



Security, Efficiency and User Interest Analysis Approach for Web Page Ranking

¹Pooja, ²Hitesh Kapoor, ³Amit Raman

¹Student M.Tech. (CSE), ²Asstt. Prof., CSE Deptt., ³Asstt. Prof., Mech. Deptt.
RPIIT, Technical Campus, Bastara, Karnal

Abstract: An Effective Web Search is one of the major part of Web architecture. The presented work is the about to define a user assisted WebPage ranking approach so that that the better view of web search will be presented to the user. The presented work is defined under three vectors. At first when the query is performed, the initial ranking is based on the relevancy vector i.e. content matching respective to the user query. Now as user view these Pages, according to the user view the ranking is modified. The second vector of ranking is the number of times a Page is visited by any user. The most important criteria of the ranking is the user recommendation. User recommendation is defined in terms of like and dislike vector. These three parameters are taken in weighted form and based on these three vectors, overall Page ranking is decided.

Keywords: Web Page Ranking, Security, Relevancy, Efficiency

I. INTRODUCTION

Web mining consists of a set operations defined on data residing on WWW data servers defines web mining as "...the discovery and analysis of useful information from the World Wide Web". Such data can be the content presented to users of the web sites such as hyper text markup language (HTML) files, images, text, audio or video. Also the psychical structure of the websites or the server logs that keep track of user accesses to the resources mentioned above can be targets of web mining techniques. Web mining as a sub category of data mining is fairly recent compared too their areas since the introduction of internet and its widespread usage itself is also recent. However, the incentive to mine the data available on the internet is quite strong. Both the number of users around the world accessing online data and the volume of the data itself motivate the stakeholders of the web sites to consider analyzing the data and user behavior. Web mining is mainly categorized into two subsets namely web content mining and web usage mining. While the content mining approaches focus on the content of single WebPages, web usage mining uses server logs that detail the past accesses to the web site data made available to public. Usually the physical structure of the web site itself which is a graph representation of all WebPages in the web site is used as a part of either method. However recent approaches that appoint more focus on the physical link structure of the web site have introduced web structure mining as a separate concept[11].

A. Web Page Ranking

The search is the web tool available over the web to search some information over the web. It basically provides the information retrieval over the web by performing the web server selection. It performs an effective search over the web by performing the url selection and the display of the information over web. The information accessed from the system is represented in the form of hyperlinks and this information is provided by Web Page Identification significantly. Now the Web Page Identification process has been improved with the inclusion of different filtration tools associated with the system. The task associated with a Web Page Identification are given as under

1. Analyze the user query and perform the search over web respective to requested query contents and present the results effectively.
2. Web Page Identification maintains an index table of extracted urls that represents what text is identified on which URL.
3. It executes the query at different combinations so that more effective results will be drawn from the system.

These all are the basic approach used by the Web Page Identification. But even then these Web Page Identifications are required to be more efficient and more accurate to user query. Because of this it is the open research area and lot of research is been done in same area. The presented work is also the improvement of this we search with the inclusion of user interest based ranking system.

B. Ranking

A search provides the information over the web by performing the effective search under the algorithmic implementation. Based on this implementation, the WebPage retrieval is performed. Once the Pages are retrieved, it is required to assign the rank to these WebPages. To assign the rank there are number of mathematical as well as algorithmic approaches. These algorithmic approaches are about to perform the calculation based on the relevance vector in terms of keyword search over the particular web domain. The next

work is about to present the result in a ranking way. The more relevant and the required information will be assigned by higher rank. The usage of a web domain is also the parameter to decide the ranking. There are number of SEO application and the professionals that can perform the allotment of the web sites or the web servers so that web information retrieval will be performed effectively. The SCO is basically the mechanism to improve the ranking of the WebPage so that the particular web domain link will be identified over the web easily and the results will be driven at the upper level.

II. REVIEW OF LITERATURE

In Year 2011, D. Choi has defined an approach to perform the query over the web and to extract the web document. The author also presented the approach to assign the ranking to these web documents. With the development of web Web Page Identifications, one of the major tasks is to retrieve these documents from web effectively. These Web Page Identifications uses the some ranking algorithm to present the result in an effective way. The author has defined a study of existing ranking algorithm used by different Web Page Identifications. The author explored the advantages and limitations of these ranking algorithms. The major contribution of author was the definition of query based information retrieval. The author defined the classification over a query and performed the query filtration. Based on this analysis, the ranking is improved and refined [1]. Another work on WebPage ranking is defined by Z. Anari in 2009. The author has defined a fuzzy based Page ranking system under the concept of Automata. The author has defined a ranking system to identify the similar Pages over the web. In his work, author had presented an algorithm based on automata learning that defined the automata concept to perform the assignment of WebPage. The author defined the function for determining weight of internal links in a WebPage. The author defined some parameters respective to the time duration and relevancy of the WebPage. In this work, author proposed a fuzzy based system that categorized the fuzzy based variable interference engine to find the WebPage rank. The author has calculated the rank of WebPage under the recursive analysis of internal links. The author presented the results in terms of improved efficiency while performing the WebPage retrieval from the system [2].

