

Query Log Search Engine Optimization

Shital Satpute

B.E Student of Computer Engineering
Atharva College of Engineering,
Mumbai, MH, India

Ashwini Shinde

B.E Student of Computer Engineering
Atharva College of Engineering,
Mumbai, MH, India

Sonali Tambe

B.E Student of Computer Engineering
Atharva College of Engineering,
Mumbai, MH, India

Santosh Dodamani

Assistant Professor,
Atharva College of Engineering,
Mumbai, India

Abstract--- With the ever increasing enormity of information over the internet, finding relevant information to a particular topic has become difficult. Search history analysis is the detailed examination of web data from different users for the purpose of understanding and optimizing web handling. In this paper, we are trying to analyze user's web history and classify it into groups through optimization algorithms. The user's search history contains browsing history and submitted queries. In this approach, we are trying to classify search history into similar query groups. The aim of this approach is to classify query into groups automatically and dynamically. The proposed system combines word similarity with document similarity for purpose of website ranking, query suggestions.

Keywords--- Query, SEO, History, Ranking

I. INTRODUCTION

World Wide Web is a huge source of information that constitutes almost all results to the submitted queries.

In this wide ocean of information, obtaining relevant result is what makes Search Engine Optimization (SEO) an indispensable part in the eyes of researchers. Hence query log mining plays a vital role in classifying user's web history. It not only helps in classifying queries but can also be combined with data mining tools to predict and suggest related results. Our log queries contains previously submitted queries, URL's and similar clicked documents.

Query log mining is also one of the branch of web analytics. Therefore an efficient search engine is the one which returns quality results ,browsing speed etc. The main job of log analyst is to find the clicked documents, submitted queries and corresponding URL's. This will help the user to predict future queries. Therefore the log query keeps a track on user's browsing history and suggests similar clicked results. This track of information enhances interaction between user and search engine. Thus SEO techniques can be considered as an effort towards improving search efficiency. Some of the examples are query spelling correction, prediction mining, query suggestion etc. Many times it happens that users are not satisfied with the information that is fetched from most search engines. In our project we are trying to make search pattern more efficient by classifying queries into different clusters. The aim of this research is to determine the suitable keywords for website visibility used on search engines by people. It happens that user clicks on a

result but does not get related information. This happens due to the insertion page breaks where SEO's try to deviate the information. A method to analyze user search history and perform user query classification in an automated and dynamic fashion. We consider a query group as a collection of queries together with the corresponding set of clicked URLs around a general information search. Each group will be dynamically updated when the user issues new queries and new query groups will be created over time. A query group can be defined as a collection of queries together with the corresponding set of user visited sites. There are different algorithms to cluster queries .Some of them includes cluster rank algorithm, Generic sequential pattern, Weighted cluster rank etc. Here the term 'data mining' is often treated as synonym for another term KDD (knowledge discovery from data) which highlights the goal of mining process. The approach with pre-mining of user logs to form cluster of queries. Out of all the words present in query, only keywords are chosen to classify into groups.

The related work is present in section II, the proposed work is present in section III along with comparison of existing system.

II. LITERATURE SURVEY

In this paper, the words cluster and group mean the same. There are many clustering algorithms, some of them are

A. Clustering Algorithms

1. Graph based query clustering
2. Concept based query clustering
3. Personalized based query clustering

The necessary information in clustering is extracted by clickthrough data. For clustering of queries, the semantics of submitted queries and clicked URL is compared. This method assumes that user clicks on highly relevant result. But it fails if user clicks on returned query results. Consider the example, "Malls in Mumbai". Here the content concept is "malls" and location concept is "Mumbai". Here it forms a bipartite graph where all concepts are at vertex. If the user clicks on result, then the concepts are merged with clicked URL. The proposed system works in two steps:

1. All the concepts from search results are extracted.
2. These concepts are used to identify related queries for that query.

Apart from these query classification algorithms, researchers have found ranking based optimization techniques which includes Google’s Page rank and Weighted page rank algorithms.

Criteria	Page Rank	Weighted page rank
Basic criteria	Graph based ranking algorithm.	Based on calculation of weight of page.
Techniques used	Web structure mining.	Web structure mining.
Complexity	O(Log(N))	< O(Log(N))
Advantage	Rank based importance of page.	Rank based importance of page
Disadvantage	Prefers old page over new page links..	Prefers popularity of websites.

Figure 1. Comparison between page rank and weighted page rank

B. Dynamic grouping algorithm

Inputs to dynamic query grouping algorithm are current singleton query group and the corresponding set of clicks, set of existing query groups, and the similarity threshold. Output of the dynamic grouping algorithm is a query group that best matches the current singleton query group or a new query group. In our approach, at first, we form a singleton query group by placing the current query.

