

## International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IMPACT FACTOR: 5.258

*IJCSMC, Vol. 5, Issue. 3, March 2016, pg.766 – 771*

# A Study on Different Ranking Factors and Methods for Search Engine Optimization

**Dr. Vandana<sup>1</sup>, Vaishali Balhara<sup>2</sup>**

<sup>1</sup>Asstt. Prof., Comp. Sc. Dept., PT. N.R.S. Govt. College, MDU, Rohtak

<sup>2</sup>Asstt. Prof., PG Dept. of Computer Science, All India Jat Heroes Memorial College, MDU, Rohtak

<sup>1</sup> [vandanamukeshmalik@gmail.com](mailto:vandanamukeshmalik@gmail.com); <sup>2</sup> [vaishali\\_leo@rediffmail.com](mailto:vaishali_leo@rediffmail.com)

---

**Abstract**— Today all the personal, public, government and organization information is available through web and provided in the form of web pages. Instead of remembering these all web links directly, some search tool can be used to identify the list of required links. There are number of such tools that provide the results based on different parameters and ranking algorithms. In this paper, the scope of these tools is been explored under the parameter specific algorithmic analysis. The study on standard search architecture is provided in light of these algorithms which can provide accurate and relevant results.

**Keywords**— Search Engine, Crawler, Index Table, Machine Learning

---

## I. INTRODUCTION

A search engine[1][2][3][4][5][6] is considered as a tool which is available to search any information on this global network. The search engine can be application specification, domain specific or the environment specific. Today each organization or the department also having the internal search engines to ensure the consistent help to the users. The relevancy and support of a search engine is considered valuable if it provide more relevant and significant results in effective time. But as the web is a global environment, it also suffers from various irregularities relative to the environment and search constraints. There also exist number of fake and irrelevant domains, pages and web information. The major challenge to search engine tool is to provide the search after filtering these spam and irrelevant pages. Another challenge to the search engine is its vast repository. As the web is having crores of documents, it is not easy to observe all the pages. Because of this, there is requirement of some fast similarity analysis method that can map the requirement from available pages in significant time and provide the efficient results on time. The basic architecture of search engine tool is shown in figure 1. The figure shows that the user query is first processed under query mining or text mining process which is generated in the form of structured query. The web information is drawn from the index table via a web crawler. The search process is about to map these crawled results taken from index table under query processor to generate the accurate search results.

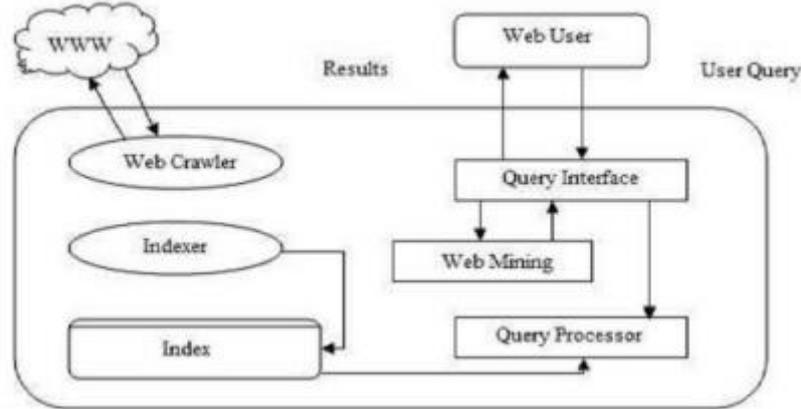


Figure 1 : Search Engine Architecture[4]

The search process also required to follow some of the basic characteristics map including the user query similarity map, user context information map etc. The server selection for performing the search based on user profile information is one such requirement. The effective server selection will not only reduce the server load but provide the region driven relevant results. The search method also required to follow the map constraint over the URL, title, headers and the contents. Each of the content part having its own significance and play effective role to generate more accurate and effective results. The hyper link based significant results can be deriving to improve the accuracy of search process. The search engine is required to generate a detailed list of feasible pages and then apply filtration at different parameters and level to generate more accurate ranking results. The tasks associated under search engine optimization are listed here under

1. Analyze the query at lower level and generate the associated keywords. Now map these keywords to the web page content and identify the similarity results.
2. Generate and maintain an index table to store the existing common content query results so that more accurate mapping of common queries will be done.
3. Apply different query combinations to generate more accurate and significant results.
4. Setup the parameters relative to the user profile or context to generate more accurate and filtered results.
5. Some time and reliability driven filters can be applied to generate accurate relevant results.
6. The application domain, user domain based analysis can be regulated to avoid the irrelevant mapping.

The search process applied over the linked pages upto deep level can be controlled and directed by using an appropriate crawler. The crawler requirement and features are discussed in next sub section.