Another work in the area of WebPage ranking is performed by H. Wu. The author has presented a collaborative ranking approach to provide the WebPage ranking respective to the recommendation process. Such kind of systems required the collaboration in terms of access time to the system. The author also introduces the WebPage ranking that will perform the substitution in terms of availability to the system. The author has performed the experimentation on the real world dataset. The obtained results show the promising effect of collaborative ranking for WebPages [3]. Another work on ranking system was defined by A. Altman in year 2005, the author defined fundamentals of basic WebPage retrieval and the ranking system. The author defined WebPage ranking under the formal theory of social choice. The author an axiom based analytical approach to perform the WebPage ranking. This ranking theory was inspired from the mathematical model where the theoretical and ordinal knowledge is collected together to present the effective WebPage ranking. Author has tested the ranking system for ecommerce web sites[4].

One of the common model for WebPage ranking and prediction system is defined by Markov Model. Such model defines the navigational behavior of web graph theory as well as defines the transitional probabilities over the ranking analysis. The author not only defined work for a single WebPage access, but also presents the work for web path generation. The web path is actually defined as a series of WebPages that a user can visit after visiting a specific WebPage. To perform this kind of analysis a Markov Model based prediction system is defined. The prediction is here defined under the web usage mining that defined the structural information for prediction of WebPages. To perform such kind of analysis, the author defined a WebPage graph and implements the markov model over it to analyze the frequency match. Based on which a acyclic web path is generated and based on the weightage assigned to this web path the prediction is performed [5]. Another work in WebPage ranking is the comparison of different WebPages and the web sites. Author M. Klein performed this comparison on two football team web sites of college team. The analysis is performed under the WebPage metrics to perform the quality assessment. The author has defined the Page comparison and the ranking system under the graph theory [6].

In year 2011, B. Gao defined a Semi Supervised ranking approach on the graph theory based on metadata information. On this graph based ranking, the social networking sites are recognized. These graph based structure defines the WebPages in the form of nodes and the transmission between these WebPages and the WebPage association is presented in the form of edges as well as edges weightage. The author also defined a comparative analysis with existing ranking approach. Based on this study, author presented a more specific and accurately scoring mechanism for WebPage ranking. The author defined the problem under a generic framework so that robustness in graph based ranking can be achieved. The proposed ranking algorithm is called semi-supervised ranking approach. In this algorithm, the markov model is been implemented at the later stage to improve the ranking results [7]. One more work on query based ranking and matching approach over the WebPage is defined by M. Murata in year 2009. The author presented the work on the matching analysis on the basis of user clicks to the specific title. The visit count to a specific WebPage is used as the main decision

parameter for deciding the WebPage ranking. Another improvement included by the author is the frequency based analysis while performing the web search. The author has defined a text based retrieval system that will include the query keywords and will return the relevant search results under the user interest analysis. The user interest is here defined as the number of clicks performed by the user for a specific title. The author performs the query filtration as well as the classification. This query filtration process reduced the size of actual query [8].

III. RESEARCH METHODOLOGY

In this present work an improved user assisted WebPage ranking approach is suggested so that the search results will be shown in an improved form. The work is presented as a Web Page Identification optimization model where the user assistance is taken in the form of user recommendation and based on the user assistance effective results are presented by the user. The work is divided in two main phases. The first phase of the work is similar to the traditional Web Page Identification where user query is performed over the web to identify the most relevant WebPages. A relevancy check is performed with searched Page and based on this relevancy vector initial ranking to the web link is performed. This obtained ranking is the key vector and stored in the database. Once the search results are presented to the user, the user input is taken in the different forms and based on this user input the Page ranking at the later stage is decided. The user assistance is taken under two main vectors called visited Page list and the user preference vector. The visited Page count is defined in terms of number of user clicks on the particular web link. Another vector considered is the user recommendation.

