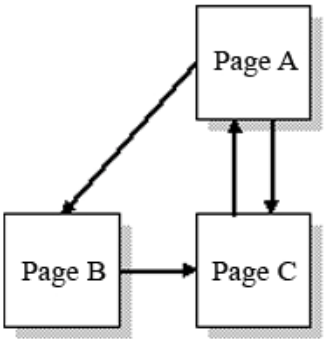

Figure 1. The site consisting of 3 pages making links to each other.

- The initial value of page rank has been assumed equal to 1

  Page A = 1

  Page B = 1

  Page C = 1

- The quantity of page rank after one stage of calculations

  Page A = (1 – 0.5) + 0.5 * (page rank of C / the number of outbound links of page C)

  Page A = 0.5 + 0.5 * (1 / 1) = 1

  Page B = (1 – 0.5) + 0.5 * (page rank of A / the number of outbound links of page A)

  Page B = 0.5 + 0.5 * (1 / 2) = 0.75

  Page C = (1 – 0.5) +

  0.5 * (page rank of A / the number of outbound links of page A + page rank of B / the number of outbound links of page B)

  Page C = 0.5 + 0.5 * (1 / 2 + 0.75 / 1) = 1.125

- New calculations according to new values of page ranks

  Page A = (1 – 0.5) + 0.5 * (page rank of C / the number of outbound links of page C)

  Page A = 0.5 + 0.5 * (1.125 / 1) = 1.0625

  Page B = (1 – 0.5) + 0.5 * (page rank of A / the number of outbound links of page A)

  Page B = 0.5 + 0.5 * (1.0625 / 2) = 0.765625

Page C = (1 – 0.5) +

0.5 * (page rank of A / the number of outbound links of page A + page rank of B / the number of outbound links of page B)

Page C = 0.5 + 0.5 * (1.0625 / 2 + 0.765625 / 1) = 1.1484375

- Repetition of calculations by new values of rankings during each stage gives the results shown in Table 1.

Table 1

The Results of Calculating Page Ranks within 50 Times of Computing the Values

Rank of page C	**Rank of page B**	**Rank of page A**	Repetition Level
1	1	1	0
1.125	.75	1	1
1.1484375	0.765625	1.0625	2
1.15283203	0.76855469	1.07421875	3
1.15365601	0.76910400	1.07641602	4
1.15381050	0.76920700	1.07682800	5
1.15383947	0.76922631	1.07690525	6
1.15384490	0.76922993	1.07691973	7
1.15384592	0.76923061	1.07692245	8
1.15384611	0.76923074	1.07692296	9
1.15384615	0.76923076	1.07692305	10
.	.	.	.
.	.	.	.
1.1538461538	0.7692307692	1.0769230769	50

If the initial value of pages is assumed equal to 0, the results after 50 times of calculations will be as follows:

Page A = 1.0769230769

Page B = 0.7692307692

Page C = 1.1538461538

According to the results, it could be stated that a) the supposed initial value has no effect on the final results. b) In order to gain an accurate and reliable value, the formula must be repeated by new quantities at each stage of calculation.

Effective Factors on Calculating Page Rank

As it was seen in the formula specifying for calculation of the page rank, the method in linkage among pages causes most affects on the page rank of a site. Table 2 briefly introduces the most effective factors on page rank. These factors will be

investigated in detail during the paper.

Table 2

The Effective Factors on Page Rank Based on the Linkage Structure

Name of Factor	Description
Internal links	Internal links factor represents the way inter-page links within a site affect the rank of each of pages and the impact of the method of linkage on the whole rank of the website
Inbound links	This factor stands for the way one link affects only the page receiving the link or the way it influences upon a collection of pages
Outbound links	This factor shows the influence of the number of outbound links on the rank of the page making the links
Dangling links	This factor stands for making links to the pages which have made no links to other pages
Number of pages	Number of pages shows the impact of either adding new pages to or deleting pages from a website, on the rank of other pages
BadRank	This factor stands for the concept of decrease in rank of pages and incurring penalties for linkage to the spam pages

Other effective criteria regarding the structure and content of web pages are listed in Table 3.

