

# A Review of Web Page Ranking Algorithm

**Parveen Rani**

**Abstract**— The World Wide Web contains large number of information. It is very difficult for a user to find the high quality information which he wants to need. When we search any information on the web, the number of URL's has been opened. User wants to show the relevant on the top of the list. So that Page Ranking algorithm is needed which provide the higher ranking to the important pages. In this paper, we discuss the Page Ranking algorithm to provide the higher ranking to important pages.

**Index Terms**— Web Mining, Web Usage Mining, Web Structure Mining, Web Control Mining, HITS algorithm, Page Rank.

## I. INTRODUCTION

The World Wide Web is a very useful and interactive resource of information like hypertext, multimedia etc. When we search any information on the Google, there are many URL's has been opened. The bulk amount of information becomes very difficult for the users to find, extract and filter the relevant information. So that some techniques are used to solve these problems.

- *Web Mining*: Web mining is the application of Data Mining technique to find useful information from web data. With the help of web, we can access multiple data. In the distributed information environment, document or objects are usually linked together to facilitate interactive access to that we can easily access information. There are some following tasks: [2]

1. Resource finding: It is the process which involve to extract data from online or resource available on the web.

2. *Information selection and pre-processing*: The automatic selection and pre- processing of particular

information from retrieved web resources and this process transforms the original retrieved data into information.

3. *Generalization*: It automatically discovers specific patterns at individual web sites as well as across multiple sites. Data Mining techniques and machine learning are used in generalization.

4. *Analysis*: It involves the validation and interpretation of the mined patterns. It plays an important role in pattern mining. A human plays an important role in information on knowledge discovery process on web.

- *Web Mining Methodologies*

- 1) Web Content Mining
- 2) Web Structure Mining
- 3) Web Usage Mining

The following terms are described below:

- 1) *Web Content Mining*: Web Content Mining is the process of retrieving the information from web document into more structure forms. It is related to Data Mining because many Data Mining techniques can be applied in Web Content Mining.

- 2) *Web Structure Mining*: Web Structure Mining deals with the discovering and modeling the link structure of the web. This can help in discovering similarity between sites or discovering web communities.

- 3) *Web Usage Mining*: Web Usage Mining deals with understanding user behavior in interacting with the web site. The aim is to obtain information that may assist web site recognition to better suit the user. The logs include information about the referring pages, user identification, time a user spend at a site and the sequence of pages visited. There are number of algorithms proposed based on link analysis. But in this paper We are defining: Page Ranking Algorithm.

*Manuscript received Feb, 2013.*

*Ms. Parveen Rani, C.S.E. Department, Guru Kashi University, Talwandi Sabo, Bathinda, India, Mobile No 9464205973,*

## II. PAGE RANKING ALGORITHM

*Overview:* The Page Rank algorithm developed at Stanford University by Larry Page and Sergey Brin in 1996. Page Rank is a link analysis algorithm which is used by the Google internet search engine. Page Ranking is a numeric value that represents the important to a page on the web. Page Rank provides a better approach that can compute the importance of the web page by counting the number of pages that are linked. These links are called as backlinks.

If the backlink comes from an important page, those link will have higher weight than those which are coming from non-important pages. If a link from one page to another page is considered as a vote. Google uses Page Rank to show the important pages move up in the result. Google calculates the importance of the pages from the votes. It can compute the importance of the web page from the votes links. Larry Page and Sergey Brin proposed a formula to calculate the PageRank as stated below:

$$PR(A) = (1-d)+d(PR(T1)/C(T1)+\dots+PR(Tn)/C(Tn))$$

1)  $PR(T_i)$  is the PageRank of the Pages  $T_i$  which links to page A

2)  $C(T_i)$  is number of outlinks on page  $T_i$

3)  $d$  is damping factor. It is used to stop other pages having too much influence. The total vote is “damped down” by multiplying it to 0.85.

