

Improved Web Page Ranking Algorithm Using Semantic Similarity and HITS Algorithm

Nisha Bansal¹ and Dr. Paramjeet Singh²

^{1,2}Giani Zail Singh College of Engineering, Bathinda
CSE Department

Abstract

People generally access the information over the internet with the help of search engines. Search engines are the programs which find the specific pages for users according to their query. Web page ranking is the most important factor on internet for search engines. Web page ranking is a technique that ranks the web pages according to their different qualities and parameters for search engines. There are various web search engines available on internet some of them are Google, Yahoo, and Bing etc. In this paper, we present a new web ranking system by using Semantic Similarity and HITS algorithm along with AI technique. These techniques work together to rank a web page from a number of web pages on the internet.

Keywords: SEO, Search engine Optimization, Web page ranking algorithm, HITS algorithm, Semantic Similarity Algorithm

1. INTRODUCTION

Now a day's searching on the internet is most widely used operation on the World Wide Web. The amount of information is increasing day by day rapidly that creates the challenge for information retrieval. There are so many tool for execute well-organized searching. Due to the volume of web and requirements of users, web search engines produce so many results in response to user query. If user sends the query for particular topic, then Web can have hundred, even thousands results in response to that query. But if the Ranking system does not present the result within the top few positions of the ranking then that is not useful for user. Following diagram shows the working of search engine.

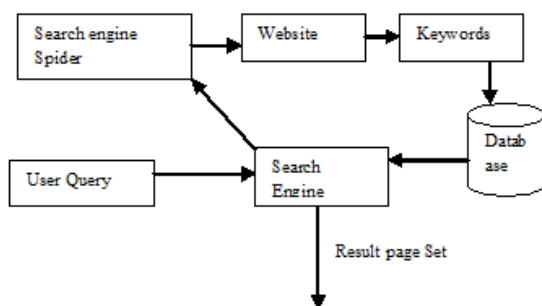


Figure 1 working of search engine

This also creates the challenge for search engine page ranking systems. Web page Ranking is the main part of any information retrieval system. Web page ranking system ranks the web pages. There are following working

flow of ranking

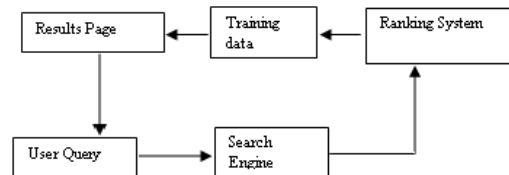


Figure 2 working of ranking system

This ranking depend upon different factors like No. of occurrences of Keyword in web Page, No. of In links and no. of Out links etc. By calculating these factors, ranking system give rank to the web page. These web pages displayed in search engine according to rank.

2. LITERATURE SURVEY

[1] P. Chahal, M. Singh and S. Kumar "Ranking of Web Documents using Semantic Similarity"

This paper proposed a novel technique which makes user search data quite efficient. This technique gives a relationship or similarity between searched document and user query. It is also consider the semantic structure of document and user query. The result set obtained from this approach gives better results than prevailing approaches. The future work can be done by using deeply semantic analysis of web pages and relevance of documents.

Gyanendra Kumar, Neelam Duhan and A. K. Sharma "Page Ranking Based on Number of Visits of Links of Web Page"

In this paper author presented a modified page ranking algorithms which is more target oriented than original page rank. The modified algorithm calculates page rank value or importance of web pages based on the visits of incoming links on a page. The paper presented a novel page ranking algorithm called VOL that provides more relevant results than original Page Rank. As a result, Author proved that VOL is far dynamic than original Page Rank algorithm and also observed that the page which has more visits of incoming links is carrying more rank value than less visited pages. The paper also presents a method to find link-visit counts of Web pages and a comparison between VOL with the Page Rank algorithm.

Parveen Rani and Er. Sukhpreet Singh, "An Offline SEO (Search Engine Optimization) Based Algorithm to Calculate Web Page Rank According to Different Parameters"

This paper describes the new algorithm for calculating web page rank according to different parameters. The proposed algorithm called M-HITS (Modified HITS) is a

new version of HITS algorithm. It is developed by extending the properties of HITS algorithm. Author present new algorithm in which six parameters are used to evaluate rank for web page. Future work can be done by using some AI techniques in addition to these proposed techniques to improve the rank of web pages.

