

# Modified Ranking Engine

Rashi Goel, Jessica Spencer, Avani Jain , Baldivya Mitra  
 Department of Computer Science and Engineering Abdul Kalam Technical University,  
 Uttar Pradesh, India

**Abstract:-** The Web consists of lots of Webpages or documents that have been published. As the internet is growing more and more day by day, it is becoming necessary to organize all the web pages in a systematic way so to make an ease for the searchers or users to get the result as per their query. So, the need of page ranking would be a great help in order to retrieve authentic data from the web. Its major purpose is to evaluate the essentiality of a web pages so, that web pages can be kept according to their priority. The proposed system is all about to remove the irrelevant pages or reducing the rank of those pages which are irrelevant so that those pages will comes at the last when someone will typed query according to their need. This proposed system is used to combine the other existing algorithm so that it covers both the old and newly introduced pages, and based on this the ranking of the web pages has to be done.

**Keywords:-** Page Rank, Damping Factor, Backlink.

## I. INTRODUCTION

There are various search engines like yandex, yahoo, bing, baidu and many more. From these the one which is utilized by most of the user or searchers is Google. Page Ranking is a part of working of search engine and we are going to focus mainly on the ranking of web pages. So, Page ranking is a way of measuring how important a

website is or a web pages. The quality of webpage is determined by the search result based on page ranking.

Page rank plays only subordinate role in the ranking of websites. The websites weighting have to done on a scale of 0 to 10. Page rank is an algorithm which was named after Larry and Brin, one of the founder of google. The main purpose is to calculate the rank of webpages i.e.PR of a page and it works by counting the number of links to a page that determines how important a website is.

## II. EXISTING WORK

As the page ranking is one of the part of search engine modules, and ranking of a page is based on the crawling and indexing of a page, so we have to understand first other modules of search engine:

- **Crawler.** A crawler crawls the web page, fetch the contents and links in the page and stores them. The developer of a website can define a file that defines how frequently a crawler is allowed to crawl and which web pages it is allowed to crawl.
- **Indexer.** Indexes the content of web page fetched by crawler. It also filters important words to be indexed out of the content.
- **Ranking.** This program basically processes the query entered by the user, fetches the indexes of the query, calculate the page rank, sort them and then displays the result.

