Vol 3 Issue 10 Nov 2013

ISSN No : 2230-7850

Monthly Multidisciplinary Research Journal

# *Indian Streams Research Journal*

**Welcome to ISRJ**

**RNI MAHMUL/2011/38595** **ISSN No.2230-7850**

Indian Streams Research Journal is a multidisciplinary research journal, published monthly in English, Hindi & Marathi Language. All research papers submitted to the journal will be double - blind peer reviewed referred by members of the editorial Board readers will include investigator in universities, research institutes government and industry with research interest in the general subjects.



**Address:-Ashok Yakkaldevi 258/34, Raviwar Peth, Solapur - 413 005 Maharashtra, India**
**Cell : 9595 359 435, Ph No: 02172372010 Email: ayisrj@yahoo.in Website: www.isrj.net**

Indian Streams Research Journal
Volume-3, Issue-10, Nov-2013
ISSN 2230-7850

Available online at www.isrj.net

# A Comparative Analysis Of Web Page RankingAlgorithms

Bhanudas Suresh Panchabhai And Marathe Dagadu Mitharam
Head of Department , R.C.Patel A.S.C. College, Shirpur
Lecturer , R.C.Patel A.S.C. College, Shirpur

**Abstrac**t:This research paper deals with analysis and comparison of web page ranking algorithms used for Information Retrieval based on various parameter to find out their Advantages and limitations for the ranking of the web pages .On the basis of analysis of different web page ranking algorithms, a comparative study is done to find out their relative strengths and limitations to find out the further scope of research in web page ranking algorithm. Algorithms like Page Rank (PR), WPR (Weighted Page Rank), HITS (Hyperlink-Induced Topic Search), and Distance Rank and Eigen Rumor algorithms are discussed and compared.

**Keywords:**Web Mining, Web Structure, Web Graph, Link Analysis, Page Rank, Weighted Page Rank, HITS, Distance Rank, Eigen Rumor

## INTRODUCTION

As the amount of information on the internet is ever-increasing day by day so there is a brave for website owner to Offer proper and applicable information to the internet user.

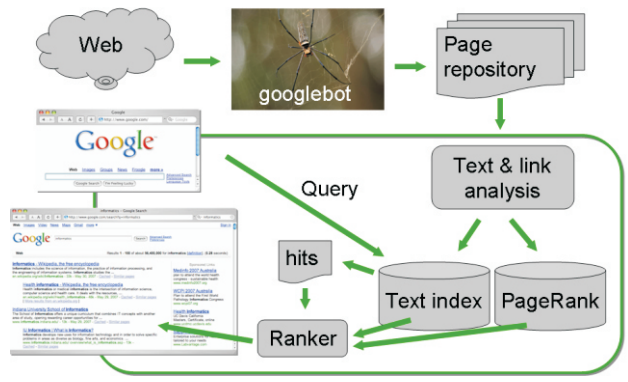


**Figure 1 .Shows a working of a typical search Engine.**

Which shows the flow graph for a searched query by a web user? An efficient ranking of query words has a major role in efficient searching for query words. There are various Challenges associated with the ranking of web pages such that some web pages are made only for navigation purpose and some pages of the web do not possess the quality of self descriptiveness. For ranking of web pages, several algorithms are proposed in the literatures. The purpose behind this research paper is to analyze the currently important algorithms for ranking of web pages to find out their relative strengths, boundaries and provide a future direction for the research in the field of efficient algorithm for ranking of the web pages [1] [2]. The remaining part of this research paper is organized as follows: Related work is Summarize in Section II. A tabular summary is presented in section III, which summarizes the techniques, working, complexity and search engine of some of the important web page ranking algorithms. Based on the literature analysis and comparison of some of various web page ranking algorithms the result is presented in section IV and a conclusion is given in section V.

## II. RELATED WORK

Web mining is the method to categorize the web pages and internet users by taking into consideration the stuffing of the page and performance of internet user in the history. Web mining helps the internet user about the web pages to be viewed in future. Web mining is made of three branches i.e. Web Content mining (WCM), Web structure mining (WSM) and Web usage mining (WUM). WCM is responsible for exploring the appropriate and related information from the contents of web. WSM is used to find out the relation between different web pages by processing the structure of web. WUM is responsible for recording the user summary and user behavior inside the log file of the web. The WCM mainly concentrates on the structure of the document whereas WSM explore the structure of the link inside the hyperlink between different documents and classify the pages of web. The number of out links i.e. links from a page and the number of in link i.e. links to a page are very important parameter in the area of web mining. The status of the web page is generally calculated by the fact that a particular page should be referred by large number of other pages and the importance of web pages may be adjudged by a large number of out links contained by a page. So WSM becomes a very important area to be researched in the field of

web mining [1] [2] [3] [4] [5]. Figure 2 shows the general classification of web mining [2].

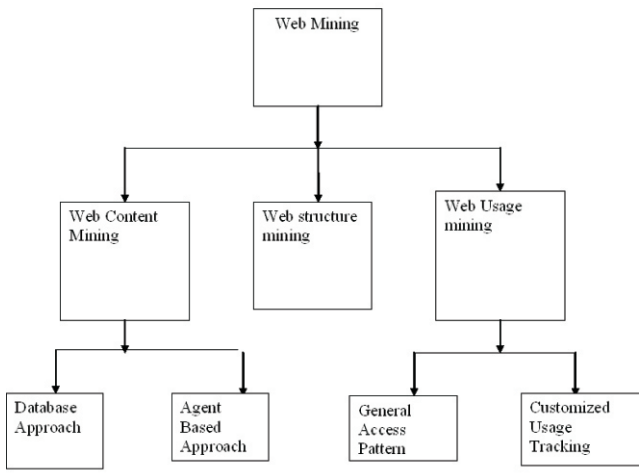


**Figure 2: Classification of Web Mining**

Here we perform the analysis of five web rank algorithms that are Page Rank, HITS, Weighted Page Rank, Eigen Rumor and Distance Rank that are used successfully and traditionally in the area of web mining.

