Research of Page ranking algorithm on Search engine using Damping factor

Punit Patel[1]

[1]*Assist. Prof., Computer Department, VGI, Mandvi-Kutch, patelpunitr@gmail.com*

Abstract: The web consists of a huge number of documents that have been published without any quality control. To retrieve necessary information from World Wide Web, search engines carry out number of tasks based on their respective structural design. Various Search engine follow, different algorithm for ranking pages & produce different result. Search engine generally returns a large number of web pages in response to user queries using algorithms. This paper discusses the idea of page rank algorithm based on the number of visits by the client or user. In this paper, a detailed study is being performed on the Page Rank, the Google's algorithm and check the damping factor value 0.85

Keywords: Search Engine, PageRank, World Wide Web, SEO

I.   INTRODUCTION

The World Wide Web (Web) is most well-liked and interactive source to broadcast information today. As on today WWW is the largest information repository and set of all nodes which are interconnected by hypertext links. With the quick growth of the Web, users get easily vanished in the rich hyperlink structure. The main aim of website owners is to providing accurate data based on the user's requirement. so, discover the content of the Web pages and retrieving the users' interests from their actions have become gradually more important.   Search Engine Optimization was a term used in the late 90s to show up the importance of a web page's position in results of the search engine. Search engine optimization (SEO) is a well defined process which is used to improve the website rank and also helps to increase traffic to a web site using search engines. SEO process also helps to increase the number of users to a Web site by high ranking in the search results of a search engine [1]. Higher page rank of websites that means that website is more visited by users. Careful optimization of web sites by Search Engine Optimization that increase the websites visibility in the different search engine Google, Yahoo, Bing and many others. The results obtained by a search engines are a combination of large amount of appropriate and inappropriate information. Normally users visit only that website which is top of the lists. SEO is one type technique which helps to find out and get page rank of website from large number of other sites in response to user's search query. So various ranking algorithm such as Page Rank, HITS are available that helps the users to navigate in the results. These ranking method uses by search engine that sort and displayed the result to users. So users can easily find the best result.

In this paper, Page rank Algorithm which works on the number of inbound links and outbound links of web pages. The main goal of the algorithm is to find out relevant information according to users requirement/query. So, this idea is very valuable to exhibit most precious pages on the top of the result list on the basis of user browsing behavior.

II.   DATA MINING OVER WEB

Now a day, the web revolution has had a profound impact on the way we search and find information at home and at work. The web has also become an enormously important tool for communicating idea, conducting business and entertainment. Web mining is a data mining technique used to extract information from World Wide Web [2]. Millions of web pages are published every day and millions of are modified or removed. Web pages are written in a different language and provide information in variety of sources such as text, video, audio, image, and animation etc.
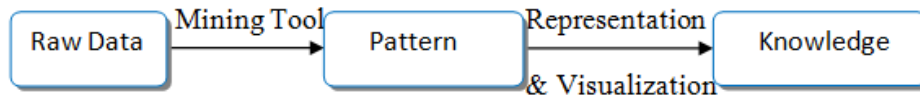
Figure 1 Process of Web Mining

A. Web Mining Categories

Web mining can be classified into three categories Web Structure Mining, Web Content Mining and Web Usage Mining as depicted in literature [3, 4].
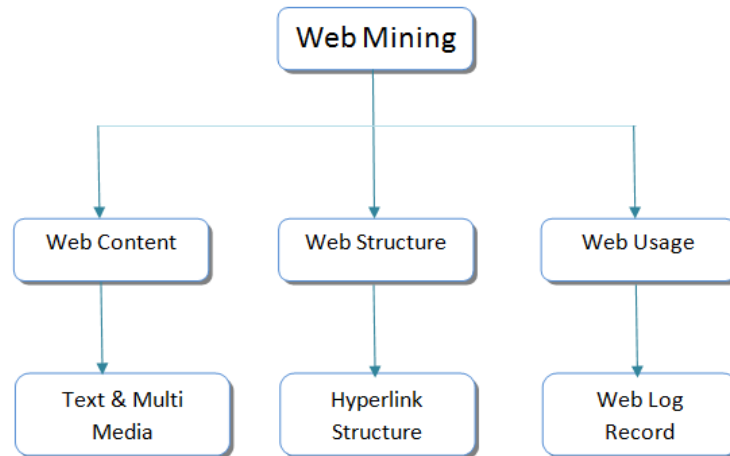
Figure 2 Classified Web Mining

Web Content Mining : Web Content Mining is the process of discovering useful information from the contents of Web pages. Web content mining involves text, images, audio, video information. It is related to text mining because most of website contents are text. It is also related to image mining. Process of Image mining is quite difficult compare to text.

