International Journal of Computer Science
and Engineering (IJCSE)
ISSN(P): 2278-9960; ISSN(E): 2278-9979
Vol. 3, Issue 4, July 2014, 37-42

International Academy of Science,
Engineering and Technology
Connecting Researchers; Nurturing Innovations

IMPROVE PAGE RANK ALGORITHM USING NORMALIZED TECHNIQUE

NIDHI K. CHITRODA[1] & KAMLESH M. PATEL[2]

[1]M.Tech Student, Department of Computer Engineering, School of Engineering, R.K University, Kasturbadham, Tramba, Rajkot, Gujarat, India

[2]Associate Professor, School of Engineering, R.K University, Kasturbadham, Tramba, Rajkot, Gujarat, India

ABSTRACT

Search engine is one kind of software, which enlists data about web sites. All the major search engines such as Google, Yahoo, Ask, Bing, etc. compute rank of web-pages by using specific algorithm. Page Rank is one of many factors that determines where your web page appears in search result ranking, but if all other factors are equal, Page Rank can have significant impact on your Google rankings. PageRank is measured on a scale of one to ten and assigned to individual pages within a website, not the entire website. Page rank can be compute by iterative, algebraic or power method. In iterative technique of page rank computation, it takes large number of iterations and memory space which can be reduce. This paper discusses iterative technique for computing PageRank, with much reduced number of iterations. And proposed algorithm gives improvement in page rank according to number of back links.

KEYWORDS: SERP = Search Engine Results Page

INTRODUCTION

Nowadays searching on the internet is most widely used operation on the World Wide Web. [2] The amount of information is increasing day by day rapidly that creates the challenge for information retrieval. There are so many tools to perform efficient searching. Due to the size of web and requirements of users creates the challenge for search engine page ranking. [6] Ranking is the main part of any information retrieval system Today's search engines may return millions of pages for a certain query It is not possible for a user to preview all the returned results So, page ranking is helpful in web searching. [3]

PageRank has been developed by Google and is named after Larry Page, Google's co-founder and president [7]. PageRank is designed to simulate the behavior of a "random web surfer" [1] who navigate a web by randomly following links. If a page with none outgoing links is achieved, the surfer goes to a randomly chosen bookmark. In more to this normal surfing behavior, the surfer occasionally spontaneously goes to a bookmark instead of following a link.

Importance of Back Links

Back Linking can be an effective method for gaining Page Rank. [8] Back Links can be described as links that are directed towards a website. The amount of Back Links is a signal of the popularity/importance of that website. Search Engines such as Google, offers extra credit to websites that have a high number of quality Back Links and consider those websites more relevant than others in the Page Ranks.

Most of search spiders include the inbound links aspect in search engine algorithm. If your web site has a more backlinks then your site is more trustworthy in SERPs eye.

"The more back links website will have, the more frequently it will be visited by search engines robots, which means that new content will be indexed much faster". [8]

In theory, if one page on a website has a lot of extremely relevant Back Links, it will be ranked high in the search engines, whereas the remainder of the pages will be ranked low. "In reality, the whole website will be ranked high, because very often, each page will have a plenty of internal back links to the main page and vice versa". [8]

It is not enough to have a lot of number of in links, you need to have quality links. The content of the sites you have incoming links with must to be related to the content on your website.

Note that valuable backlinks and relevant incoming links are two very essential scrutinizes. Inward links from a website that has good content and content linked to the theme of your website will give you much better stead than a link from a badly designed web site.

Getting Quality Back Links

If you exchange backlinks with a good and truthful site then it is OK but if you link with bad site then it would be dangerous to your site. The good practice of backlinks exchange will be making stand. There are following types of incoming links to your site:

- One way inbound links: This type of in links gives you more value and better results. One-way incoming links are links that peak to your site from other websites.
- Reciprocal inward links: It is not popular now days. Search spiders do not like these links like they do. I would be better to get one-way backlinks which give you better results and functionality.

One Way Back Links

Making one way inbound links to your site, guarantees you better search positions, which gives you better and valuable visitors mainly from search robots to your site. But, how to force other webmasters to link to your site.

One of the possible ways is to keep your site filled with good and unique contents. When some web master comes to your site, and get something useful, they would likely link to your site on their pages.

Page Rank

Page Rank is a "vote", by all the other pages of site on the Web, about how important a page is. [1] Any link to a web page determines as a vote of support. [1] If there's no link there's no support (but it's only an abstention from voting rather than a vote against the page).