Table 3

Effective Factors on Page Rank based on the Structure and Content of Web Pages

Name of factors	Description
Subdirectories	This factor represents the hierarchical structure of sorting the pages inside a website
Content analysis	This factor stands for investigating the accuracy of links made among the pages based on the content of both link-maker pages and pages receiving the links
Domain	The domain factor regards the impact of the method used for selecting URLs, on the page rank
Google Bomb	This factor concerns making virtual (unreal) links among pages in order to deceive search engines in page ranking

Internal links

The highest amount of page rank within a site equals to the number of available pages on the site multiplied by 1. Increase in the number of pages of a site will cause increase in the amount of page rank. It means that the number of the pages of a site

conveys a direct correlation with the rank given to the site. In other words, the more the number of pages and the more accurate the linkage amongst pages are, the more the final rank of site will promote (Hirai, Raghavan, Garcia-Molina, & Paepcke, 1999). This cannot be ignored that making weak links among the pages will end in not achieving the highest rank for the site. On the other hand, adding new pages into the site is advised for increasing the page rank of whole site. But it deserves noting that some kinds of pages will decrease the value of the site. Those pages with identical or the same content could be mentioned as an example. Search engines call such pages "Spam". Presence of these spam pages could result in assigning penalties for the site.

It should be notified that pages could be indexed in search engines such as Google only if they have already received link(s) at least from one page in www. There will be presented some examples of how to calculate the page rank of a site in different modes of linkage among its pages. Two kinds of modes are considered by the search engines in scaling the page rank: Simple mode and Real mode. [Ignoring the presence of inbound or outbound links,]. In the simple mode, it is assumed that all pages are indexed in Google. Based on the real mode, those pages without inbound links will not be covered by the indexing system of the search engines. The samples presented in this paper have been considered in the simple mode. The software of "markhorrel" (Horrell, 2008) and "WebWorkshop" (WebWorkshop, 2008) were used to make and repeat calculations of page rank.

As an example, we suppose five pages of A, B, C, D and E available at one site in which none of these pages have received any links from outside (Figure 2).

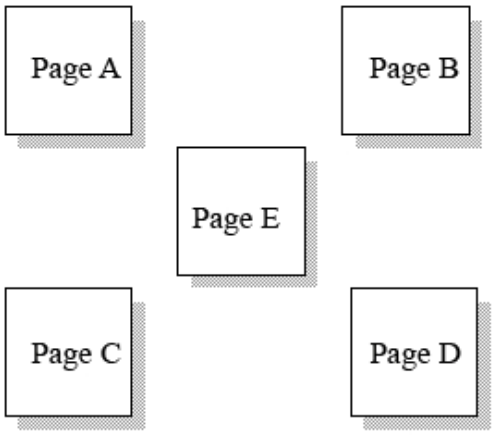

Figure 2. The pages inside a site with no links among them.

As you know, the rank value of 1 is considered for each of the pages in the first round of calculations. It was mentioned that this preliminary value makes no effect on computing such model. Regardless of the accuracy of calculations, it could be stated

that the value of page rank after several repetition of calculations comes to a fixed amount. Starting with the value of 1 will result in fewer times of calculations.

The highest page rank of a site is a total of each page's rank. Therefore, if there are five pages in a site each of which embodies the rank of 1, the page rank of the site will equal to 5. In case there is no link among the pages, the final rank of the site will be extremely fewer than the maximum amount of rank (Figure 2).

After one round of calculations, the rank amount of each page will be 0.15. Repeating calculations for more times, the amount of 0.15 remains for each page but the total page rank of the site will be 0.75 while it could achieve the number of 5. This example shows how the page rank of the site could stay lower than the maximum.

In the next mode, for example, we will make a link between two links in which page A makes a link to page B and conversely B makes a link to A but the page C stays still without any inbound or outbound links (Figure 3).

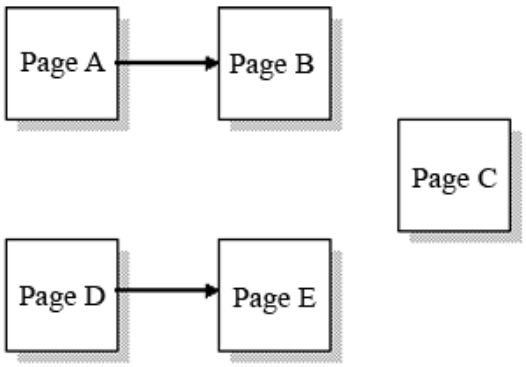

Figure 3. The impact of links on changing the page rank.