Web Structure Mining: It deals with searching and modeling the web's link structure. Web structure mining consists of nodes (Web pages), as well as edges (hyperlinks) linking between two related pages it is the process of discovering structure information from the Web. Web structure also consist of In-degree and Out-degree Hyperlinks: A Hyperlink is used to connect a different Web page to other web page of different location. Different technique of web structure is Page Rank, HITS and so on.

**Web Usage Mining:** Web Usage mining has been used for various purposes:

➢ A knowledge discovery process for mining marketing intelligence information from web data.

➢ In order to improve the performance of the website, web usage logs can be used to extract useful web traffic patterns. Web usage mining provides valuable knowledge about user behavior on WWW. One of the major goals of web usage mining is to reveal interesting trends and patterns which can be provide useful information about the user of a system. It includes web server log such as user's IP, referral URL, response status and HTTP request and other.

III.   PAGE RANK

Google has the most well known ranking algorithm called the Page Rank algorithm that has been claimed to supply top ranking pages that are relevant. The Page Rank algorithm was used and enhanced by Lawrence Page and Sergey Brin [5]. Page Rank algorithm describes the popularity of web page or website. This Page Rank algorithm is depend on the link Analysis in which ranking of web page is decided based on outbound links and inbounds links[6]. That means it's totally bed on link of WWW and Google uses this algorithm for searching the web pages based on number of hyperlinks such as Inbound and outbound.

Inbound Links: Inbound links are those links that is comes from other site to your website, it is also known as "backlinks". Google consider only relevant links point to your site but you cannot control which sites point to your site. If your website content is unique and rich then there are much chances those links will be "do follow" otherwise links will be consider as "no follow"

Outbound Links: Outbound links are those links that is pointing to other site from your website and you have more control over these links.
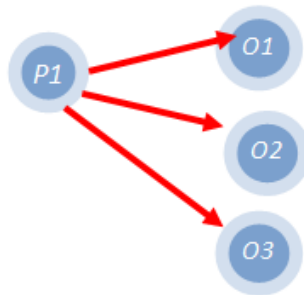
Figure 3. Outbound links pointing to other site

A page has high rank if the other pages with high rank linked to it [7]. It is given by:-

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

- Let A be the page and whose page rank is PR(A).
- Let PR (Ti) is the Pagerank of pages Ti which link to page A,
- C (Ti) is the number of outbound links going out from page Ti and
- d is a damping factor assume to be between 0 and 1 usually 0.85. Sometimes does not click on any links & jumps to another pages at random. It follows the direct links.

- (1-d) is the probability of jumping off to some random pages; every page has a minimum page rank of (1-d). It follows the non-direct links.

To calculate the Page Rank of any Page We required to know the Page Rank of each page that point to it and number of the outbound links from each of those pages.

IV. IMPLEMENTATION AND RESULT ANALYSIS

Let us consider a simple example of three web page A,B and C shown in figure.
  1. Page A contains 1 outbound link that is pointing to Page B.
  2. Page B contains 2 outbound links that is pointing to Page A and Page C.
  3. And Page contains 1 outbound link that is pointing to Page A
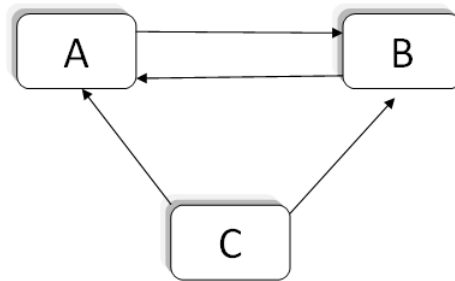 4. The initial page Rank of each page is considered to be 1.