HOW TO MEASURE PAGE RANK

Suppose for instance, that we have a small Internet consisting of just 4 web sites www.page1.com, www.page2.com, www.page3.com, www.page4.com, referencing each other.

We translate this scenario into a directed graph with 4 nodes, one for each web site. When web site i references j, we include a directed edge between node i and node j in graph. For instance, Page1 links to all of the other pages, so node 1 in the graph will have outgoing edges to all of the other nodes. Page3 has only one link, to Page 1, therefore node 3 will have one outgoing edge to node 1.

In our model, each page should transfer evenly its importance to the pages that it links to. Node 1 has 3 outgoing edges, so it will pass on of its importance to each of the other 3 nodes. Node 3 has only one outgoing edge, so it will pass on all of its importance to node 1. In general, if a node has k outgoing edges, it will pass on of its importance to each of the nodes that it links to. Better visualize the process by assigning weights to each edge is shown in figure.

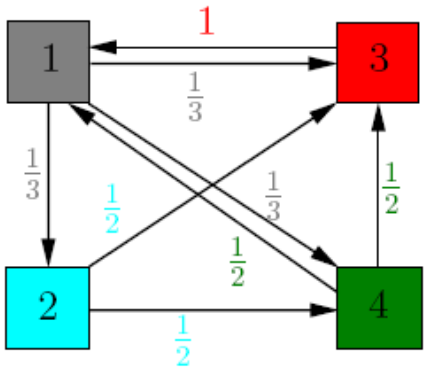

Figure 1: Importance Distribution

So, denote by A the transition matrix of the graph, A = $\begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}$.

So M = $[a_{ij}]$

Where i,j = 1 to number of webpages

a_{ij} = vote to i^{th} page by j^{th} page

After making such a matrix one can follow any algorithm to compute page rank of web page.

PAGERANK ALGORITHM

Page ranking algorithms are the heart of search engine and give result that suites best in user expectation. Need of best quality results are the main reason in innovation of different page ranking algorithms, HITS, PageRank, Weighted PageRank, Distance Rank, Dirichlet Rank Algorithm, Page content ranking are different examples of page ranking used in different scenario.

Traditional Page Rank Algorithm

- Calculate page ranks of all pages by following formula:

  PR (A) = (1-d) + d (PR(T_1)/C(T_1) +..... + PR(Tn)/C(Tn))

  Where PR (A) is the PageRank of page A,

  C (Ti) is the number of outbound links on page Ti

  PR (Ti) is the PageRank of pages Ti which link to page A and

  d is a damping factor which can be set between 0 and 1, but it is usually set to 0.85

- Repeat step 1 until values of two consecutive iterations match.

Proposed Page Rank Algorithm

- Initially assume PAGE RANK of all web pages to be any value, let it be 1.
- Calculate intermediate matrix by following formula

  Mi = (d * M) + (((1 - d) / N)

  Where M is input matrix according to web graph

  N is number of web pages in particular web page.

  d is damping factor which is set to 0.85
- Calculate page ranks of all pages by following formula

  PR(A) = Mi * PR(T_1) …. PR(Tn)

  Where T_1 through Tn are pages providing incoming links to Page A

  PR(T_1) is the Page Rank of T_1

  PR(Tn) is the Page Rank of Tn
- Repeat step 3 until page rank values of two Consecutive iterations are same. The pages which have the highest page rank are more significant pages. So, number of iterations for calculating the page ranks in the Proposed Page Rank algorithm are reduced.

Experiments

The page rank value for the both of algorithm are almost same. We are using different equation in both of case but this computation difference is in <1. That's why the final resultant page rank does not vary large.

In traditional algorithm there is summation of two numbers which are less than one where in proposed algorithm there is multiplication of two numbers which are less than one. So multiplication of two numbers which are less than one get tends to zero faster as compare to summation of those same numbers.

In Matlab R2009b for specific number of nodes/pages, generated random web graph or link matrix. According to it generated vote matrix and input matrix for page rank algorithm. For 10 webpages random generated web graph is shown in figure 2.

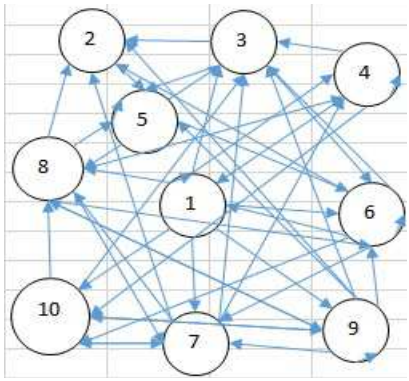

Figure 2: Web Graph for 10 Web Pages

For this web graph page rank values by both the algorithm are shown in table 1.