#### A. Web Crawler

Web crawler is the integrated tool for any web processing to move between the web pages under relevancy vector. It processed in a graph formed structure to identify the interconnected links. The seed URL is taken by the crawler as the primary search map which is later on processed by direction given by robot.txt file. The permission and the information file processing is applied to gain the crawler impact. The pages are crawled to extract the URL from the content page and process them recursively. The extracted links are stored in a queue so that the same process will be applied on internal links also. The challenge of this crawler is to process the high size repository and generate the linked relevant index table. The topic driven mapping can be applied on online search which can be applied on web portals to generate the profile specific information. There are different form of web crawler based on the application and domain requirement. The outcome of a search engine is considered in the form of index table. The number of levels decided to generate the link identify the size of index table. For the accurate and efficiency search process, an accurate index table is required as the input database.

In this paper, an exploration to the search engine is provided with specification of the impact of ranking algorithms and the associated parameters. The paper has challenges and scope of search engine for real time applications. In this section, the search engine tool, its features and the basic architecture is discussed. In this section, the search impact and associated taxonomy with crawler is discussed. In section II, the work defined by earlier researchers is presented. In section III, the ranking factors and some of ranking algorithms are discussed with feature exploration. In section IV, the conclusion of the work is presented.

## II. RELATED WORK

Search engine optimization is required to obtain the effective page access based on requirement observation and prediction. A filtered search and intelligent ranking method is required to avoid the bad, duplicate and spam pages. In this section, the methods and the parameters considered by different researchers for improving the ranking mechanism are discussed. Ashlesha *et. al.*[1] has provided the work on a new page ranking algorithm based on the feedback and user preference analysis. Author applied the structured similarity analysis with domain observation and user time spend on a page to generate the page ranking. The quality method has provided the deeper observation to generate appropriate user specific ranking. Mercy Paul Selvan *et. al.*[2] included various parameters including the rechability, feedback and value for improving the ranking method. Author applied the interest level observation to improving the search results and incorporates them under personalized requirement. Multiple analysis factors are also combined to generate the search result. Author identified the accurate context map to generate the organized information to ensure the personalized ranked search modified under different vectors. Vijay Chauhan *et. al.* [3] integrated the machine learning approach to generate the page priority by observing the results of different search engines. A study work is provided to control the search method and improving the search results more accurately. Author used the connectivity driven mapping between the pages and the visit to generate the session driven results. Nandnee jain *at. al.*[4] used the response time as the major vector to improve the reliability of search results. Author tried to improve the efficiency of searched pages as well as tried to rectify the problem that occur because of overload. Jayendra singh Chouhan *et. al.* [5] combined different search and context factors in single algorithm. The clustering is applied to generate the data categorization and applied a walk theory to generate the ranking based on content mapping. Author also applied the computational resource analysis to apply the relevancy driven map and to generate the effective page ranking. The query criteria also filtered using history search so that more relevant results will be obtained.

Sweah Liang Yong *et. al.*[6] provided a work on graph net based mapping with different parameters including trust, hitmap and the content map. Author generated the flexible graph based ranking scheme to provide the connecting map analysis. The predictive observation is applied to generate the next query results. Jianmei Huang *et. al.* [7] provided a discussion on ranking algorithm and provided the categorization based ranking. Author applied the score driven mapping of web pages and generated the optimized map results. The search quality is improved for search engine by better cross mapping and the search ranking is improved upto an extent. Gyanendra Kumar *et. al.* [8] provided a new page ranking model based on the link visit method which was further devised under different search engines. Author used the concept of meta search engine to improve the search results. An inbound search analysis is provided along with intelligent observation on page visit to achieve the results of search space. Author provided the link visit based mapping to generate the effective web results to improve the page ranking. Tian Chong[9] used a classification tree based method to construct the relational observation on user visit and the requirement. The page ranking problem is drafted and provided a step by solution to achieve the reliable and efficient results. Author captured the quality information to generate effective search results. The dynamic crawling is provided with index mode specification to improve the search results. Sumita Gupta *et. al.*[10] provided a query based mapping to navigate the search results based on the content information observation. Author used the ranking method with comparative context information analysis. Author analyzed these algorithms to generate the results based on relevancy map. Author also improved the contextual performance based on the query map. The extensive growth of the filtered map is provided by the author generate the effective map results.