Assuming the first page rank of 1 for each of the pages and after one time of calculation, the following numbers are resulted:

Page A = 0.15

Page B = 1

Page C = 0.15

Page D = 0.15

Page E = 1

As it is seen, the page rank of B has increased. Repetition of computations –about 100 times of calculations– will change the results as follows:

Page A = 0.15

Page B = 0.2775

Page C = 0.15

Page D = 0.15

Page E = 0.2775

At this stage, the page rank of B and E pages has improved and the total page rank of the site gets number 1.005. This rank value of the site is better than the previous stage but there is still a gap of 3.995 to reach the maximum rank. It should be considered that these facts are technically inaccurate because the specified factors such as dangling links –which are concerned in google– have been ignored here for the ease of calculations.

Now, we suppose another mode in which there are outbound and inbound links to each of the pages (Figure 4).

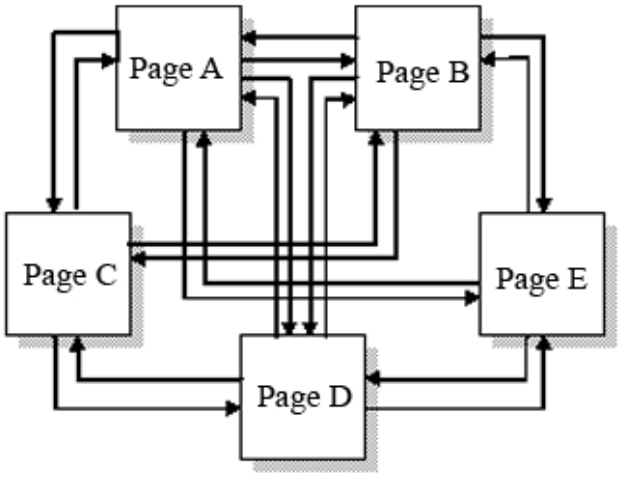

Figure 4. The amount of page rank of the site by maximum number of links amongst the pages.

Supposing the first rank of 1, calculations of this mode show the following results:

Page A = 1

Page B = 1

Page C = 1

Page D = 1

Page E = 1

This mode achieves the maximum rank. Even if the linkage was in the circular method, that is A to B, B to C, C to D, D to E and E to A, such a result could be achieved. It is seen that weak linkage will simply decrease the page rank and conversely, strong linkage could result in gaining the complete potential of the site. It should be reminded that we are not to let the pages of the site have equal share of ranking. It means that some of the pages inside the site must embody more shares in scaling. For example the index pages and hub pages are matters of great importance. By the sample site already investigated –a site with 5 pages– we are to steer the page rank to the index page of A.

Suppose a mode in which page A has made links to B, C, D and E pages and has received links from them (Figure 5).

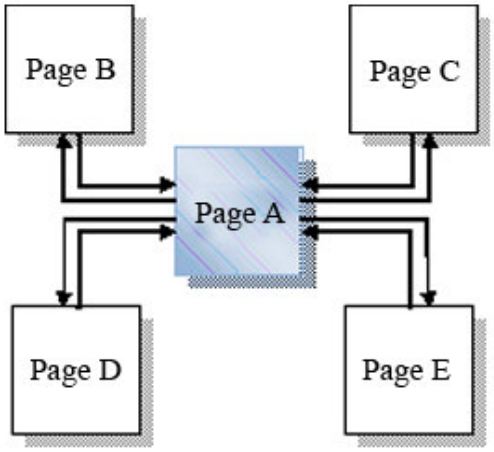

Figure 5. Steering the page rank to a special page.

At this mode, one time calculation with the firs page rank of 1 produces the following results:

Page A = 3.55

Page B = 0.3625

Page C = 0.3625

Page D = 0.3625

Page E = 0.3625

And after 100 times of computation we have:

Page A = 2.3783

Page B = 0.6554

Page C = 0.6554

Page D = 0.6554

Page E = 0.6554

As it is seen in the two previous examples, the total page rank of the site reaches the maximum number of 5. Figure 5 shows that page A possesses more shares of the whole page rank of the site than other pages. This is how to steer the page rank to one special page of the site.

