ABSTRACT

Google index in 1998 had 26 million pages. The underlying technology behind optimized search among 26 million pages was the page rank algorithm. Page Rank stated that a document ranks high if other HIGH ranking documents link to it. Thus, rank of document is determined by the rank of document which link to it. Their rank again is given by the rank of documents which link to them. Hence, the Page Rank of a document is always determined recursively by the Page Rank of other documents. And thus, Page Rank is, in the end, based on the linking structure of the whole web. This paper describes the page rank algorithm and the major factors affecting this algorithm namely: inbound links, outbound links and the number of pages in a website.

Keywords: Inbound links, Outbound Link, Random Surfer, Page Rank.

1. INTRODUCTION

Google gained its popularity due to its unique Page Rank algorithm. Originally this algorithm was given by the Stanford graduates Lawrence Page and Sergey Brin. This algorithm has undergone quite a number of modifications to fit in the requirements as and when required.

Conventional approach employed by various search engines was to search for a phrase within the document. In the approach occurrences of words and phrases was weighted. This weightage was decided on basis of density of the phrase in the document and the emphasis given to the phrase. In order to check the emphasis on the phrase its HTML tags were checked. The problem with the above approach was automatic website were generated (spam websites) based on analysis of content specific ranking criteria (doorway pages). Although the approach of the algorithm sees broad and complex, Page and Brin were a practice by a relatively trivial algorithm. [1]

2. PAGE RANK ALGORITHM

The original Page Rank algorithm was described Page and Sergey Brin in several publications.

\[
PR(A) = (1-d) + d \left( \frac{PR(T1)}{C(T1)} + \ldots + \frac{PR(Ti)}{C(Ti)} \right)
\]

Where, \(PR(A)\) is the PageRank of page \(A\), \(PR(Ti)\) is the PageRank of pages \(Ti\) which link \(C(Ti)\) is the number of outbound links on page damping factor which can be set between 0 and 1. So, first of all, we see that PageRank does not as a whole, but is determined for each page further, the PageRank of a page, suppose recursively defined by the PageRanks of the link to this page \(A\). The Page Rank of pages \(Ti\) which link to \(p\) influence the Page Rank of page \(A\) uniformly. Rank algorithm, the Page Rank of a page weighted by the number of outbound links \(C\). This means that the more outbound links a page will page \(A\) benefit from a link to it on page \(T\).

The weighted Page Rank of pages \(Ti\) is th outcome of this is that an additional inbound will always increase page \(A\)'s Page Rank. Finally, the sum of the weighted Page Rank of multiplied with a damping factor \(d\) which can and 1. Thereby, the extent of Page Rank benefit another page linking to it is reduced. [2]

3. THE RANDOM SURFER MODEL

In their publications, Lawrence Page and Sergey Brin give a very simple intuitive justification for the Page Rank algorithm. They considered Page Rank as a model of user behavior, where a surfer clicks on links at random with no regard towards content [6].

The random surfer visits a web page with a certain probability which derives from the page's Page Rank. The probability that the random surfer clicks on one link is solely given by the number of links on that page. This is why one page's Page Rank is not completely passed on to a page it links to, but is divided by the number of links on the
page. So, the probability for the random surfer reaching one page is the sum of probabilities for the random surfer following links to this page. [2]

Now here we introduce the term ‘d’ or the damping factor. The probability for the random surfer not stopping to click on links is given by the damping factor d, which is, depending on the degree of probability therefore, set between 0 and 1. The higher d is, the more likely will the random surfer keep clicking links. [2]

The surfer jumps to another page at random after he stopped clicking links. Regardless of inbound links, the probability for the random surfer jumping to a page is always (1-d), so a page has always a minimum Page Rank.

4. IMPLEMENTATION

In google Page Rank algorithm alone was not implemented for searching purpose. An IRscore was also attached with it for effective search. Following three factors determine the IRscore of a website.
1. Page specific factors
2. Anchor text of inbound links
3. Page Rank

Page specific factors are, besides the body text, for instance the content of the title tag or the URL of the document. In order to provide search results, Google computes an IR score out of page specific factors and the anchor text of inbound links of a page, which is weighted by position and accentuation of the search term within the document. This way the relevance of a document for a query is determined achieve better rankings than pages with high means of classical search engine optimization.

