An efficient algorithm for web search optimization

Dr.Ashit Kumar Dutta

Associate Professor

Department of Computer Science

Shaqra University

Abstract

Information overload is the common problem in the web information retrieval system. Web search optimization is the active area of research because of size of the web is increasing infinitely. Users' are not satisfied with the available search engine. The notion of the paper is to address an efficient algorithm for the better web search optimization which leads to user satisfaction. The research is based on link analysis technique and uses web cache and log of the users system.

Keywords: Web search, Link analysis, log, page rank, local rank

I.Introduction

Searching becomes more common for the people to navigate the related information. Yahoo, google and altavista are the major search engines for the internet users. Search engines index almost all websites exist in the internet. There are millions of websites are present in world wide web and they are indexed in unique way by the search engine. Web pages are ranked by the search engines using HITS and page rank algorithm.[1][2] If a user is trying particular information through a keyword by using a search engine then it will be searched through hyperlinks and related pages are displayed and organized by page rank / HITS algorithm. In the past, the keywords were searched by content used in the webpage; later found that it was not a correct way because bogus content created to escalate the page by the site developers.

Link analysis techniques are widely used in finding fraudulent activities occurred in web, statistical analysis and web search optimization.[3][4] Knowledge discovery is the process of acquiring meaningful information from the huge pool of data in web. Users tend to use search engine to get their related information but information overload will be a problem for the user to select the exact link contains related information.

Search engine uses page rank and HITS algorithm for the ranking mechanisms of websites to help the user to select the appropriate sites. Users are not well versed in querying for their needed information. Users used to get dilemma in selecting websites because of large numbers of sites returned by the search engines. Search engine keeps record of user activities and return the sites with the use of recorded information. Some search engine uses cookies to monitor the user activity and deliver the related information according to their mind set.[5][6] The keyword "cookies" related to food as well as files stored in the system and it is depend upon the user uses

it. If the search engine recognize the past usage of the user and it automatically escalates the rank o the keyword related site for the user for the easier navigation.[7][8]

The notion of the paper is to optimize the search results according to the user. The following part of the paper will discuss the literature and algorithm framed for the optimization.

II. Review of the literature

Link analysis is the method used for marketing in web and web search optimization. Web mining is the trend setter in the world of internet. Link analysis is one of the technologies of web mining. Many researches are revolving around web search optimization. The following are the recent works executed for search optimization.

A.K.Sharma et.al., have proposed a method performs query clustering for the sequential pattern of previously visited web pages by using sequential pattern algorithm. The generated pattern in applied to re – rank the pages given by page rank algorithm.

Ping – Tsai et. al., have designed a web server by applying search engine optimization technique for the purpose of web intelligence. They have analyzed the web services for customers' data and demographics. They have deployed the top k- keywords for the search engine optimization.

Iman Rasekh has proposed a research on dynamic search optimization by using imperialistic competitive algorithm. The research discuss about the graph structured search for the semantic web search.

Rajaram et.al., proposed a research using web cache and eviction method for semantic web search engine. The research discussed about current algorithms for web cache optimization. The authors have used least grade replacement method for their work.

Andrew y.Ng et.al., have proposed stable algorithms for link analysis. They have analyzed the existing link analysis algorithms HITS and page rank. They have extended the analysis to give insight into ways of designing stable link analysis methods.

III. Results and Discussion

Search engine optimization should be carried out in a proper way for the optimum result. Link analysis is the channel to carry out the optimization.[9][10] The past research shows that there were numerous techniques exist for the optimization but ends in vain.[11][12] Link analysis shows good results for optimization. The proposed research uses the cookies, web cache and browser cache to collect the past activity of the user. Cookies are normally deposited by search engine to capture the user navigational pattern. The research gives priority to a particular site if the user visits more than 3 times and according to that the local rank will be set for the particular site. Page rank algorithm is used to generate the general page rank for the web page. The local

rank and page rank are compared and if local rank is lesser than or equal to page rank then actual rank will be the local rank otherwise it will be equal to page rank.