In the next mode, we add a new link from C to B (Figure 6).

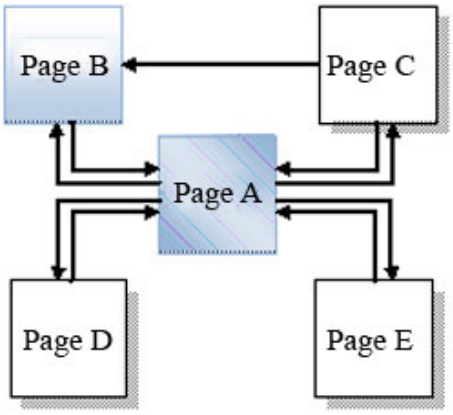

Figure 6. Sharing the page rank among the pages receiving link.

Having the preliminary rank of 1 for all pages, one round of calculation gives following results:

Page A = 3.125

Page B = 0.7875

Page C = 0.3625

Page D = 0.3625

Page E = 0.3625

Comparing the results of one round of calculations with the results of the previous mode, it is seen that page A has lost some of its page rank, page B has increased its page rank and page C has not faced any change. This is so because page A, in the previous mode, had received all the votes of other pages but here the vote of page C in making links has been shared between pages A and B.

After about 100 times of repeating calculations there are:

Page A = 2.23482

Page B = 0.89048

Page C = 0.62489

Page D = 0.62489

Page E = 0.62489

An increase in the page rank of C is seen at this mode because page C has shared its vote between A and B. Compared to the previous mode in which C votes completely for A, in other words, page A here can add less value to the page C in the correlation of A→C. Therefore it could be concluded that making extra links by a page to other pages could indirectly decrease its page rank if it receives link from any of these pages – presence of mutual links. But in absence of converse links –mutual links– we do not

expect decrease in page rank of the extra link-maker pages.

Inbound links

It is seen that receiving links from other pages is a cause of increase in the page rank. Considering the algorithm used in calculation of page rank, if we imagine page X making an inbound link to page A, the page rank of A, according to the following formula, will increase:

$d \times PR(X) / C(X)$

PR(X) represents the page rank of X and C(X) stands for the number of outbound links from X (Efactory, 2003). But, on the other hand, making links by A to other pages will increase the rank of these pages. If each of pages receiving link from A, makes link to A, the page rank of A will increase due to more number of inbound links.

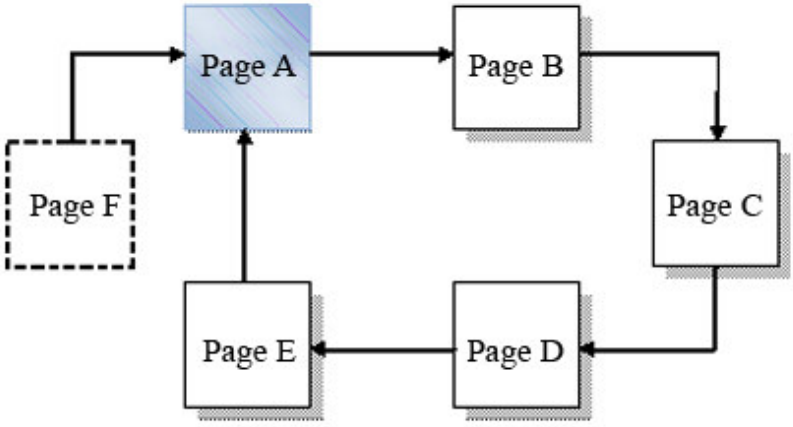

Figure 7. The affect of inbound links on the page rank of the site.

Giving an example will clarify the issue. A, B, C, D and E pages have made a circular link among each other (Figure 7). Regardless of inbound links from pages outside the site, each of the five pages has the page rank of 1. If page F from another site with the same page rank of 1 is added to our sample 4-page site and makes an inbound link to A, the page rank of pages after 100 rounds of calculations will be as following:

Page A = 2.5279
Page B = 2.2987
Page C = 2.1039
Page D = 1.9383
Page E = 1.7976

It is seen that how the impact of the inbound link from an external page, that is F, has been distributed on all pages inside the site. The page rank of A increases for two

reasons: a) due to receiving direct link from page F and b) increase in rank of A will increase the rank of pages receiving inbound link from A and, in turn as the linkage among the pages in our sample is in the circular format, the impact of increase in the rank of pages will again return to page A.