If pages are optimized for highly competitive s is essential for good rankings to have a high Pa if a page is well optimized in terms of classical optimization. The reason therefore is that the score diminishes the more often the keyword o document or the anchor texts of inbound lin by extensive keyword repetition. Thereby, the classical search engine optimization are limited becomes the decisive factor in highly competition areas.[3]

4.1 Inbound Links

Backlinks, also known as incoming links, inbound links, inlinks, and inward links, are incoming links to a website or web page. In basic link terminology any link received by a web node (web page, directory website, or top level domain) from another web node [4] whereas, outbound links start from your site an external site.

The Effect of inbound link:

It has already been said that each additional in web page always increases that page's Page look at the Page Rank algorithm , which is given by \( PR(A) = (1-d) + d \left( \frac{PR(T1)}{C(T1)} + ... + \frac{PR(Tn)}{C(Tn)} \right) \) One may assume that an additional inbound link increases the Page Rank of page A byd \( \times \) PR(X)/C(X) is the Page Rank of page X and C(X) in total number of its outbound links. But page A other pages itself.[5] Thus, these pages get a Page Rank benefit also. If these pages link back to page A, page A will have Page Rank benefit from its additional inbound link.
Considering the above figure (Fig 3), let a web site consist of 4 pages A, B, C, D circularly linked without any external inbound links. Thus, Page Rank of each page A, B, C, D will be 1.

Now consider page X as an inbound link to page A of the website. Consider page rank of X as 10, denoted by PR(X). Let the damping factor d to be 0.5. Now we get the following equations for each of the page of the website (Referring to eq.1)

\[
\begin{align*}
PR(A) &= 0.5 + 0.5\times(PR(X) + PR(D)) = 5.5 + 0.5\times PR(D) \\
PR(B) &= 0.5 + 0.5\times PR(A) \\
PR(C) &= 0.5 + 0.5\times PR(B) \\
PR(D) &= 0.5 + 0.5\times PR(C)
\end{align*}
\]

Since the total number of outbound links for each page is one, the outbound links do not need to be considered in the equations. Solving the above equations we get:

\[
\begin{align*}
PR(A) &= 19/3 = 6.33 \\
PR(B) &= 11/3 = 3.67 \\
PR(C) &= 7/3 = 2.33 \\
PR(D) &= 5/3 = 1.67
\end{align*}
\]

We see that the initial effect of the additional inbound link of page A, which was given by \(d \times PR(X)/C(X) = 0.5 \times 10/1 = 5\) is passed on by the links on our site.

The higher the damping factor, the larger is the effect of an additional inbound link on the Page Rank of the page that receives the link, and the distribution of Page Rank over the other pages of the site is more even. Suppose in the above example we took 0.75 as damping factor then rank of each page would increase.

At a damping factor of 0.5, the accumulated Page Rank of all pages of our site is given by

\[
PR(A) + PR(B) + PR(C) + PR(D) = 14
\]

At a damping factor of 0.75 the accumulated Page Rank of all pages of the site is given by

\[
PR(A) + PR(B) + PR(C) + PR(D) = 34
\]

As for a website with no outbound links the accumulated page rank increases by a factor of: \((d / (1-d)) \times (PR(X)/C(X))\)

Where X is a page additionally linking to one page of the site, PR(X) is its Page Rank and C(X) its number of outbound links. The formula presented above is only valid, if the additional link points to a page within a closed system of pages, in other words a website without outbound links to other sites.

For the actual Page Rank calculations at Google, Lawrence Page and Sergey Brin claim to usually set the damping factor ‘d’ to 0.85. Thereby, the boost for a closed system of web pages by an additional link from page X is given by

\[
(0.85/0.15) \times (PR(X)/C(X)) = 5.67 \times (PR(X)/C(X))
\]

So, inbound links have a far larger effect than one may assume. It is not necessary for a page to have many inbound links to rank well. A single link from a high ranking page is sufficient.

