Review Article Impact Factor: 0.621 ISSN: 2319-507X
Ashwini Gaurkhede, IJPRET, 2014; Volume 2 (9): 351-356 IJPRET

INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

A REVIEW ON QUERY BASED HISTORY ORGANIZATION

ASHWINI R. GAURKHEDE[1], PROF. D. R. DHOTRE[2],

1. M. E. Student, Dept of CSE, SSGMCE, Shegaon, Dist: Buldhana, M. H. , India.
2. Professor, Dept of CSE, SSGMCE, Shegaon, Dist: Buldhana, M.H., India.

Accepted Date: 27/02/2014 ; Published Date: 01/05/2014

Abstract: The primary means of accessing information online is still through keyword queries to a search engine. The complex task oriented goals are continuously pursuing on the web by users. To find new information, People often repeat Web searches on topics they have previously explored and to re-find information they have seen in the past. The repeated search query may differ from the initial query but can lead to clicks on the same results. In addition to viewing their search history, users can manipulate it by manually editing and organizing related queries and clicks into groups. The query group corresponds to the atomic information. While searching, the search engine can keep their old queries and clicks. Grouping of related queries in the search history is useful for a variety of search engine applications. Query grouping allows the search engine to better understand a user's session and tailor that user's search experience according to their needs. Hence this system presents a mechanism that automatically identifies query groups in the search history. These features are helpful and the manual efforts involved can be disruptive and will be untenable as the search history gets longer over time. In this paper, our goal is to automatically and dynamically organize a user's search history into query groups containing one or more related queries and their corresponding clicks.

Keywords: User History, Search History, Query Clustering, Search Engine, Click Graph.

Corresponding Author: MR. ASHWINI R. GAURKHEDE

Access Online On:

www.ijpret.com

How to Cite This Article:

Ashwini Gaurkhede, IJPRET, 2014; Volume 2 (9): 351-356

PAPER-QR CODE

INTRODUCTION

How the search results are organized and presented, it directly affects search engine utility. Ideally, the organization should make it easier for users to fulfill their information needs. The variety and the complexity of tasks that users try to accomplish online are growing day by day like the richness of information on the web. Hence, the behavior of the user is unpredictable and untraceable as in a user may perform many different search terms over small period of time or may perform many similar searches at different times. Users are issuing informational and transactional queries to the search engine. This is because users now pursue much broader informational and task-oriented goals such as arranging for future travel, managing their finances, or planning their purchase decisions. A complex task has to be broken down into a number of co-dependent steps over a period of time. To access the information over the internet is through the keywords and queries using a search engine. For viewing their search history, users can manipulate it by manually editing and organizing related queries and clicks into groups. These features are helpful but the manual efforts involved can be disruptive and will be untenable as the search history gets longer over time. To overcome the above problem and to enable services and features that can help users during their complex search quests online is the capability to identify and group related queries together. With the help of user's previous activities of document relevance we can predict his/her reaction to the current retrieved documents. For example, if the user searched with the same query "python" before and clicked on Python language website's link, we have high confidence that the user would do it again this time, and it makes good sense to list that webpage in the top. Even when there is no exact occurrence of the current query in history, user may still find similar queries like "python doc" helpful i.e. discovering that the user prefers results from the www.python.org site.

The most important, query grouping allows the search engine to better understand the user's session and potentially tailor that user's search experience according to their needs. This will help to improve the quality of the key components of search engines such as result ranking, query suggestions, query alterations, sessionization, and collaborative search. People often repeat Web searches, both to find new information on topics they have previously explored and to re-find information they have seen in the past. The query associated with a repeat search may differ from the initial query but can nonetheless lead to clicks on the same results. Thanks to the ubiquity of the Internet search engine box, users have come to depend on Web search engines both to find new information and to re-find previously viewed information [1].

GOAL

Our main goal is to organize the user search histories into query groups, each containing one or more related queries and their corresponding clicks. The main objective is to analyze the query log generated by the user and then use them for further operations like, generating query group, extracting semantics relations from query log, clustering them, etc.

Query logs

Query logs records the queries and the actions of the users of search engines, and as such they contain valuable information about the interests, the preferences, and the behavior of the users, as well as their implicit feedback to search engine results.

Query clustering

Query clustering [6] is a process used to discover frequently asked questions or most popular topics on a search engine. Because of the short lengths of queries, approaches based on keywords are not suitable for query clustering. The similarity between two queries may be deduced from the common documents the users selected for them. Despite the fact that keywords are not always good descriptors of contents, most existing search engines still rely solely on the keywords contained in documents and queries to calculate their similarity. In many cases, the answers returned by search engines are not relevant to the user's information need, although they do contain the same keywords as the query. The query clustering approach is based on two principles that are (1) If users clicked on the same documents for different queries, then these queries are similar. (2) If a set of documents is often selected for the same queries, then the terms in these documents are; to some extent, related to the terms of the queries.