Different researchers also provided the improvement to search algorithm using the agent or the optimization inclusion. This kind of sample driven analysis is also provided by research to reduce the search data. Nagappan *et. al.*[11] provided work in same direction. Author used the multifactor based weighted algorithm to improve the content retrieval. Author applied agent based method integration to assign the larger ranking to the pages and improved the ranking results. Author derived the effective similarity information which was mapped with user contextual information to generate more accurate results. Divjot *et. al.* [12] also used the weighted ranked algorithm in relational aspect. Author used the approach to improve the HITS algorithm and provided more predictive unbiased scoring to web pages. The deep link analysis also improved the validity of work. Vedpal Singh *et. al.* [13] provided the analysis on internal mechanism of business organization and provided the myth and fallacies analysis on different ranking algorithms. Author improved the search method and provided the internal assessment to generate more accurate search results. Neelam Duhan *et. al.*[14] has provided the work on application driven analysis to provide control content mining, usage mining and structure mining. Author defined the ranking response observation with navigational behavior specification to generate more context driven results. Author improved the link search relevancy respective to user context. Different issues[15] associated to search system and the intelligent web mining was discussed by the author. Author applied the incorporated back propagation method to optimize the search results. Ashish Jain *et. al.*[16] identify the strengths and weaknesses of different algorithms including the topic based methods, weighted methods *etc.* Author defined a topology driven architecture to generate the relevant search results. Zhanzi Qiu *et. al.*[17] suggested

the preoperational filtration method for link type estimation and relevancy estimation to generate the filter rules. Author identified the technical issues to improve the search.

### III. RANKING METHODS

The search engine[3][4][5][6] information extracted from the web is in the form of relative URL that are proven based on the query similarity and user context map. The algorithmic specification and implementation is applied to generate more appropriate web links as results. But, as these query based pages are searched, they suffer from web search irregularities. In these irregularities some of the fake sites and spam pages stores the query information exactly and map to the user requirement but these pages are not appropriate to the user search. Because of this, there is requirement some parameter driven map to generate effective search results. These parameters are considered to improve the reliability and efficiency of search results. The ranking[6][7][8][9][10] to the search system is applied by different researchers in different ways. Different priorities and weights to different vector differentiate the significance of these parameters. The application and domain specification and mapping also support the important role while assigning the ranking to these pages. The ranking based web page mapping is required to achieve more accurate and significant results. The upper level analysis can be applied to gain more significant results. The ranking parameters include reliability which includes the page relevancy, Domain reliability, fake probability etc. The user context is another parameter that analyzes the user work behavior over the web for generating the ranking. The work behavior includes the number of hits to the particular page, page visit time, page forwarding, page bookmark etc. Efficiency is the vector that deals with the speed of generating results. The server load, accesses time, request process time are the parameters based on which the page ranking can be decided. These all parameters can be used in one or other way to generate the web page ranking. Some of the common methods for web page ranking are listed here under

#### A. *Name Entity Method*

This method is very much inspired from the concepts of natural language processing to deal with selective information or text. The domain specific and application specific search can be controlled by using these methods. The profile based search or the component based search are mostly directed by this method. The selective information processed by these methods includes the person name, location, firm name etc. These key points are considered as the information controlled to generate the mapping set so that more effective information map will be obtained. The process ratio can be applied in simple form to generate the contextual information. Different constraints and database map can be applied to identify the relative content vector. The rule based derivation provides more filtered mapping so that the results will be more appropriate to the application.

#### B. *Rule Based Approaches*

In these kind of approaches some rules or the constraints can be formulated to generate the different faces of effects under named entity identification and classification. The token based mapping can be applied on different speech words. The stemming, tagging, annotation assignment can be applied at the preprocessing stage to generate the categorized map. The database lookup is used to generate these rules. The boundary driven map with structural information derivation is applied to identify the mapped area and the object. The classification method can be incorporated at the earlier stage to map with specification of generated rules. In automated form, the training process can be applied to setup these rules on verified data. Later on, the real time data can be processed to generate the data class. The partial and full mapping can be applied to improve the accuracy of relevancy process.

#### C. *Machine Learning Methods*

Machine learning methods can be applied to generate the mapping results based on feature analysis. At the earlier stage, the features are acquired from the index table list as well as user requirement analysis. Later on some rule will be applied to identify the effective dataset features. Finally, the feature map will be applied on the dataset to generate the accurate map results. The training process with label specification can be applied to define the requirement data class and query class. The class can be identify based on the query type, domain requirement and the application specification. These all parameters can be applied individually and aggregative to generate more significant and relevant results. The learning method can be supervised learning, semi supervised learning or unsupervised learning based on the requirement specification.

Supervised learning is favorable content analysis based method applied on large annotated web pages. The common approach used to apply the specific rule based method such as Hidden Markov Model (HMM), Decision Tree, SVM (Support vector machine) etc. These parameters can be applied to generate the hidden associated information and provide the results generation with rule specification. Semi supervised learning used

a corpus to prepare the annotated set. The recognition processed is weakly supervised and the bootstrapping process is applied at the earlier stage to generate more accurate results. The sentence level and contextual information can be combined to generate the appropriate results. Unsupervised learning can be applied as the clustering method to use the available dataset and its associated statistics to divide in the category form. These categories are formed to identify the most mapped results.

