

A Survey on Query Suggestion Techniques and Evaluation of Location Aware Keyword Query Suggestion

Dipali M. Mane ^[1], Mr.Hemant A. Tirmare ^[2]

Department of Technology
Shivaji University, Kolhapur
India

ABSTRACT

Keyword suggestion in web search helps user to access the information without knowing how to express in queries. The main concept of query suggestion used to retrieve documents from the related server. Search engine provide the platform for users to describe their information need more neatly by using query recommendation. Previously there has been lot of work done for retrieving relevant data of users to meet their information need and improving performance of search engines. This paper reviews and compares different available methods in query log processing for information retrieval. Then we conclude that Existing keyword suggestion techniques are not considering the locations of the users and the query results. The spatial factor is not considered to retrieve result. The approach based on location aware keyword query suggestion is better to understand user's interaction process with search engines to find the appropriate information need.

Keywords: — Query suggestion, Location aware keyword framework

I. INTRODUCTION

Keyword suggestion is most fundamental features of the web search engine. The major problem of current web search engine is that search queries are short and ambiguous. Users normally submit short queries to web search engine, and short queries are mostly ambiguous. Users try different queries to retrieve the relevant information because may have little knowledge about information of searching. The list of keywords displayed by the web search engine is not always good descriptor of the information needs of user.

In order to overcome this problem, many search engines have implemented the query suggestion method, Also known as keyword suggestion. The keyword suggestion helps user to search or to access the relevant information. The keyword suggestion module of the search engine recommends a set of keywords queries that are most likely to refine the users search. Effective method for keyword suggestion is based on information from query log. In this proposed method is based on a query clustering process in which groups of semantically similar queries are identified

II. LITERATURE SURVEY

The clustering process uses the content of historical preferences of users registered in the query log of the search engine. If the related queries are not present in the query log then apply the location aware keyword query suggestion.

The location aware keyword(LKS) query suggestion method provide the suggested queries retrieve documents which is related to user information and located near to users location.LKS framework, it construct and use keyword document bipartite graph(KD graph) that connect to keyword queries with their relevant document. LKS adjust weight on edges in KD graph to capture the semantics relevance between keyword queries and spatial distance between document location and user location. For distance calculation the Personalized PageRank(PPR) algorithm is used, it uses Random walk with restart(RWR) on KD graph, starting from user supplied query to find the set of keywords and spatial proximity to the user location. But RWR search has high computational cost on large graph to address this issue; a new portion based algorithm is used to reduce the cost of RWR search.

Shuyao Qi, Dingming Wu, and Nikos Mamoulis [1] designed a location-aware keyword query suggestion framework. They propose a weighted keyword-document graph, which captures

both the semantic significance between keyword queries and the spatial distance between the resulting documents and the user location. The graph is browsed in a random-walk-with-restart fashion, to select the keyword queries with the highest scores as suggestions. To make framework scalable, authors propose a partition-based approach that outperforms the baseline algorithm by up to an order of magnitude. And design the first ever Location-aware Keyword query Suggestion framework, for suggestions relevant to the user's information needs that also retrieve relevant documents close to the query issuer's location. Also extend the state-of-the-art Bookmark Colouring Algorithm (BCA) for RWR search to compute the location-aware suggestions.

Royi Ronen and et.al [2] introduced novel, domain-independent and privacy preserving methods for enhancing MF models by expanding the user-item matrix and by imputation of the user-item matrix, using browsing logs and search query logs. They introduced two approaches to enhancing user modelling using these data. Authors show that CF systems can be enhanced using Internet browsing data and search engine query logs, both represent a rich profile of individuals' interests. They demonstrate the value of their approach on two real datasets each comprising of the activities of tens of thousands of individuals. The first dataset details the download of Windows Phone 8 mobile applications and the second - item views in an online retail store. Both datasets are enhanced using anonymized Internet browsing logs.