Table 1: Page Rank Values

Webpage	Traditional	New	Difference
1	0.0729	0.0718	0.0011
2	0.0705	0.0695	0.001
3	0.163	0.1602	0.0028
4	0.0809	0.0797	0.0012
5	0.1453	0.1429	0.0024
6	0.0554	0.0547	0.0007
7	0.1083	0.1066	0.0017
8	0.0747	0.0736	0.0011
9	0.1225	0.1205	0.002
10	0.1225	0.1205	0.002
Number of Iterations	39	5	34

By practical performing both of the algorithm we get results as shown in graph 3.2

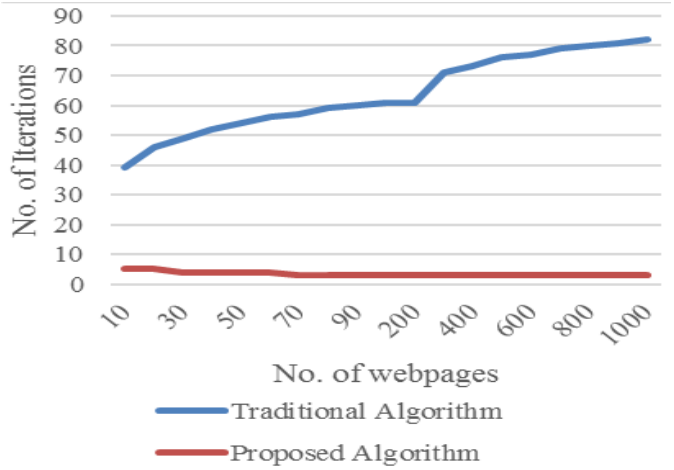

Figure 3: Traditional vs. Proposed Algorithm

We can see that there is vast difference in number of iterations for particular number of webpages. We have generated web graph randomly for defined number of webpages. So number of iteration gets reduce in proposed case and after one value of webpages it gets constant.

CONCLUSIONS

SEO basically relies on user behavior, social engagements, visitors and other publishers. Page Rank is, in fact very simple apart from one scary seeming formula. But when a simple calculation is applied hundreds (or billions) of times over the results can seem complicated. The average Actual PR of all pages in the index is 1.0.

Getting inbound links to site is the only way to increase average PR of site. How that PR is distributed amongst the pages on our site depends on the details of our internal linking and which of our pages are linked to. If we give outbound links to other sites then our site's average PR will decrease. [5] This principle is being followed by proposed algorithm if we analyze the difference in page rank value. Also proposed algorithm gives less number of iteration by changing in computation method. And number of iteration gets much reduced in proposed case and after one value of webpages it gets constant. So one can save the time and memory for large computation (hundreds/thousands of nodes).

REFERENCES

1. Kaushal K., Abhaya, Fungayi D., 2013: 'PageRank algorithm and its variations: A Survey report', IOSR Journal of Computer Engineering (IOSR-JCE), 14.
2. Ali H. Al-Badi, Ali O. Al Majeeni, Pam J. Mayhew and Abdullah S. Al-Rashdi, 2011: 'Improving Website Ranking through Search Engine Optimization', Journal of Internet and e-business, 2011, Article ID 969476.
3. 'Google Page Rank Algorithm', http://www.sirgroane.net/google-page-rank/ (Accessed on 18-10-2013)
4. Hema D., Prof. B. N. Roy, 2011: 'An Improved Page Rank Algorithm based on Optimized Normalization Technique', (IJCSIT) International Journal of Computer Science and Information Technologies, 2(5).
5. Sharma D, and A. K. Sharma, 2010: 'A Comparative Analysis of Web Page Ranking Algorithms', International Journal on Computer Science and Engineering, 1, 2670-2676.
6. Dr. Khanna S., Om Prakash V., 2011: 'CONCEPT OF SEARCH ENGINE OPTIMIZATION IN WEB SEARCH ENGINE', IJAERS, 1,235-237.
7. Michael David, 2013: 'Search Engine Optimization in 2013'.
8. SEO Tips, 2011, "Backlinks", accessed January 2013, http://www.seotipsy.com/backlinks/ (Accessed on 12-12-2013)
9. http://en.wikipedia.org/wiki/PageRank (Accessed on 23-3-2014)