Outbound links

As the page rank is based on the linkage structure available among all pages of the web, the outbound link from a page could cause impacts on the link-maker page itself. According to the formula used in calculation of page rank, the outbound links from a page make no direct influence on the page itself but due to the linkage structure in the web, outbound links could indirectly affect the rank of the link-maker page (Efactory, 2003). As an example, we imagine two websites one of which including A and B pages and the other including C and D pages. The pages A and B have made mutual links to each other and conversely C and D as well. In absence of any link between the two websites, the first page rank of the pages equals to 1 and the total page rank of all pages will be 4 (Figure 8).

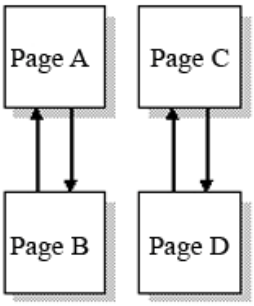

Figure 8. Two website with no links to each other.

In case A makes a link to C, we investigate the impact of this link on page A itself (Figure 9). New computations of page rank gives the following results:

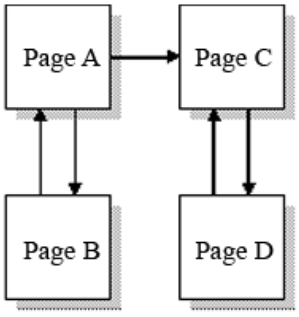

Figure 9. Making outbound link from one site to an other site.

Page A = 0.60869
Page B = 0.47826
Page C = 1.52173
Page D = 1.39130

It is evident that the total page rank of both websites still equals to 4 but the page rank of A and B has decreased whilst that of C and D has increased. A and B pages lose some amount of their ranks, for increase in the number of outbound links from A will affect the rank of page B in a way that the share of B from the rank of page A decreases and consequently the rank of page B drops to some extent down. As the page rank of B decreases, page A which has received link from B, will face a decrease in its rank. Increase in the number of inbound links to C and as a result increase in the rank of page C and the pages receiving link from C are why C and D pages reach more amount of rank. It is clear that outbound links may not cause any impact on the rank of whole pages but these links could change the distribution of ranking within pages.

Adding new pages

One of the important ways of increasing the page rank of the site is to add new pages to the site because adding each new page will bring an average amount of 1 to the general rank of the site. By adding new pages you can steer the page rank to the chief pages of the sites (Rogers, 2005). If you add new pages to the Figure 5 and make links from these new pages to the specified page, that is A (Figure 10) the general amount of the page rank will increase and the page rank of A will be doubled.

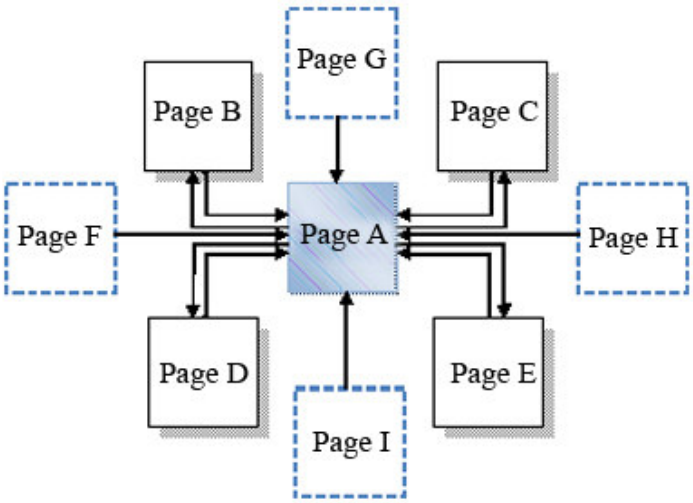

Figure 10. The impact of pages without inbound link on the page rank of other pages.