Algorithm

Input: Cookies data, web cache data and page rank
Output: Updated rank of pages

1. wc=0, c=1,lr=1;
2. l = requestuser(p);
3. wc++, c++;
4. If new_requestgoto 2
5. if(wc> 0 && c > 0)
6. find(webcache)
7. if (B(l) > 3) // Browser cache data
8. lr++;
9. else
10. lr = 0;
11. for (each page == l)
12. pr=pagerank(l)
13. if (lr< = pr) // local rank and page rank
14. ar = pr(l) // actual rank
15. else
16. ar = lr(l);
17. end

The following table shows the result generated by the proposed algorithm. The LR, PR and AR are local rank, page rank and actual rank.

Results:
Keyword – Database

S.No.	Webpage	LR	PR	AR
1	en.wikipedia.org	3	1	1
2	www.webopedia.com	2	2	2
3	www.tutorialspoint.com	4	5	4
4	www.mysql.com	1	7	1
5	www.salesforce.com	7	3	3
6	Database.oxfordjournals.org	5	4	4
7	www.databasejournal.com	6	6	6

The result describes the working style of the algorithm. The keyword "Database" listed out the above 7 web pages and LR, PR are generated by the algorithm. The comparison of LR and PR has generated the AR and according to that the pages will be re arranged for the user. The user

requested page will be displayed in a manner that the user can easily navigate for the required information.

Conclusion

Search engine optimization plays a vital role in web search, web marketing and web traffic. Link analysis is the method used for the optimization. Earlier researches have shown good result and the proposed research is a step towards customization of the search result for the users. Cookies and web cache are utilized to find the users' navigation details. The results of the research show that the pages are well optimized and users are satisfied with the performance.

References

[1]. A.K.Sharma, Neha Aggarwal, NeelamDuhan and Ranjna Gupta, Web search result optimization by mining the search engine query logs, IEEE, 978 – 1- 4244 – 9703 – 4/10.

[2]. Ping – Tsai Chung, Sarah H.Chung and Chun – Keung Hui, A web server design using search engine optimization techniques for web intelligence for small organizations, IEEE LISAT 2012, Long Island Systems, Applications and Technology conference, May 2012.

[3]. M. Rajaram and Lathashanmugavadivu, web caching in semantic web based multiple search engines, IEEE, 978 – 1- 4244 – 5967 – 4 / 10.

[4]. Iman Rasekh, Dynamic search optimization for semantic webs using imperialistic competitive algorithms, IEEE 978 – 1- 4673 – 1401 – 5 / 12.

[5]. S.Chakrabarti, Mining the web: Statistical analysis of hypertext and semi structured data, Morgan Kaufmann, 2002.

[6].Z. Markov and D.T. Larose, Data mining the web: uncovering patterns in web content, structure and usage. 2007.

[7]. K.Bharat and M.Henzinger, Improved algorithms for topic distillation in a hyperlinked environment. In proc. 21st annual intl. ACM SIGIR conf. pages 104 -111, ACM 1998.

[8]. D.Cohn and H.Chang, Probabilistically identifying authoritative documents. In proc. 17th international conference on machine learning, 2000.

[9]. G.H. Golub and C.F. Van Loan, Matrix computations. Johns Hopkins Univ. press 1996.

[10]. S.Brin and L.Page, The anatomy of a large – scale hypertextual (web) search engine. In the seventh international world wide web conference 1998.

[11]. Andrew Y.Ng, Alice X.Zheng and Michael I.Jordan, Stable algorithms for link analysis, SIGIR 01, Sep 9 – 12, 2001, New Orleans, Lousiana, USA.

[12]. Alex zolotovitski and Richard castro, Usage number of pages found by search engines for business analytics, International conference on internet computing (ICOMP '09), las vegas, 2009.