International Journal of Advanced Engineering & Innovative Technology (IJAEIT)

Volume 1, Issue 5, December -2014

ISRA IMPACT FACTOR: 0.42

Personalized Search Based On User Search Histories

Pawar Ganesh, Mulani Mohasin, Dagade Anil

Abstract:- Users are increasingly pursuing complex task-oriented goals on the Web, such as making travel arrangements, managing finances or planning purchases. To this end, they usually break down the tasks into a few co-dependent steps and issue multiple queries around these steps repeatedly over long periods of time. To better support users in their long-term information quests on the Web, search engines keep track of their queries and clicks while searching online. In this paper, we study the problem of organizing a user's historical queries into groups in a dynamic and automated fashion. Automatically identifying query groups is helpful for a number of different search engine components and applications, such as query suggestions, result ranking, query alteration and collaborative search. We go beyond approaches that rely on textual similarity or time thresholds, and we propose a more robust approach that leverages search query logs

Index Term: - User History, Search History, Query, Clustering, Query Reformulation, Task Identification

INTRODUCTION:-

AS the size and richness of information on the Web grows, so does the variety and the complexity of tasks that users try to accomplish online. Users are no longer content with issuing simple navigational queries. Various studies on query logs (e.g., Yahoo's [1] and AltaVista's [2]) reveal that only about 20% of queries are navigational. The rest are informational or transac- tional in nature. 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. However, the primary means of accessing information online is still through keyword queries to a search engine. A complex task such as travel arrangement has to be broken down into a number of co-dependent steps over a period of time. For instance, a user may first search on possible destinations, timeline, events, etc. After deciding when and where.

RELATED WORK:-

While we are not aware of any previous work that has the same objective of organizing user history into query groups, there has been prior work in determining whether two queries belong to the same search task. In recent work, Jones and Klinkner [4] and Boldi et al. [5] investigate the search-task identification problem. More specifically, Jones and Klinkner [4] considered a search session to consist of a number of tasks (missions), and each task further consists of a number of sub-tasks (goals). They trained a binary classifier with features based on time, text, and query logs to determine whether two queries belong to the same task. First, the query-log based features in [4], [5] are extracted from co-occurrence statistics of query pairs. In our work, we additionally consider query pairs having common clicked URLs and we exploit both co- occurrence and click information through a combined query fusion graph. [4] will not be able to break ties when an incoming query is considered relevant to two existing query groups. Additionally, our approach does not involve learning and thus does not require manual labeling and re-training as more search data come Finally, our goal is to provide users with useful query groups on- the-fly while respecting existing query groups. On the other hand, search task identification is mostly done at server side with goals such as personalization, query suggestions [5] etc. as users may be multitasking when searching online [3], thus resulting in interleaved query groups. The notion of using text similarity to identify related queries has been proposed in prior work. He et al. [24] and Ozmutlu and C¸avdur [28] used the overlap of terms of two queries to detect changes in the topics of the searches. Lau and Horvitz [29] studied the different refinement classes based on the keywords in queries, and attempted to predict these classes using a Bayesian classifier. Unlike online query grouping, the

Pawar Ganesh, Mulani Mohasin, Dagade Anil
Research Scholar at Department of Computer Engineering ,
Navsahyadri Education Society's Group of Institutions,
University of Pune, Naigaon, Pune 412213 India

queries to be clustered are provided in advance, and might come from many different users. The query clustering process is also a batch process that can be accomplished offline. While these prior work make use of click graphs, our approach is much richer in that we use the click graph in combination with the reformulation graph, and we also consider indirect relationships between queries connected beyond one hop in the click graph. Graphs based on query and click logs [3 5] have also been used in previous work for different applications such as query suggestions [5] query expansion [36], ranking [37] and keyword generation [14] we take advantage of the stationary probabilities computed from the graph as a descriptive vector (image) for each query in order to determine similarity among query

PROPOSED METHODOLOGY:-

Collecting and Storing User Search Histories, For Search Engine And Organize It In Database. Optimizing Query Using Search Histories Providing Password Security Database Maintenance To Better Support Users In Their Long-term Information Quests On The Web.Search Engines Keep Track Of Their Queries And Clicks While Searching Online. Automatically Identifying Query Groups Is Helpful For A Number Of Different Search Engine Components

SYSTEM ARCHITECTURE:-

To ensure that each query group contains closely related and relevant queries and clicks, it is important to have a suitable relevance measure sim between the current query singleton group sc and an existing query group si ∈ S.

There are a number of possible approaches to determine the relevance between sc and si. Below, we outline a number of different relevance metrics that we will later use as baselines in experiments (see Section 5). We will also discuss the pros and cons of such metrics as well as our proposed approach of using search logs (see Section 3). Time. One may assume that sc and si are somehow relevant if the queries appear close to each other in time in the user's history. In other words, we assume that users generally issue very similar queries and clicks within a short period of time.

PAGE RANK ALGORITHM:-

Steps:

1. Make a new output file, R.
2. Read L and I in parallel (since they're all sorted by URL).
3. For each unique source URL, determine whether it has any outgoing links:
4. If not, add its current PageRank value to the sum: T (terminals).
5. If it does have outgoing links, write (source-url, dest-url, Ip/|Q|), where Ip is the current PageRank value, |Q| is the number of outgoing links, and dest_url is a link destination.
6. Sort R by destination URL.
7. Scan R and I at the same time.The new value of Rp is generated
8. Check for convergence
9. Write new Rp values to a new I file.

MATHEMATICAL MODULE:

1. U is a main set of users like u1,u2,u3….
U={u1,u2,u3…..} 2 .
u1={s1,s2,s3….}

u1 is a set of existing user query group

3. s1={q1,clk1}when

s1 is a set of current query & clicks

q1 is a single query

Clk1 is a single click for that query

CONCLUSION:

How such information can be used effectively for the task of organizing user search histories into query groups. This intend to investigate the usefulness of knowledge gained from these query groups in various applications such as providing query suggestions and ranking of search results.

REFERENCES:

1. J. Teevan, E. Adar, R. Jones, And M. A. S.Potts, Information Retrieval repeat queries in yahoo's logs," in SIGIR. New York, NY, USA: ACM, 2007, pp. 151–158.
2. E. Sadikov, J. Madhavan, L. Wang, and A. Halevy, "Clustering query refinements by user intent," in WWW, 2010.
3. A. Spink, M. Park, B. J. Jansen, and J. Pedersen, "Multitasking during Web search sessions," Information Processing and Manage- ment, 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. J. Han and M. Kamber, Data Mining: Concepts and Techniques. Morgan Kaufmann, 2000.
6. A. Broder, “A Taxonomy of Web Search,”SIGIR Forum,vol. 36,no. 2, pp. 3- 10, 2002.
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,” Interna- tional Journal of Neural Systems, vol. 18, no. 3, pp. 185–194, 2008
10. M. Berry and M. Browne, Eds., Lecture Notes in Data Mining. World Scientific Publishing Company, 2006.
11. V. I. Levenshtein, “Binary codes capable of correcting deletions, insertions and reversals,” Soviet Physics Doklady, vol. 10, p. 707, 1966.