Page A = 4.21621
Page B = 1.04594
Page C = 1.04594
Page D = 1.04594
Page E = 1.04594
Page F = 0.15
Page G = 0.15
Page H = 0.15
Page I = 0.15

There is a point in the last model which is not correct. New pages are not considered in the indexing system of search engines such as Google and consequently their impact is not considered effective on the ranking of page A. That is so because one page, at least, must be available from which links have been made to new pages.

Decrease in the page rank due to adding new pages

Adding new pages may cause negative effects on page rank of the site. As one page is added, its rank must be extracted from already available pages. If old pages lose some quantity of their ranks for giving some rank to the new pages, then the general page rank of the site will increase. More new pages are added, more old pages will lose the amount of scaling. Therefore, arranging new pages into the site should be carried out in a way that the main pages of the site do not lose their rank (Efactory, 2003). Now, we add 20 new pages to the example illustrated in the Figure 4 each of which has made link only to page A and has received link only from page A. The general page rank of the site reaches the number of 21 which is a conclusion of adding new pages. The inbound links to page A increase the page rank of A, but increase in the number of outbound links from A will decrease the share of pages receiving link from A (Figure 11). After 100 rounds of calculations the following results are given:

Page A = 9.7297
Other pages = 0.5635

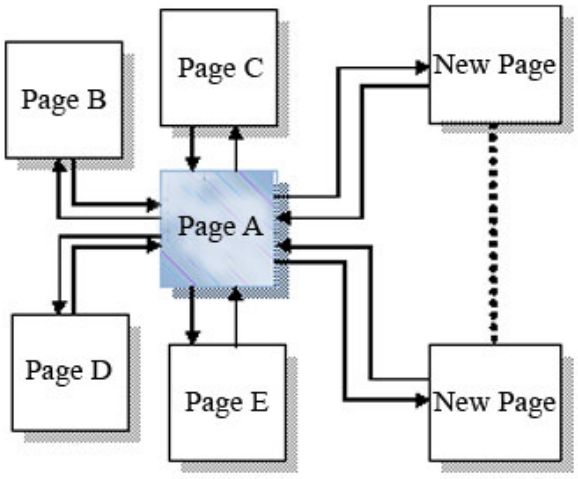

Figure 11. The impact of adding new pages on the page rank of each page and on the general page rank of the site.

Dangling links

Those links entering the page without any kind of link to other pages are known as dangling links. In other words, dangling link enters the page which has no connection to other links. These links could be effective for the page rank because it is not clear how the impact of such links is distributed and moreover, there are a huge number of such links within the www. Such links will be usually ignored in calculating the page rank of pages. It means that search engines such as Google do not consider dangling links in calculation of page rank (TwoSpots, 2008). For example, in Figure 12, page B has received a link from page A but B has no other kinds of link to other pages. Therefore, the link from A to B is a kind of dangling one and it will not be considered in calculations. The page rank of each of the pages in presence of a dangling link is as follows:

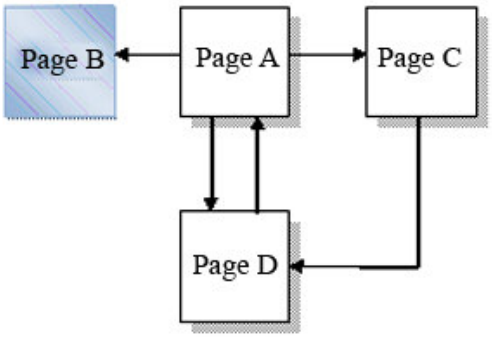

Figure 12. Dangling links and their impact on the page rank of the site.

Page A = 1.1633
Page C = 0.6444
Page D = 1.1921

BadRank against page rank

There will be introduced a technique such as page ranking which is used in analysis of the structure of pages' links, while contrary to page ranking, this technique does not identify the importance of a page but measures the negative attributes of the pages. This technique is called BadRanking. If page A embodies any kinds of link to page B with an amount of high BadRank, page A will also find some amount of BadRank. The difference between these two techniques is that BadRank is not computed based on investigating the inbound links to pages but is calculated according to the outbound links from pages. The following formula shows how to calculate the BadRank:

BR(A) = E(A) (1-d) + d (BR(Ti)/C(Ti) + … + BR(Tn)/C(Tn))

In this formula, BR(A) is the BadRank related to page A, BR(Ti) is the BadRank related to Ti pages consisting outbound links from page A, C(Ti) is the number of inbound links of Ti links and d is the moderator factor (Efactory, 2003).