The PageRank forms a probability distribution over the web pages so the sum of PageRanks of all web pages will be one. The PageRank of a page can be calculated without knowing the final value of PageRank of other pages. It is an iterative algorithm which follows the principle of normalized link matrix of web. PageRank of a page depends on the number of pages pointing to a page. [2]

## III. LITERATURE SURVEY

Analysis of Data Mining Techniques for increasing search speed in web is proposed by “B.Chaitanya Krishna, C.Niveditha, G.Anusha, U.Sindhu, Sk Silar.”. In this paper , they are explaining the full detail of PageRank and HITS algorithm and also define the limitations, problems and comparison analysis of both algorithms.[1]

Page Ranking Algorithms for Web Mining is proposed by “Rekha Jain, Dr. G. N. Purohit”. In this paper, they are discussing and comparing the PageRank, weighted PageRank and HITS algorithm. They are compared Mining techniques, IP parameters, working, complexity, limitations and search engine of all the algorithms. HITS are used in structure Mining and Web Content Mining. PageRank and Weighted PageRank calculates the score at indexing time and sort them according to importance of page where as HITS calculates the hub and authority score of  $n$  highly relevant pages. Complexity of PageRank algorithm is  $O(\log N)$  where as complexity of Weighted PageRank and HITS algorithms are  $<O(\log N)$ . [2]

Web Mining Research: A Survey is proposed by “Raymond Kosala and Hendrik Blockeel”. In this paper, they survey the web mining categories, its methods, applications and some research issues. They also explore the relation between web mining categories and related agent paradigm. [3]

Web Mining: Methodologies, Algorithms and Applications is proposed by “Bussa V.R.R.Nagarjuna, Akula Ratna babu, Miriyala Markandeyulu, A.S.K.Ratnam”. They presented an overview of web mining, its methodologies, algorithms and applications. This paper explains methodologies and two most popular algorithms of web mining: HITS and Page Rank. By using web mining algorithms, significant patterns about the user behavior on the web can be extracted and thus improve the relationship between the website and its users. [4]

## IV. CONCLUSION

This is the survey paper of PageRank algorithm. PageRank is a better approach for calculating the page value which is a numeric value that represents the importance of a page on the web. There is given a formula for calculating the total number of pages which are linked together and counting these links as a vote. Those page will go on the top of the list which has higher number of numeric value or votes.

## V. REFERENCES

[1] Analysis of Data Mining Techniques for increasing search speed in web, “B.Chaitanya Krishna, C.Niveditha, G.Anusha, U.Sindhu, Sk Silar.”, International Journal of Modern

Engineering Research (IJMER), Vol.2, Issue.1, Jan-Feb 2012 pp-375-383.

[2] Page Ranking Algorithms for Web Mining, “Rekha Jain, Dr. G. N. Purohit”, International Journal of Computer Applications (0975 – 8887), Volume 13– No.5, January 2011.

[3] Web Mining Research: A Survey, “Raymond Kosala and Hendrik Blockeel”, ACM SIGKDD Explorations, Volume2, Issue-1, 2000.

[4] Web Mining: Methodologies, Algorithms and Applications, “Bussa V.R.R.Nagarjuna, Akula Ratna babu, Miriyala Markandeyulu, A.S.K.Ratnam”, International Journal of Soft Computing and Engineering (IJSCE), ISSN: 2231-2307, Volume-2, Issue-3, July 2012.

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

[6] Web Structure Mining Exploring Hyperlinks and Algorithms for Information Retrieval, “P Ravi Kumar and Singh Ashutosh kumar”, American Journal of applied sciences, 7 (6) 840-845 2010.

[7] D. Achlioptas, A. Fiat, A.R. Karlin and F. McSherry, “Web search via hub synthesis”, Proc. Symp. on Foundations of Computer Science, 2001.

[8] K. Bharat and M. R. Henzinge, “Improved algorithms for topic distillation in a hyperlinked environment”, ACM Conf. on Research and Develop. in Info. Retrieval (SIGIR'98), 1998.

[9] Laxmi Choudhary and Bhawani Shankar Burdak, “Role of Ranking Algorithms for Information Retrieval”.

[10] Nidhi Grover and Ritika Wason, “Comparative Analysis Of Pagerank And HITS Algorithms”, International Journal of Engineering Research & Technology (IJERT), Vol. 1 Issue 8, October – 2012 ISSN: 2278-0181



Ms. Parveen Rani , I have received my B-Tech degree in computer Engineering from Yadavindra College of Engineering, Talwandi Sabo (Bathinda) in 2011 and pursuing M-Tech degree in computer science & Engineering from Guru Kashi University, Talwandi Sabo (Bathinda). My Research areas are Internet technology and Web Mining.