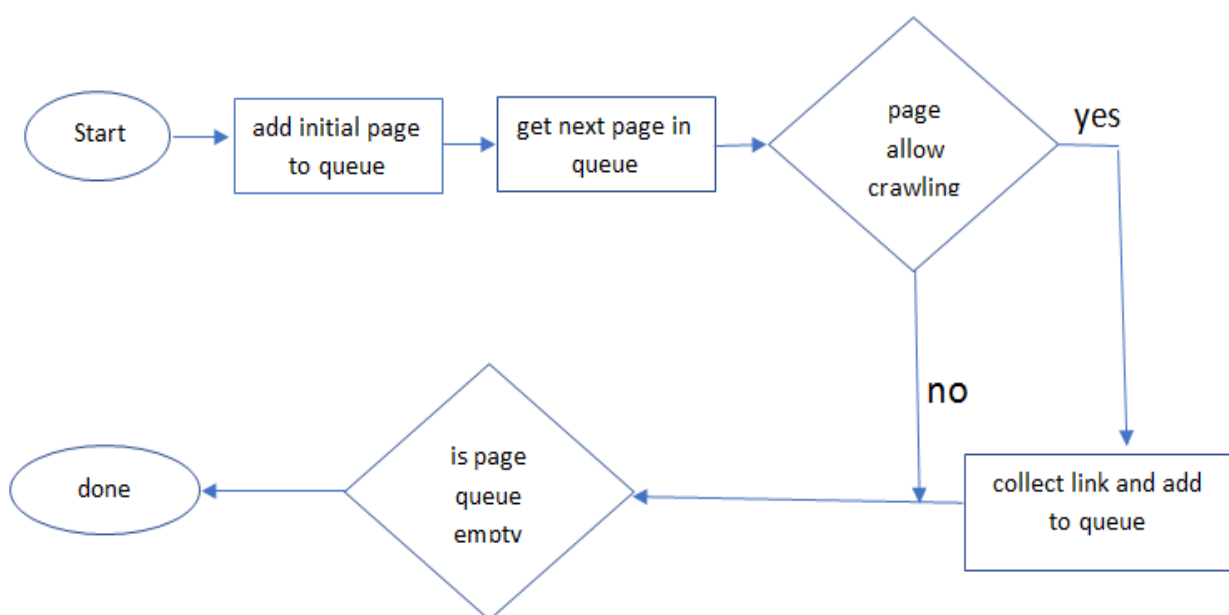


Fig 1

### III. PROPOSED WORK AND ALGORITHM

The Proposed ranking system, is all about to increase the performance of the existing ranking system, to make it more user friendly and user get a result according to their typed query with minimum latency. So, the proposed system is all about to remove the irrelevant pages or reducing the rank of those pages which are irrelevant so that those pages will comes at the last when someone typed query according to their need. This proposed system is used to combine the other existing algorithm so that it covers both the old and newly introduced pages, and on the basis of that ranking of the pages has to be done. The main aim of this algorithm is that it combine the two formulae of existing algorithm to improve the efficiency and the ranking of paged calculated with more accuracy and more faster.

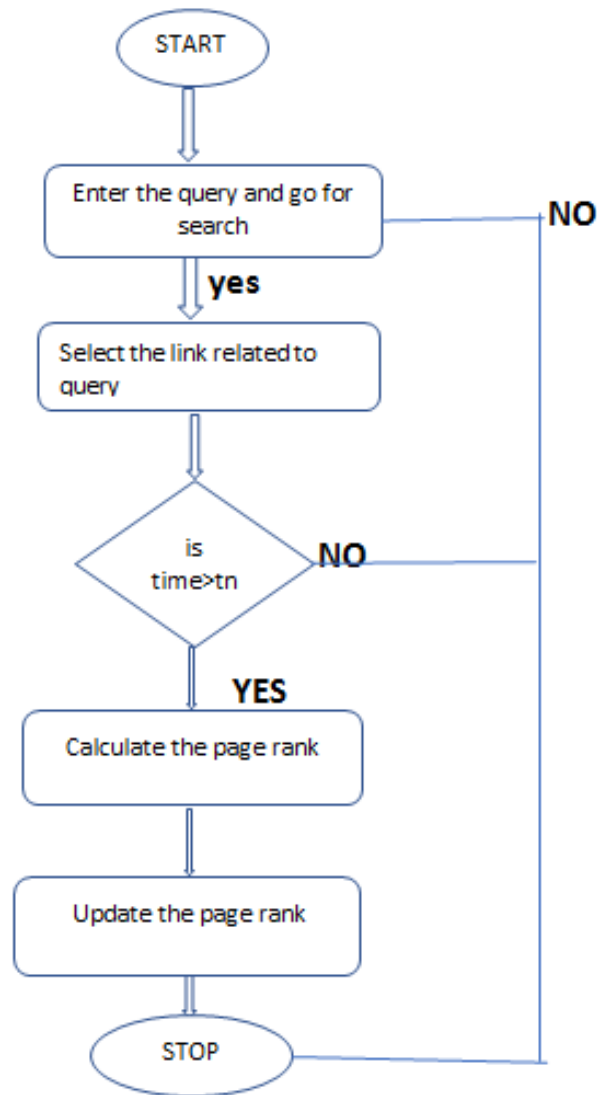


Fig 2

### IV. EXPERIMENTAL RESULT

The proposed technique is implemented in our local system. In our experiment we have created a java program for all the working. The java program is running as expected. Page rank is calculated by making a directed graph of webpages where each webpage refers to a node and a directed edge between two nodes determines that two nodes or webpages have a link between them. If a webpage has no link to other page than it become a sink and therefore will terminate the random surfing process.

```

Problems @ Javadoc Declaration Console
<terminated> PageRanking [Java Application] C:\Program Files\Java\jre1.8.0_211\bin\javaw.exe
Enter the connection links of each webpage with the other webpages :

1 0 1 1 0
0 1 0 0 1
0 0 1 1 0
1 1 1 1 0
0 1 0 1 0
    
```

Fig 3

```

Problems @ Javadoc Declaration Console
<terminated> PageRanking [Java Application] C:\Program Files\Java\jre1.8.0_211\bin\javaw.exe (02-May)
Initial PageRank Values , Zeroth Step
Page Rank of A is : 0.2
Page Rank of B is : 0.2
Page Rank of C is : 0.2
Page Rank of D is : 0.2
Page Rank of E is : 0.2
    
```