### 4.2 Outbound Links

Since Page Rank is based on the linking structure of the whole web, it is inescapable that if the inbound links of a page influence its Page Rank, its outbound links also have some impact.

Both pages of each site solely link to each both page has page rank one.

Now we link page A of the first website to page assuming the damping factor to be 0.75. We therefore get the following equations for the Page Rank values:

\[
\begin{align*}
PR(A) &= 0.25 + 0.75\times PR(B) \\
PR(B) &= 0.25 + 0.375\times PR(A) \\
PR(C) &= 0.25 + 0.75\times PR(D) + 0.375\times PR(A) \\
PR(D) &= 0.25 + 0.75\times PR(C)
\end{align*}
\]

Solving the equations gives us the following P for the first site:

\[
PR(A) = 14/23
\]
We therefore get an accumulated Page Rank first site. The Page Rank values of the second site:

PR(C) = 35/23
PR(D) = 32/23

So, the accumulated Page Rank of the second site's total Page Rank for both sites is 92/23 = 4. Each link has no effect on the total Page Rank of the additional site, the Page Rank benefit for one Page Rank loss of the other. As it has already been shown, the Page Rank closed system of web pages by an additional in given by:

(\frac{d}{1-d}) \times (\frac{PR(X)}{C(X)})

back to that system, since it otherwise gains by lost Page Rank.

The intuitive justification for the loss of p addition of outbound links is that if a person is external page, from say any page 'A', then the person to remain in page 'A' diminishes. Thus, the Page Rank of website containing pages.

4.3 Effect Due to number of pages on page rank

Since the accumulated page rank of a website of the individual page rank of the pages, one w normally draw a conclusion that addition of a increase the overall page rank of the website. But interestingly, this need not necessarily be b

Consider the following example.

Solving the equations gives us the following Page Rank values:

PR(A) = 260/14
PR(B) = 101/14
PR(C) = 101/14

Now we add a new page hierarchically on the lower level of the site. After adding page D, the equations for the pages' Page Rank values are given by:

PR(A) = 0.25 + 0.75 (10 + PR(B) + PR(C) + PR(D))
PR(B) = PR(C) = PR(D) = 0.25 + 0.75 (PR(A) / 3)

Solving these equations gives us the following Page Rank values:

PR(A) = 266/14
PR(B) = 70/14
PR(C) = 70/14
PR(D) = 70/14

As expected since our example site has no outbound links, after adding page D, the accumulated Page Rank of all pages increases by one from 33 to 34. Further, the Page Rank of page A rises marginally. In contrast, the Page Rank of pages B and C depletes substantially.

By adding pages to a hierarchically structured websites, the consequences for the already existing pages are no uniform. The consequences for websites with a different structure shall be shown by another example.
Referring to the above figure (Fig 6), in this example if we add page D without disturbing the structure of the website the accumulated page rank of website increases as in previous example but the page rank of ALL individual pages decreases including page A (note in previous example page rank of A had increased).

As we have seen that addition of pages to a website individual page rank of the pages in the website generally, this algorithms favors websites with lesser number of pages, however can counter this effect by adding attractive content and number of inbound links.

5. CONCLUSION

Thus, page rank algorithm provided an innovative unconventional way of link analysis to optimize results. However, this algorithm was criticized for not making any significant use of the quality of backlinks. The subsiding of page rank caused the declining of page rank and newer algorithms like Panda and Penguin algo taken up by Google which ranked websites as individual pages further they also took the quality of content into consideration.

REFERENCES

[6] The Page Rank Citation Ranking: Bringing it Web (PDF, 1999) by Lawrence Page, Rajeev Motwani and Terry Winograd.

AUTHORS

Mr. Romit D. Jadhav
Received Bachelor’s Degree in Computer Science And Engineering from Sant Gadge Baba Amravati University in 2012 & Pursuing Master Degree In CSE from P.R. Patil College of Engg. Amravati-444602.

Prof. Ajay B. Gadicha
Received the Master Degree in Information Technology from Gadge Baba Amravati University in 2011. Working as Assistant professor in Department of Information Technology at P. R. Patil College of Engg. Amravati-444602.