Di Jiang and et.al [3] proposed a new query suggestion paradigm, Query Suggestion with Diversification and Personalization that effectively integrate diversification and personalization into one unified framework. In the QS-DP, the suggested queries are successfully diversified to cover different facets of the input query and the ranking of the suggested queries are personalized to ensure that the top ones that align with a user's personal preferences. They propose a new representation for query log. The proposed multi-bipartite-graph representation comprehensively captures different kinds of relations between search queries in query log. Based on the multi-bipartite-graph representation, they design two strategies to identify the most relevant suggestion candidate. Then an algorithm based on cross-bipartite hitting time is proposed to explore the latent structure of the multi-bipartite-graph representation and identify the remaining suggestion candidates with diversity-awareness. Finally, the User Profiling Model (UPM) is proposed to personalize the ranking of the query suggestion lists for each individual user.

Bin Zhang, Shubo Zhang and et.al [4] proposed a method that computes likeness among queries based on "Query- Clicked Sequence" model. This model counts weight of clicked document term by density of documents containing this term

on clicked sequence, and filters content of unrelated documents during similarity computation. Based on the characteristics of different concentration on relevant and irrelevant documents occurring on clicked document sequence, this paper proposed a query similarity computing method based on irrelevant feedback analysis, and recommended queries based on this method. This method constructs a relevant term collection for each clicked sequence of one query, from relevant document and computes similarity among queries by relevant term collection offline with recommendation of online queries based on the computation result. Query recommendation based on their method can effectively decrease the negative effect on query similarity computation, and increase accuracy of query similarity computation, therefore increase accuracy of query recommendation, especially for informational queries.

Magdalini Eirinaki and et.al [5] developed the QueRIE system for personalized query recommendations. QueRIE monitors the user's querying behavior and finds matching patterns in the system's query log, identifying "similar" users. Subsequently, these queries are used to recommend queries that the current user may find useful. They explore the use of latent factor models when, instead of ratings, the input consists of database-query log data. And explored how latent factor models, and in particular matrix factorization using ALS, affect the quality of the recommendations and computational efficiency of their framework. Such techniques have become very popular in traditional rating-based recommender systems, and in this work authors verified that they capture latent similarities between users and "items" even when the input is not explicit.

Taiki Miyanishi and et.al [6] proposed time aware structured query suggestion which clustered query suggestion along timeline so the user can narrow down his search from a temporal point of view. When the suggested query is clicked the method presents web pages from query-URL bipartite graph. After ranking those according to click count within a particular time period this method helping user to access relevant web pages. It free the users from burden of entering a specific time constraint with query, this method can be used in the context of real user search tasks.

R. Bhushan et al. [7] Explained a web recommender approach based on learning from web logs it recommends user a list of pages that are relevant to the users proposed query by comparing with historic pattern and also rerank the result pages. This system proves to be efficient as the pages desired by the users are on the top in the result list and this method reduces the search time of the user. In this the recommendation is based on the feedback of users and web log analysis.

P. Goyal et.al [8] proposed a snippet based method to facilitate users with query recommendations. The concepts related to the users information needs are suggested to the users to satisfy their current information need, extracted the concepts from the web snippet. Authors proposed two weight functions to measure the relevance between query and concept. Related concepts with different meaning are selected and recommended as query suggestions to the users.

Y. Liu et.al [9] presented an approach based on the users search behaviour. Their suggested query recommendation framework follows the fact that if user clicks certain result returned by search engine then it does not necessarily mean that the user is interested in that result but it probably reflects that the user is instead interested in the snippets of the result. This is because that up to that time the user clicks certain result just by viewing the snippet, the resultant document has not shown to user by that time.

D. Broccolo et al. [10] has explained a dynamic knowledge based approach which gets updated by continuously as queries are issued, to keep record of possible variations of user interest. This model extensively guesses the real hidden intent of user behind a submitted query and proves its effectiveness by dropping the effect of aging by updating & rebuilding the query recommendation model incrementally. In this the update operation runs in parallel with the query processor. Thus this dynamic knowledge based approach is better than that of all static models based on query log.