Figure 3. Three web pages links between each other

The Page Rank of each page is computed by following equation
PR (A) = 0.2 + 0.4PR (B) + 0.8PR (C)
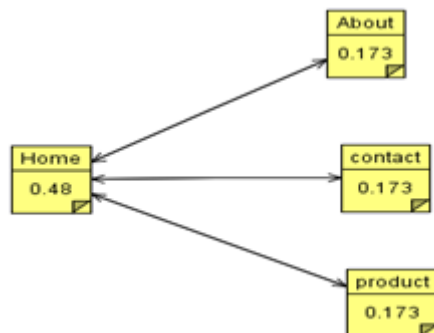PR (B) = 0.2 + 0.8PR (A)
PR (C) = 0.2 + 0.4PR (B)
The result of above equation is given
PR (A) = 1.2
PR (B) = 1.0
PR (C) = 0.66
Evolution for page rank of 4 pages with different damping factor. Here outbound link Is constant but every time damping factor change

Evolution for page rank of 7 pages with different damping factor. Here outbound link Is constant but every time damping factor change.

No	Damping factor	Home	About	Product	Contact	Gallery	Facebook	Twitter
1	0.6	0.295	0.105	0.148	0.127	0.150	0.105	0.070
2	0.7	0.313	0.102	0.146	0.127	0.153	0.102	0.058
3	0.75	0.321	0.100	0.145	0.126	0.155	0.100	0.052
4	0.8	0.329	0.099	0.144	0.126	0.157	0.099	0.046
5	0.85	0.336	0.098	0.143	0.126	0.159	0.098	0.041
6	0.9	0.344	0.097	0.142	0.126	0.160	0.097	0.035

Table 1: Damping factor on famous sites

No	Damping factor	Home	About	contact	Product
1	0.00	0.25	0.25	0.25	0.25
2	0.05	0.274	0.242	0.242	0.242
3	0.1	0.295	0.235	0.235	0.235
4	0.2	0.333	0.222	0.222	0.222
5	0.3	0.365	0.212	0.212	0.212
6	0.4	0.393	0.202	0.202	0.202
7	0.5	0.417	0.194	0.194	0.194
8	0.6	0.438	0.188	0.188	0.188
9	0.7	0.456	0.181	0.181	0.181
10	0.75	0.464	0.179	0.179	0.179
11	0.8	0.472	0.176	0.176	0.176
12	0.85	0.48	0.173	0.173	0.173
13	0.9	0.487	0.171	0.171	0.171

Table 2: Damping factor analysis on different input

V.   CONCLUSION

The Page Ranking algorithms which are an application of web mining play a vital role to easier navigation for users. In this literature review we have discussed about Web Mining and its categorization, beside this we have explained page rank algorithm and how it employ with different concept such as number of users that visit the web pages. And also analyze the page

rank of web pages for search engine. Based on the survey, we take the different damping factor for analysis and find out that the general scenario of damping factor is 0.85 that means most probably 0.85 values is used for page ranking of web page. In the traditional Google PageRank algorithm, the damping factor is the major element to change the page ranking in hyperlink diagrams. Analysis results indicate four categories of PageRank based on the damping factor d. All websites have Minor changes in PageRank value regardless of how d changes after 0.85 value of pagerank is change slightly.

VI.   REFERENCES

[1] Parveen Rani, Er. Sukhpreet Singh: *An Offline SEO (Search Engine Optimization) Based Algorithm to Calculate Web Page Rank According to Different Parameters, INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY Vol 9, No 1, July 15 ,2013*

[2] Tamanna Bhatia," Link Analysis Algorithms For Web Mining ", IJCST Vol. 2, Issue 2, June 2011.

[3] R.Cooley, B.Mobasher and J.Srivastava, *"Web Mining: Information and Pattern Discovery on the World Wide Web". In Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), 1997.*

[4] Zdravko Markov and Daniel T. Larose*, "Mining the Web: Uncovering Patterns in Web Content, Structure and Usage Data". Copyright 2007 John Wiley & Sons, Inc*

[5] Neelam Duhan, A. K. Sharma, Komal Kumar Bhatia.*Page Ranking Algorithms: A Survey* (2009).

[6] Brin, Sergey and Page Lawrence. *The anatomy of a Large-scale hypertextual Web search engine. Computer Networks and ISDN Systems*, April 1998.

[7] Tushar Atreja, A. K. Sharma, Neelam Duhan. *A comparison study of Web Page Ranking Algorithms,IPAJOURNALS*