Query group and Dynamic query group

Each query group is a collection of queries by the same user that are relevant to each other around a common information need. These query groups are dynamically updated as the user issues new queries, and new query groups may be created over time. A query group is an ordered list of queries, q_i, together with the corresponding set of clicked URLs, clk_i of q_i. A query group is denoted as $s = (\{q_1; clk_1\}; \ldots; \{q_k; clk_k\})$.

The specific formulation of our problem is as follows:

Given: a set of existing query groups of a user, $S = \{s_1, s_2, \ldots, s_n\}$, and their current query and clicks, $\{q_c, clk_c\}$,

Find: the query group for $\{q_c, clk_c\}$, which is either one of the existing query groups in S that is most related to, or a new query group $s_c = \{q_c, clk_c\}$ if there does not exist a query group in S that is not sufficiently related to $\{q_c, clk_c\}$.

Algorithm to select the query group which is similar to the given query and clicked URLs.

SelectBestQueryGroup

Input:

1) The current singleton query group sc containing the current query q_c and set of clicks clk_c

2) A set of existing query groups $S=\{s_1,..,s_m\}$

3) A similarity threshold T_{sim}, $0 <= T_{sim} <= 1$

Output: The query group s that best matches sc, or a new one if necessary

(0) $s = \phi$

(1) Tmax = Tsim

(2) for i = 1 to m

(3) if $sim(s_c, s_i) > Tmax$

(4) s = si

(5) Tmax = sim(sc, si)

(6) if $s = \phi$

(7) $S = S \cup s_c$

(8) $s = s_c$

(9) return s

RELATED WORK

A. Broder in 2002 came with "A taxonomy of web search" [2] The main aim of the research is to point out difference between informational, navigational and transactional queries and introduce and analyze a taxonomy of web searches. An understanding of this taxonomy is essential to the development of successful web search. Current search engines deal well with informational and navigational queries.

J. Teevan, E. Adar, R. Jones, and M. A. S. Potts in 2007 published "Information reretrieval: repeat queries in yahoo's logs," [1]. The query associated with a repeat search may differ from

the initial query but can nonetheless lead to clicks on the same results. The observation of some users includes that repeat searches and repeat clicks were very common and it was possible to predict which queries were navigational and what results were likely to be clicked.

"Multitasking during Web search sessions" by A. Spink, M. Park, B. J. Jansen, and J. Pedersen, in 2006 [3] learned in this research, during web search session, multitasking is done. Multitasking is the ability of humans to simultaneously handle the demands of multiple tasks through task switching. There are a broad variety of topics in multitasking search sessions, and three or more query sessions sometimes contained frequent topic changes. Multitasking is found to be a growing element in Web searching..

Jones and Klinkner [4] and Boldi et al. [5] investigated the search-task identification problem. More specifically, Jones and Klinkner [4] considered a search session to consist of a number of tasks, and each task further consists of a number of sub-tasks. They trained a binary classifier with features based on time, text, and query logs to determine whether two queries belong to the same task.

R. Baeza-Yates and A. Tiberi proposed the technique of "Extracting semantic relations from query logs" [7]. The graph mining techniques are proposed in this research, applied to larger query logs. The problem of online query grouping is also related to query clustering [6], [7]. It is found that the query clusters to be used as possible questions for a FAQ feature relying on both text and click features.

CONCLUSIONS

In this paper, we show how the information can be used effectively for the task of organizing user search histories into query groups. We described the process of generation of query log by a user using a web search engine accessing any information over period of time. This query log can be grouped. We further described the process of query clustering. Here we studied the basic concepts about organizing a user search histories for better performance.

REFERENCES

1. J. Teevan, E. Adar, R. Jones, and M. A. S. Potts, "Information reretrieval: repeat queries in yahoo's logs," in *SIGIR*. New York, NY, USA: ACM, 2007, pp. 151–158.

2. A. Broder, "A taxonomy of web search," SIGIR Forum, vol. 36, no. 2, pp. 3–10, 2002.

3. A. Spink, M. Park, B. J. Jansen, and J. Pedersen, "Multitasking during Web search sessions," Information Processing and Management, vol. 42, no. 1, pp. 264–275, 2006.

4. R. Jones and K. L. Klinkner, "Beyond the session timeout: Automatic hierarchical segmentation of search topics in query logs" in *CIKM*, 2008.

5. P. Boldi, F. Bonchi, C. Castillo, D. Donato, A. Gionis, and S. Vigna, "The query-flow graph: Model and applications" in *CIKM*, 2008.

6. D. Beeferman and A. Berger, "Agglomerative clustering of a search engine query log" in *KDD*, 2000.

7. R. Baeza-Yates and A. Tiberi, "Extracting semantic relations from query logs" in *KDD*, 2007.

8. J. Han and M. Kamber, Data Mining: Concepts and Techniques. Morgan Kaufmann, 2000.

9. W. Barbakh and C. Fyfe, "Online clustering algorithms", International Journal of Neural Systems, vol. 18, no. 3, pp. 185–194, 2008.

10. J.-R. Wen, J.-Y. Nie, and H.-J. Zhang, "Query clustering using user logs," ACM Transactions in Information Systems, vol. 20, no. 1, pp.59–81, 2002.