III. PROPOSED SYSTEM

The proposed system aims at classifying user search history into clusters in an automated and dynamic fashion. The automatically classified query groups will help in different search engine optimization techniques like query suggestion, search result re-ranking, query alterations. It is query and content independent algorithm that assigns a value to every document independent of the query [3]. It is concerned with the static quality of web page. It computes page rank using web graph. This algorithm cluster forming techniques to group queries of similar types and a new result is entered in web browsing history.

The following table demonstrates the clustering of queries

Table I User Query Session

Number	Query text
1	Gmail sign-in
2	Email services
3	Linkedin login
4	Linkedin profile
5	Word to Pdf
6	Pdf to word
7	.doc to .pdf
8	Gmail account
9	Converters
10	Adobe reader

1	Gmail sign-in
2	Email services
3	Gmail account
1	Word to pdf
2	Pdf to word
3	.doc to .pdf
4	Converters
1	Linkedin
2	Profile

Figure 2. User query session

GSP algorithm :-

This algorithm for deciding the best matching query group is given below[1]:

1. Select Best Group g.
2. Input: Current query and URL.
T indicates threshold value.
3. The current query and the set of clicks as a singleton query group, gc.
4. The set of already formed query groups,
G = { g1, g2,...,gn}
Similarity threshold value, Tsim.
g = φ is an empty set of cluster..In the next step,the search query is matched with URL.
5. Then we make use of a while loop for continuing the iterations.

The word similarity of GSP algorithm is calculated as shown below:

$$q1=\{\text{Maruti,swift,dzire}\}$$

$$q2=\{\text{Maruti,swift,price}\}$$

$$\text{sim}(q1,q2)=\frac{1+1}{\sqrt{1+1+1}.\sqrt{1+1+1}}$$

$$= 0.66$$

If $\text{sim}(q1,q2)>0.53$,place queries in same group

Result of proposed system:

Total number of URLs	Page Rank time (s)	Cluster Rank time (s)	Difference
9325	386.5	340.2	12%
22371	840.4	916.6	-9%
26864	1305.9	1130.5	13%
33928	1968.1	1564.5	20%
41148	2195.8	1722.6	21%
69439	4455.6	3251.1	27%
105674	15914.5	10477.6	34%
290561	31246.1	26562.54	15%

Figure 3. Query Search time chart

IV. CONCLUSION

In this paper, we have tried to optimize the search results using cluster rank and GSP algorithm to give better results and also boost the implementation. We are trying to make efficient search engine in order to obtain relevant information. Search engines can be combined with data mining tools to predict alternative search results so that user can also get additional results in high ranking. Automatic and dynamic grouping is the essence of the proposed system. After classification of queries, it can be grouped into query suggestion, query alteration and other optimization techniques as the future work.

V. REFERENCES

- [1] Archana Kurian, "Analyzing and classifying user search history for web search engine optimization", 3rd international Conference on Eco-friendly Computing and Communication Systems, 2014 IEEE.
- [2] Fawaz AL Zaghoul, Osama Rabahah and Hussam Fakhouri, "Website search engine optimization: Geographical and Cultural point of view", 2014 UKSim-AMSS 16th International Conference on Computer modelling and simulation.
- [3] Tsung-Fu Lin, Yan-Ping Chi, " Application of webpage optimization for clustering system on search engine-Google study" ,2014 IEEE International Symposium on Computer, Consumer and Control.
- [4] Ashish Kumar Kushwaha, Prof. Nitin Chopde, " A comparative study of algorithm in SEO & approach for optimizing the search engine results using Hybrid of query recommendation and document clustering, Genetic algorithm" .
- [5] Sowmya Ravi, Neeraja Ganesan, "Search engines using Evolutionary algorithms" , International Journal of Communication Network Security ISSN: 2231 – 1882, Volume-1, Issue-4, 2012.
- [6] Bhupesh Gupta, Sandip Kumar Goyal, "A Review on clustering algorithm for search engine optimization", International Journal of Advanced Research in Computer Science and Software Engineering.
- [7] Venkat N. Gudivada and Dhana Rao, East Carolina University, Jordan Paris, CBS Interactive, Published by IEEE Computer Society, Oct' 15.
- [8] JOHN B. KILLORAN, "How to Use Search Engine Optimization Techniques to Increase Website Visibility", IEEE Transactions on Professional communication, Vol.56, NO.1 MARCH 2013.