Remarkable Notes in Calculating Page Rank

Other than inter-site structure, factors such as choosing the Domains URLs, different levels of pages inside the site and the content of pages could cause positive and negative affects on page rank.

Domains and files' URLs

In view of search engines, the URLs presented in Table 4 are not the same and consequently they are investigated differently.

Table 4

Different Modes of Displaying URLs

www. domain.com/
domain.com/
www.domain.com/index.html
domain.com/index.html

The difference in using URLs will cause troubles in calculating site rank. It would be better to standardize the addresses related to the home page and other pages; otherwise, each address could bring different page rank.

Subdirectories

The principles used in subdirectories represent the importance of arrangement as well as accuracy of chief pages of the site. In other words, the size of web avoids calculating and validating all levels available at a site. Therefore, the first or high level pages must be well taken into consideration so that the rank of the site could achieve a deserving status. Someone –as Google does not follow a distinct system due to its commercial aspects– believes in that Google subtracts one unit from the page rank in lieu of each level of subdirectories arranged in relation to the first page. For example, if the value of root pages equals 4, then the value of about 3 will be specified to the pages arranged at the next level. Anyway, as search engines tend not to search through the very low levels, it would be better to keep the structure of subdirectories only at one or two levels. It should also be stated that the principle of decrease in the rank of subdirectories is common in sites in which ranking is based on the main page. If most of the links of other sites are made to subdirectories, the subdirectory page could reach a higher rank than the main page (Smashing Magazine, 2007; Grossman and Frider 2005).

Weighting links based on content analysis

Some of the algorithms used by search engines in the analysis of pages accept the relationship among the pages only if they keep content relationship with each other. Therefore, the content analysis is done to control the accuracy of inbound links and validate the pages ignoring irrelevant links. Stata, Bharat and Maghoul (2000) offered a method based on Topic Vectors. As a matter of fact, topic vectors are "term vectors" in a more extended usage. In other words, topic vectors not only cover the words within a page but they concern the words of several pages of the same subject fields (Efactory, 2003). Therefore, search engines act here intelligently in a way that only if link-maker and link-receiver are connected by relevant content, they could be considered for validation. As a result, no scores will be given to irrelevant links (Stata, et al. 2000).

Google bombs

Google bomb which is also known as "link bomb" is an internet term which relates to a method of affecting the rank of a specific page within the ranking issued by Google (Wikipedia, 2008). Based on the structure of the algorithm calculating page rank in Google, if a large number of sites make inbound links to a specific page, this page gains a high rank and it will become one of the most relevant retrieved sites. As a matter of fact, Google bomb is a collection of remarkable sites making links to a specific page and drive this page to the peak level of ranking (Wired Magazine, 2008).

Google Toolbar

Most of search engines add to their browsers some search facilities in the frame of toolbars one of which is the Google Toolbar shown in Figure 13.

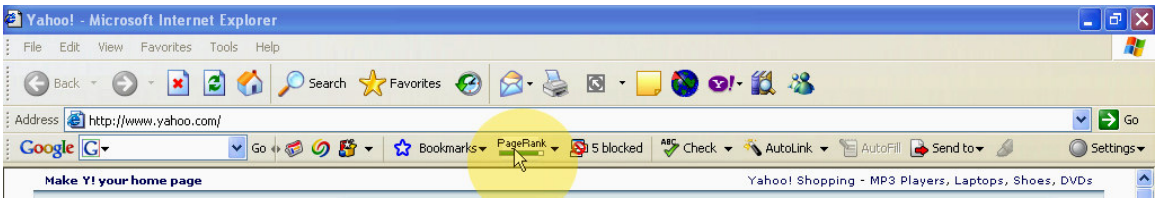

Figure 13. Searching page rank within the Google Toolbar.

One of the abilities prepared by this toolbar is to show the rank of a page which is visible in browser. It should be noted that the ranking shown by the Google toolbar is not exactly equal to the real rank of the page. According to the calculations and views of Google experts, millions of pages through the web are with the average page rank of 1. The Google toolbar ranking is from 1 to 10 on the basis of which Google ranking system divides the real page rank of the whole web into 10 segments. Each segment is shown by the number available in the toolbar. So, the amounts shown in the toolbar display the limit in which the page ranks are set while they lack exact accuracy.