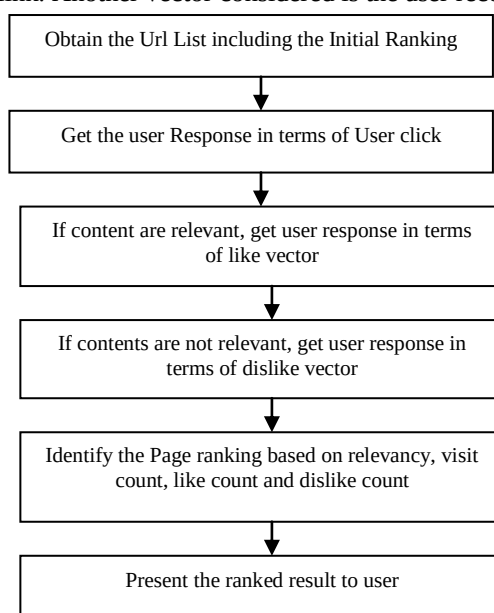


Figure 1: Proposed Model

IV. RESULTS

The first step of proposed work is to perform the user query on the web. To obtain the relevant web links at the initial phase, the search engine based query is performed. ASP.net provides such library to perform the web request and obtain the relevant links. These links are taken using the Google search engine. The obtained results from Google query are the initial page list on which the relevancy check is performed. The work is performed dynamically, it means so specific limit on keywords or the pages is implemented. The work will be implemented on the Google directly and obtain the web links and present it as the initial relevant search. The work is then applied with modified ranking approach under security, efficiency and user interest parameters. The results are here shown as under

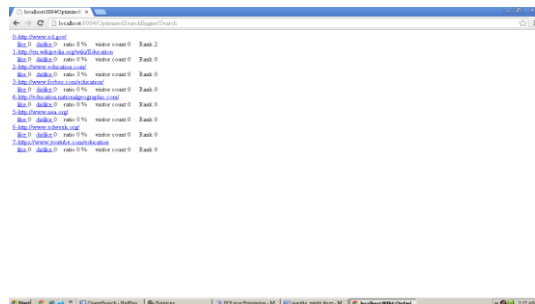


Figure 2 : Web Page Ranking

Here figure 2 is showing the ranked results obtained from the work.

IV. CONCLUSION

The presented work is about to perform the effective search in the Web environment based on the user query relevancy factor. The relevancy of the query is here analyzed under three main factors called Keyword based Analysis, User Recommendation Analysis and the User Web Page visit analysis. Based on these all factors a ranking criterion is decided and based on these ranking vectors the Web Pages are ordered. The user can get the best Web Page as well as recommend other for the best Page selection.

REFERENCES

- [1] Donjung Choi, "An Approach to Use Query-related Web Context on Document Ranking", ICUIMC '11, February 21–23, 2011, Seoul, Korea. ACM 978-1-4503-0571-6
- [2] Zohreh Anari, "WebPage Ranking based on Fuzzy and Learning Automata", MEDES 2009, October 27-30, 2009, Lyon, France. ACM 978-1-60558-829-2/09/0010
- [3] Honghan Wu, "Ranking Domain Objects by Wisdom of WebPages", WIMS'12, June 13-15, 2012 Craiova, Romania ACM 978-1-4503-0915-8/12/06
- [4] Alon Altman, "Ranking Systems: The PageRank Axioms", EC'05, June 5–8, 2005, Vancouver, British Columbia, Canada. ACM 1-59593-049-3/05/0006
- [5] Magdalini Eirinaki, "Web Path Recommendations based on Page Ranking and Markov Models", WIDM'05, November 5, 2005, Bremen, Germany ACM 1-59593-194-5/05/0011
- [6] Martin Klein, "Comparing the Performance of US College Football Teams in the Web and on the Field", HT'09, June 29–July 1, 2009, Torino, Italy. ACM 978-1-60558-486-7/09/06
- [7] Bin Gao, "Semi-Supervised Ranking on Very Large Graphs with Rich Metadata", KDD'11, August 21–24, 2011, San Diego, California, USA. ACM 978-1-4503-0813-7/11/08
- [8] Masaya Murata, "Query-Page Intention Matching using Clicked Titles and Snippets to Boost Search Rankings", JCDL'09, June 15–19, 2009, Austin, Texas, USA. ACM 978-1-60558-322-8/09/06
- [9] Elisabetta Fersini, "Granular Modeling of Web Documents: Impact on Information Retrieval Systems", WIDM'08, October 30, 2008, Napa Valley, California, USA ACM 978-1-60558-260-3/08/10
- [10] Rajeev Motwani, "Evolution of Page Popularity under Random Web Graph Models", PODS'06, June 26–28, 2006, Chicago, Illinois, USA. ACM 1-59593-318-2/06/0006
- [11] Srinivas Vadrevu, "Identifying Primary Content from WebPages and its Application to Web Search Ranking", WWW 2011, March 28–April 1, 2011, Hyderabad, India. ACM 978-1-4503-0637-9/11/03