Overview and Analysis of different Page Ranking Algorithms

Siddhant More[1], Vibha Bharuka[2], Prof. Tanuja Lonhari[3]

[1,2] Student, Dept. of Computer Engineering, Dr.D.Y.Patil Institute of Engineering Management and Research, Akurdi, Pune, Maharashtra, India
[3]Assistant Professor, Dept. of Computer Engineering, Dr.D.Y.Patil Institute of Engineering Management and Research. Akurdi, Maharashtra, India

---***---

Abstract – *Internet is growing so far and so wide with new reached user on internet based on mobile devices. The internet is flooding with information both relevant and irrelevant. It is challenging for web developer to implement proper links to right information using web contents and hyperlinks. This paper deals with the analysis and comparison of such Algorithms in web page.*

Key Words: HTML, CSS, JS, Page ranking, Links.

1.INTRODUCTION

Searching for a document starts with a basic keyword for querying. A search engine takes the query as an input to search the web for relevant articles. Search engine query adds to the mass of analytical data on the Internet. The more data search engines collect, the more accurate the search results become. Efficient ranking of query is important in searching of a web page. Linking and ranking webpages provide better interconnection in the intra-webs and greater usability. There are several algorithms proposed for ranking web pages that are discussed in this paper.

2. RELATED CONCEPTS

2.1 Page Ranking Algorithm

   Page Rank algorithm is the most commonly used algorithm for ranking the various pages. It is named after Larry Page. Working of the Page Rank algorithm depends upon link structure of the web pages. It weighs in the importance of the web pages by measuring them. It checks the rank score as the deciding factor. If the addition of the all the ranks of the back links is larger than the page, it gets assigned with a large rank.

Rank of a web page X is calculated by the page rank of those pages that links to page X using formula given below:

$$PR\ X = 1 - d + d\ PR(Y1)\ C(Y1) + \ldots \ldots + PR(Yn)\ C(Yn)$$

where, PR (X) = Page Rank of web page X,

PR (Yi) = Page Rank of pages Yi that links to a web page

X C (Yi) = Number of outbound links on web page

Yi d = Damping Factor (value between 0 and 1, but usually value is 0.85).

2.2 HITS (HYPERLINK-INDUCES TOPIC SEARCH)

This algorithm calculates the page rank at the query time and solves the problem of indexing time page rank calculation faced in previous page ranking algorithms. HITS algorithm ranks the web page by processing in links and out links of the web pages.

Most of the web pages act as hubs as well as authorities simultaneously. HITS algorithm is having following two steps:

   (i) Sampling Step

   (ii) Iterative Step

The weight of Hub (Hp) and the weight of Authority (Ap) can be calculated using following formulae:

$$Hp = Aq\ q \in I(p)$$

$$Ap = Hq\ q \in B(p)$$

Where, H q = Hub score of a web page

A q = Authority score of a web page

 I (p) = set of reference pages of page

 P B (p) = set of referrer pages of page

The hub weight of a web page = sum of authority weights of pages that it links to.

2.3 DISTANCE RANK ALGORITHM

In distance rank algorithm the technique of reinforcement learning is used. A factor named punishment factor is used to consider the distance between pages. Shortest logarithmic distance between two pages is used to rank them. The Advantage of this algorithm is that it can find pages with high quality more efficiently and faster by the use of distance based solution. The Limitation of this algorithm is that the crawler should perform a large calculation to calculate the distance vector, if new page is inserted between the two pages.

2.4 TIME RANK ALGORITHM

In time rank algorithm the visit time of the page after applying original web page rank algorithm to know about the degree of importance to the users is calculated. This algorithm utilizes the time factor to increase the accuracy of the web page ranking. Because of the methodology used in this algorithm, it can be called as a combination between content and link structure. The results of this algorithm are very close and regarded as perfect in agreement with the applied theory for developing the algorithm.

2.5 WEIGHTED LINK RANKING ALGORITHM

A modification of the standard page rank algorithm is given by Ricardo Baez -Yates and Emilio Davis named as weighted links rank (WL Rank). This algorithm provides weight value to the link based on three parameters i.e. length of the anchor text, tag in which the link is contained and relative position in the page. Simulation results show that the results of the search engine are improved using weighted links. The length of anchor text seems to be the best attributes in this algorithm. Relative position, which reveal that physical position does not always in synchronism with logical position is not so result oriented. Future work in this algorithm includes, tuning of the weight factor of every term for further evolution.

Table -1: Comparison Table

COMPARISON TABLE			
ALGORITHM	**PARAMETER**		**QUALITY**
	Technique	**Input Parameter**	**Relevance**	
Page rank	Web Structure Mining	Back links	Less	Medium
HITS	Web structure & content mining	Content, Back and forward links	More	Less than Page Rank
Time Rank	Web usuage mining	Original pagerank and server log	High	Moderate
Distance rank	Web structure mining	Forward Link	Moderate	High

3. CONCLUSIONS

By going through the analysis of some of the important web page ranking algorithms, it is concluded that each algorithm has been development in such a way that it targets specific users. Each algorithm has its own advantages and disadvantages. The web now growing with content of varying degrees popping up, it becomes necessary for a web developer to have knowledge of major ranking algorithm to better implement the task in hand. This paper successfully helps to summarize the techniques, advantages and

limitations of some of important web page ranking algorithms.

REFERENCES:

[1]   Dr. Daya Gupta, "User Preference Based Page Ranking Algorithm,"2016.