It is stated that changing the page rank in the limit of ranking available at and the toolbar will be done by calculations based on logarithm models and not based on linear models. Giving the logarithm calculations and considering the base 10 for this, linkage of 10 new pages is needed for one unit increase in the links of the toolbar. In order to gain one more unit increase, linkage of 100 new pages is needed, gaining one more unit, linkage of 1000 new pages and reaching another unit of increase need linkage of 10000 new pages to our site. This is why promotion at low levels –for example from page rank of 2 to page rank 3– is much easier than promotion at higher levels –for example from rank 8 to 9.

Conclusion

Page rank is a numeric amount which represents the importance of a page within the web. The toolbars of different search engines such as Google give these numeric amounts while accessing to the pages. Search engines display the result of retrieved pages respectively based on the priority of ranking. Inter-pages links are the most effective factor on page rank. The more the number of inbound links into a page, the more the page rank is. The total page rank of a site depends on the number of pages inside the site. Making optimum links among pages and avoiding creation of pages

lacking accessibility could cause the maximum internal potential of the site. The inbound links into the pages of the site could increase the page rank and conversely, the outbound links to the pages punished by search engines will result in losing page rank and incurring special penalties for the whole site. At the end, methods such as content analysis of pages could be useful in investigating the accuracy of inbound links to the page as well as assessing the accuracy of validating pages.

References

Brin, S., & Page, L. (1998). *The Anatomy of a Large Scale Hypertextual Web Search Engine*. Stanford University. Retrieved November, 2008, from http://infolab. stanford.edu/~backrub/google.html

Dwork, C., Kumar, R., Naor, M., & Sivakumar, D. (2001). Rank aggregation methods for the web. *Proceedings of the Tenth International World Wide Web Conference*. Hong Kong.

Efactory. (2003). *A survey of Google's pagerank*. Retrieved September, 2008, from http://pr.efactory.de/e-pagerank-themes.shtml

Grossman, D., & Frider, E. (2005). *Information retrieval: Algorithms and exploration methods*. (Jafar Mehrad & Sara Koleini, Trans.). Mashhad: Digital Library publications & Shiraz: RLST.

Hirai, J., Raghavan, S., Garcia-Molina, H., & Paepcke, A. (1999). *Web Base: A repository of web pages*. Stanford University Technical Report.

Jeh, G., & Widom, J. (2002). *Scaling personalized web search*. Stanford University Technical Report. Retrieved October, 2008 from http://ilpubs.stanford.edu: 8090/530/

Horrell, M. *Page Rank Calculator*. Retrieved November, 2008, from http://www.markhorrell.com/seo/pagerank.asp

Mehrad, J., & Binesh, M. (2007). Investigation of the library and information science websites retrieved by Yahoo and ODP based on content and efficiency. *Faslnameye Ketabdari va Ettela-Resani*. *10*(1), 39-60.

Rogers, I. (2005). *The Google pagerank algorithm and how it works*. Retrieved September, 2008, from http://www.iprcom.com/papers/pagerank/

Smashing Magazine. (2007). *Google pagerank: What do we know about it?* Retrieved October, 2008, from http://www.smashingmagazine.com/2007/06/05/google-pagerank-what-do-we-really-know-about-it/

Stata, R., Bharat, K., & Maghoul, F. (2000). The Term Vector Database: Fast access to indexing terms for Web pages. *9th International World Wide Web Conference (WWW.)*. May, 15-19, 2000, Amsterdam, The Netherlands.

TwoSpots, *Google's PageRank explained and how to make the most of it*. Retrieved

September, 2008, from http://www.twospots.com/web-articles/23/

WebWorkshop. *Google's PageRank – Calculator*. Retrieved September, 2008, from http://www.webworkshop.net/pagerank_calculator.php

Wikipedia. (2008) *Google bomb*. Retrieved October, 2008, from http://en.wikipedia.org/wiki/Google_bomb

Wired Magazine. *Remembering the first Google bomb*. Retrieved September, 2008, from http://blog.wired.com/monkeybites/2007/01/earlier_today_m.html