#### D. Duplicate Page Detection

This scheme is used to filter the web results from the duplicate link generation. The intelligent observation on page content is required to identify the duplicate page generation. The individual information analysis with authorized information analysis is required to map this information and to generate more accurate and the unique information. The equivalence information processing can be applied and considered to generate the reason results so that the index table will be filtered. The source driven map can be applied to identify the valid web results. The method is appropriate avoid to ranking on spam pages and sites.

#### IV. CONCLUSIONS

In this paper, a study to the web search engine and associated search method is provided. The paper has discussed the scope of web search method and its specification in public and private domain. The basic model is provided with specification of integrated components. The role of web crawler and the ranking parameters is also discussed. The main work is defined to explore the scope of different ranking algorithms. These automated algorithms are here described with feature vector specification.

#### REFERENCES

- [1] A. Gupta, A. Dixit and P. Devi, "A novel user preference and feedback based Page Ranking technique," Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on, New Delhi, 2015, pp. 1335-1340.
- [2] M. P. Selvan, A. C. Shekar, D. R. Babu and A. K. Teja, "Efficient ranking based on web page importance and personalized search," Communications and Signal Processing (ICCSP), 2015 International Conference on, Melmaruvathur, 2015, pp. 1093-1097.
- [3] V. Chauhan, A. Jaiswal and J. Khan, "Web Page Ranking Using Machine Learning Approach," Advanced Computing & Communication Technologies (ACCT), 2015 Fifth International Conference on, Haryana, 2015, pp. 575-580.
- [4] N. Jain and U. Dwivedi, "Ranking web pages based on user interaction time," Computer Engineering and Applications (ICACEA), 2015 International Conference on Advances in, Ghaziabad, 2015, pp. 35-41.
- [5] J. Singh Chouhan and A. Gadwal, "Improving web search user query relevance using content based page rank," Computer, Communication and Control (IC4), 2015 International Conference on, Indore, 2015, pp. 1-5.
- [6] S. L. Yong, M. Hagenbuchner and A. C. Tsoi, "Ranking Web Pages Using Machine Learning Approaches," Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT '08. IEEE/WIC/ACM International Conference on, Sydney, NSW, 2008, pp. 677-680.
- [7] Jianmei Huang, Guoren Wang and Zhiqiong Wang, "Cross-subject page ranking based on text categorization," Information and Automation, 2008. ICIA 2008. International Conference on, Changsha, 2008, pp. 363-368.
- [8] G. Kumar, N. Duhan and A. K. Sharma, "Page ranking based on number of visits of links of Web page," Computer and Communication Technology (ICCCT), 2011 2nd International Conference on, Allahabad, 2011, pp. 11-14.
- [9] C. Tian, "A kind of algorithm for page ranking based on classified tree in search engine," Computer Application and System Modeling (ICCASM), 2010 International Conference on, Taiyuan, 2010, pp. V13-538-V13-541.
- [10] S. Gupta, N. Duhan, P. Bansal and J. Sidhu, "Page ranking algorithms in online digital libraries: A survey," Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), 2014 3rd International Conference on, Noida, 2014, pp. 1-6.
- [11] Nagappan V. K and P. Elango, "Agent based weighted page ranking algorithm for Web content information retrieval," Computing and Communications Technologies (ICCCT), 2015 International Conference on, Chennai, 2015, pp. 31-36.
- [12] Divjot and J. Singh, "Effective Model and Implementation of Dynamic Ranking in Web Pages," Communication Systems and Network Technologies (CSNT), 2015 Fifth International Conference on, Gwalior, 2015, pp. 1010-1014.

- [13] V. Singh, A. Chaudhary and P. Punia, "OSA - PR: Optimized searching algorithm based on page ranking: Proposed algorithm," Engineering Education: Innovative Practices and Future Trends (AICERA), 2012 IEEE International Conference on, Kottayam, 2012, pp. 1-4.
- [14] N. Duhan, A. K. Sharma and K. K. Bhatia, "Page Ranking Algorithms: A Survey," Advance Computing Conference, 2009. IACC 2009. IEEE International, Patiala, 2009, pp. 1530-1537.
- [15] D. Malhotra, "Intelligent web mining to ameliorate Web Page Rank using Back- Propagation neural network," Confluence The Next Generation Information Technology Summit (Confluence), 2014 5th International Conference -, Noida, 2014, pp. 77-81.
- [16] 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, Gwalior, 2013, pp. 640-645.
- [17] Zhanzi Qiu, M. Hemmje and E. J. Neuhold, "Using link types in Web page ranking and filtering," Web Information Systems Engineering, 2001. Proceedings of the Second International Conference on, 2001, pp. 311-320 vol.1.