Ashish Jain, Rajeev Sharma, Gireesh Dixit and Varsha Tomar “Page Ranking Algorithms in Web Mining, Limitations of Existing methods and a New Method for Indexing Web Pages”

This paper proposed a new method called Intelligent Search Method (ISM). Author developed new method to index the web pages using an intelligent search strategy in which meaning of the search query is interpreted and then indexed the web pages based on the interpretation. This Paper also described the limitations of existing methods and discussed the different algorithms used for link analysis like Page Rank (PR), Weighted Page Rank (WPR), Hyperlink-Induced Topic Search (HITS) and CLEVER algorithm. The new method can be integrated with any of the Page Ranking Algorithms to produce better and relevant search results.

3. PROPOSED WORK

In this paper, author proposed a ranking system with improved HITS and Semantic Similarity techniques.

It is used to rates the web pages. It is also known as Hubs and Authorities. A good hub represented a page that pointed to many other pages and a good authority represented a page that was linked by many different hubs. So its authority value, which estimates the value of the content of the page, and its hub value, which estimates the value of its links to other pages. We will use C# and ASP.net to implement the algorithm. A Graphical User Interface (GUI) will be created to display the results.

Steps of semantic similarity Algorithm:

- Step 1:** Firstly construct a Text-List (by links).
- Step 2:** Then acquire query as a text: a String.
- Step 3:** For each Text in Text-List do:
 - (a) Create Text-Vector-Space.
 - (b) Create Domain-Dictionary of words.
 - (c) Using Statistical-Model () and Domain-Dictionary, Compute relevance-value of Text corresponding to user Query.
 - (d) Make Domain-Ontology of the Text.
 - (e) Compute Domain-Similarity of Text value with Domain- Ontology.
 - (f) Verify the maximum of Domain-Similarity value and relevance-value and call it Relevance-Score.

Step 4: Then Go to step 3 until there are no text left in the Text-List or else no more text is to be considered.

Step 5: Organize the links according to the decreasing order of relevance-score and assign the rank to them.

Step 6: And then finally display the contents according to their ranks.

Steps of HITS Algorithm

Step 1: Enter the adjacency matrix of the web pages.

Step 2: Then enter the frequency of different types of parameters (Bold, Italic, Keyword, and No. of Unique Click).

Step 3: Compute the Hubs and Authorities for each web page.

Step 4: Normalize all these values for every web page and then compute the partial rank for every web page.

Step 5: And then add weights of the parameters to the calculated partial rank.

Step 6: Sort the web pages positions according to the calculated ranks corresponding to both Hub Values and Authority Values of web pages.

Step 7: Exit.

4. Results and Discussion:

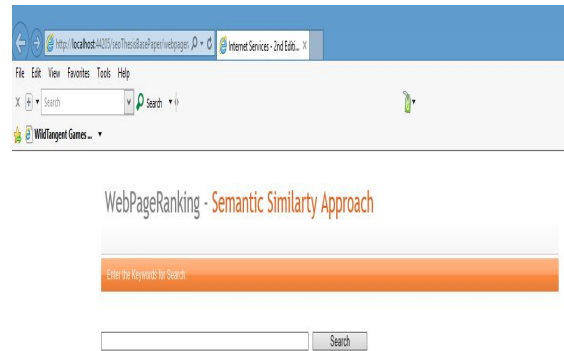


Figure 3 Input Screen of proposed system

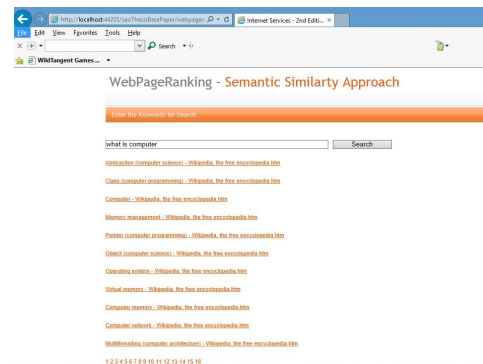


Figure 4 Proposed system with result

Above figure shows the result of proposed system in which when user query “What is computer” is enter in search box, it provide a set of web pages corresponding to given query.