Fig 4

<pre> Problems @ Javadoc Declaration &lt;terminated&gt; PageRanking [Java Appl] After 1th Step Page Rank of A is : 0.06 Page Rank of B is : 0.16 Page Rank of C is : 0.16 Page Rank of D is : 0.4 Page Rank of E is : 0.2         </pre>	<pre> Problems @ Javadoc Declaration &lt;terminated&gt; PageRanking [Java Applic] After 2th Step Page Rank of A is : 0.1333 Page Rank of B is : 0.2333 Page Rank of C is : 0.1666 Page Rank of D is : 0.3000 Page Rank of E is : 0.1666         </pre>
--	--

Fig 5

```

Problems @ Javadoc Declaration Console
<terminated> PageRanking [Java Application] C:\Program
Final Page Rank :
Page Rank of A is : 14.00 %
Page Rank of B is : 23.00 %
Page Rank of C is : 17.00 %
Page Rank of D is : 29.00 %
Page Rank of E is : 17.00 %
    
```

Fig 6

**V. LITERATURE SURVEY**

- *Weighted Page rank.*
- Weighted Page Rank algorithm is an extension of the Page-Rank algorithm.
- This algorithm allocates a higher rank values to the more significant pages rather than dividing the rank value of a page evenly among its outgoing linked web pages.
- This algorithm takes into account the importance of both the in links and out links of the pages and distributes rank scores based on the popularity of the pages.

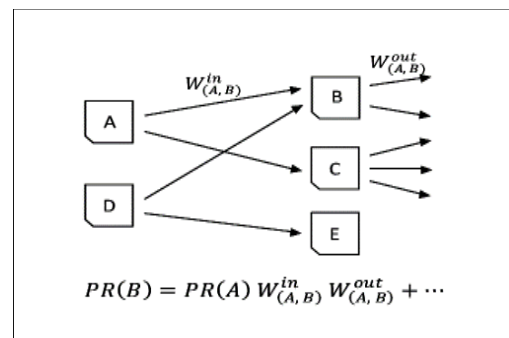


Fig 7

➤ *HITS Algorithm.*

- HITS algorithm is a link analysis algorithm.
- John Kleinberg proposed an algorithm which is a link analysis algorithm that rates the web pages .
- This algorithm is also known as Hubs and Authorities.
- Hubs are webpages that have outlinks and Authorities are webpages that have inlinks.
- This Algorithm iteratively find hub and authority scores at the webpage.
- HITS is a query dependent algorithm.
- The conclusion of this algorithm is that it assigns two scores for each webpage, its authority which estimate the value of the content of the webpage and its hub value which will find the value of its links to the other webpages.

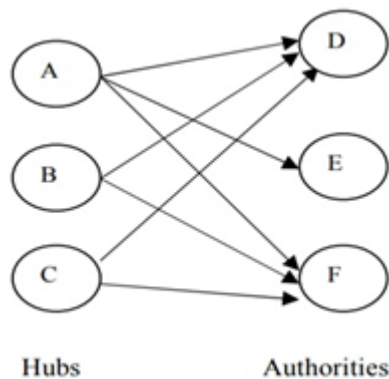


Fig 8:- Hubs and Authorities Link Structure

## VI. CONCLUSION

Finally, we can say that we have seen the working, purpose, algorithm, of the proposed work. The basic idea is to improve the performance of the algorithm as the proposed work combines the two formulae of damping factor so to get ranking of web pages in a more accurate way and also many different algorithms has been introduced and also their implementation. The other algorithms focuses on different things such as weighted page rank focuses on forward and backward links and quantum page rank focuses on how to retrieve the links.

## REFERENCES

- [1]. [www.google.org](http://www.google.org)
- [2]. [www.geeksforgeeks.org/ranking-google-search-works/amp/](http://www.geeksforgeeks.org/ranking-google-search-works/amp/)
- [3]. [www.en.m.wikipedia.org](http://www.en.m.wikipedia.org)
- [4]. Ridings and M. Shishigin, "page rank unrecovered", technical report 2002
- [5]. C.Riding "webpage rank unrecovered" technical report,2002
- [6]. R.Cooley,"Web mining: Information and pattern discovery on the world wide web proceeding of the 9<sup>th</sup> IEEE .
- [7]. [www.checkpagerank.net](http://www.checkpagerank.net)
- [8]. facts about google competition archived from the original on November 4, 2011

- [9]. Weighted page rank algorithm based on in-out weight of web pages, B.Jagannathan and kalyani desikan, Indian Journal of Science and Technology, december 2015.
- [10]. Effective utilization of hits in significant information retrieval, Shailendra G.Pawar and Pratiksha Natani, International conference for Conversions of Technology, 2014.