## III. COMPARISON OF VARIOUS WEB PAGE RANKING ALGORITHMS

By extensive study and literature analysis of some of the important web page ranking algorithms, it is concluded that each algorithm has some relative strengths and limitations. A detailed comparison of ranking algorithms studied is shown in Table 1. Comparison is done on the basis of some Measures such as main techniques used, methodology, input Parameters, Complexity and search engine.

| **Algorithm** | **Page Rank** | **HITS** | **Weighted Page Rank** | **Eigen Rumor** | **Distance Rank** |
|---|---|---|---|---|---|
| **Main Technique** | Web Structure Mining | Web Structure Mining, Web Content Mining | Web Structure Mining | Web Content Mining | Web structure Mining |
| **Methodology** | This algorithm Computes the score for pages at the time of indexing of the pages. | It computes the hubs and authority of the relevant pages. It relevant as well as important page as the result. | Weight of web page is calculated on the basis of input and outgoing links and on the basis of we ight the importance of page is decided | Eigen rumor use the Adjacency matrix, which is constructed from agent to object link not page to page link. | Based on reinforcement learning which consider the logarithmic distance between the pages |
| **Input Parameter** | Back links | Content, Back and Forward links | Back links and Forward links. | Agent/Object | Forward links |
| **Search Engine** | Google | Clever | Research Model | Blog Community model | Research Model |
| **Complexity** | O(log N) | <O(log N) | <O(log N) | $0(n^2)$ | 0(log N) |
| **Invention And Year** | Larry Page & Sergey Brin in 1998 | Jon Kleinberg, 1998 | Wenpu Xing And Ali Ghorbani 2004 | Ko Fujimura et al. 2005 | Ali Mohammad Zareh Bidoki et al. 2007 |

Table 1. Comparison of five Various Web Page Algorithms

## IV. RESULTS

Based on the algorithm used, the ranking algorithm provides a definite rank to resultant web pages. A typical search engine should use web page ranking algorithms based on the Specific needs of the users. After going through exhaustive analysis of above five algorithms for ranking of web pages we got the some significant outcome as Quality of Result, Advantages, significance and limitations of web page rank algorithms. Those are shown in tabular form in Table 2.

| **Algorithm** | **Page Rank** | **HITS** | **Weighted Page Rank** | **Eigen Rumor** | **Distance Rank** |
|---|---|---|---|---|---|
| **Relevancy** | Less (this algo. rank the pages on the indexing time) | More (this algo. Uses the hyperlinks so according to the hyperlinks so according to Hen zinger, 2001 it will give good results and also consider the content of the page) | Less as ranking is based on the calculation of weight of the web page at the time of indexing. | High for Blog so it is mainly used for blog Ranking. | Moderate due to the use of hyperlinks |
| **Quality of results** | Medium | Less than PR | Higher than PR | Higher than HITS And PR | High |
| **Significance** | High. Back links are considered. | Moderate. Hub & authorities scores are utilized. | High. The pages are sorted according to importance. | High for blog ranking. | High. It is based on distance between the pages |
| **Limitation** | Results come at the time of indexing and | Topic drift and efficiency problem | Relevancy is ignored. | It is most specifically used for blog ranking not for web page ranking as other ranking like page rank, HITS | If new page inserted between two pages then the crawler should perform a large calculation to calculate the distance vector. |
| **Advantages** | Rank is calculated on the | Returned pages have high | It gives higher accuracy in terms of | Useful for ranking of blog as well as web pages because | Algorithm consider real user by |

**Table 2. Results of five Various Web Page Algorithms**

## V. CONCLUSION

After going through exhaustive analysis of algorithms for ranking

of web pages against the various parameters such as methodology,

Input parameters, relevancy of results and importance of the results,

It is concluded that existing techniques have limitations particularly in terms of time reply, correctness of results, magnitude of the results and relevancy of results. An efficient web page ranking algorithm should meet out these

Challenges powerfully with compatibility with universal standards of web technology.

## REFERENCES

[1] Wenpu Xing and Ali Ghorbani, "Weighted PageRank Algorithm", In Proceedings of the 2nd Annual Conference on Communication
Networks & Services Research, PP. 305-314, 2004.

[2] Neelam Duhan, A. K. Sharma and Komal Kumar Bhatia, "Page
Ranking Algorithms: A Survey", In proceedings of the IEEE International Advanced Computing Conference (IACC), 2009.
[3] R. Kosala and H. Blockeel, "Web Mining Research: A survey", In
ACM SIGKDD Explorations, 2(1), PP.1–15, 2000.
[4] S. Madria, S. S. Bhowmick, W. K. Ng, and E.-P. Lim, "Research
Issues in Web Data Mining", In Proceedings of the Conference on
Data Warehousing and Knowledge Discovery, PP. 303–319, 1999.
[5] S. Pal, V. Talwar, and P. Mitra, "Web Mining in Soft Computing
Framework: Relevance, State of the Art and Future Directions:, In
IEEE Trans. Neural Networks, 13(5), PP.1163–1177, 2002.

BHANUDAS SURESH PANCHABHAI
Head of Department , R.C.Patel A.S.C. College, Shirpur

MARATHE DAGADU MITHARAM
Lecturer , R.C.Patel A.S.C. College, Shirpur