Parameter\Technique	HITS Algorithm	Semantic Similarity Algorithm	Proposed System
Time Efficiency	72%	87%	91%
Accuracy	79%	91%	95%
User specific Page Generation	No	No	Yes
Relevance Ratio	90%	92%	96%
High Relevance Ratio	30%	41%	51%

Table 1: Comparison of Techniques

Graph 1:

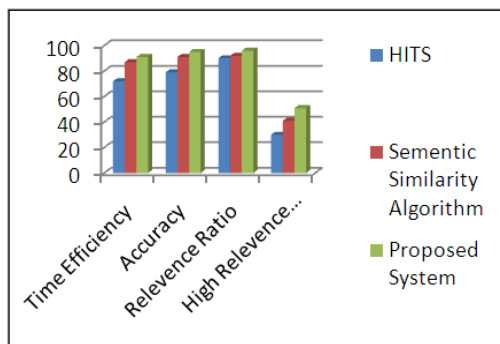


Figure 6: Graph showing different parameters of techniques.

5. Conclusion and Future Scope

Proposed system presents an improved semantic similarity technique to rank a web page from a set of given web pages. System is tested on 1000 web pages comes under various categories like education, computer, programming, chemistry etc. Various input queries are given as an input to the system and results are checked. System can be further improved by implementing on the cloud servers and by using multithreading techniques to improve the time efficiency. System can be further checked by increasing the number of web page categories.

References

[1] G. Kumar, N. Duhan and A. K. Sharma “Page Ranking Based on Number of Visits of Links of Web Page” Computer and Communication Technology (ICCT), 2011 2nd International Conference on ©2011 IEEE, DOI 10.1109/ICCT.2011.6075206 Page(s): 11 – 14

[2] P. Rani and 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

[3] N. Batra, A. Kumar, Dr. Dheerendra Singh and Dr. R.N. Rajotia “Content Based Hidden web Ranking Algorithm(CHWRA)” Advance Computing Conference (IACC), 2014 IEEE International, © 2014 IEEE, DOI 10.1109/IAdCC.2014.6779390 Page(s): 586 – 589

[4] A. Jain, R. Sharma, G. Dixit and V. Tomar, “Page Ranking Algorithms in Web Mining, Limitations of Existing methods and a New Method for Indexing Web Pages” Communication Systems and Network Technologies (CSNT), 2013 International Conference on, © 2013 IEEE, DOI 10.1109/CSNT.2013.137 Page(s): 640 – 645

[5] H. Dubey, Prof. B. N. Roy “An Improved Page Rank Algorithm based on Optimized Normalization Technique” (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (5), 2011, 2183-2188

[6] R. Kumar and S. Saini “A Study on SEO Monitoring System Based on Corporate Website Development” International Journal of Computer Science, Engineering and Information Technology (IJCEIT), Vol.1, No.2, June 2011

[7] K. ur Rehman and M. N. Ahmed Khan “ The Foremost Guidelines for Achieving Higher Ranking in Search Results through Search Engine Optimization” International Journal of Advanced Science and Technology Vol. 52, March, 2013

[8] P. Chahal, M. Singh and S. Kumar “Ranking of Web Documents using Semantic Similarity” Information Systems and Computer Networks (ISCON), 2013 International Conference on ©2013 IEEE, DOI 10.1109/ICISCON.2013.6524191 Page(s): 145 – 150

[9] N. V. Pardakhe, Prof. R. R. Keole “Analysis of Various Web Page Ranking Algorithms in Web Structure Mining” International Journal of Advanced Research in Computer and Communication Engineering Vol.2, Issue 12, December 2013

[10] M. Cui, S. Hu “Search Engine Optimization Research for Website Promotion” Information Technology, Computer Engineering and Management Sciences (ICM), 2011 International Conference on (Volume:4) © 2011 IEEE, DOI 10.1109/ICM.2011.308 Page(s): 100 - 103