III. PROPOSED METHOD

A. Keyword query suggestion

There are different types of approaches for keyword query suggestion. This can be classified into three categories: random walk based, cluster based and learning to rank approaches. We briefly review the other methods from our observation any of given method cannot consider the user location in query suggestion

1. Random walk based approaches

This method uses graph structure for modeling the information that is provided by query log and then applies the random walk process on graph for query suggestion.

2. Cluster based approaches

In this method the query log is viewed as query URL bipartite graph. By applying the clustering algorithm on vertices in the graph, query cluster can be identified. Then, user supplied query q and queries that are belonging to same cluster as q does not returned to the user as suggestion

3. Learning to Rank Approaches

This approach is trained based on different type of query features like query performance prediction. Given query q, a list of suggestion is produced based on their similarity to q in topic distribution space.

B. Location aware keyword query suggestion

The proposed system contains location aware keyword query suggestion based on document proximity. In LKS (location aware keyword suggestion), it constructs and uses a bipartite graph. This graph is connected to keyword queries with their relevant documents. LKS adjusts weight on edges in graph to capture semantics relevance between queries and also spatial distance between document location and query issuer location. The partition based algorithm can be used to calculate the graph distance with highest semantics relevance to inserted keyword and near to query issuer location

The proposed system suggests the keyword queries using query log if and only if it is contained in the query log. When the keyword query is not present in query log then maximally contained rewriting is used to find approximate keyword. This system is effective if the database is big or distributed.

Figure 1 shows the architecture of the proposed system.

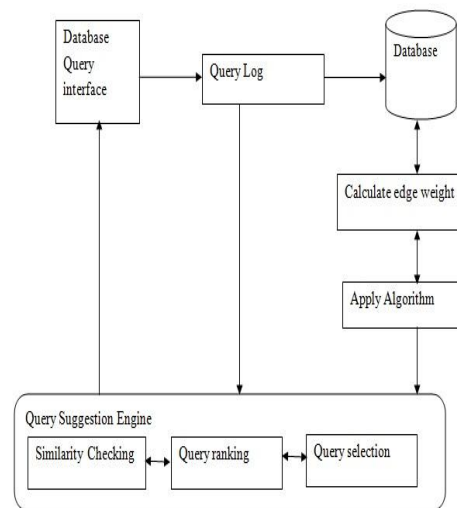


Figure 1: Architecture

Too many connection problems can be occurred during query search process, the speed of processing may be slow. Search queries are usually short and ambiguous. Users typically submit short queries to web search engine, and short queries are mostly ambiguous. Users searching for the same information may phrase their queries differently. Often, users try different queries until they are satisfied with the results. In order to formulate effective queries, users may need to be familiar with specific terminology in a knowledge domain. So to monitor users querying behaviour and find matching pattern

in the system's query log, attempt to identify users with similar information needs and use this information to recommend queries.

IV. CONCLUSIONS

User satisfaction plays very important role in information retrieval. Query recommendation is best method for helping users to satisfy the users information need by suggesting queries related to current users need by maintaining query log processing files, by using past historical navigation patterns, by updating the records of query processing so that by using dynamic and static log data and so on. This paper helps to review some of these query recommendation techniques. So for further research we plan to make use of the query log for keyword query suggestion method. This method is useful when the database is big or distributed. This method reduces the cost of data communication. In our future work we want to prepare synonym suggestion for given keyword query using offline dictionary interface.

REFERENCES

- [1] Shuyao Qi, Dingming Wu and Nikos Mamoulis, "Location aware keyword Query suggestion based on document proximity", IEEE transaction paper on knowledge and data engineering, Vol 28, No 1, pp 82-97, 2016.
- [2] Royi Ronen and et.al, "Recommendations Meet Web Browsing: Enhancing Collaborative Filtering using Internet Browsing Logs", IEEE 32nd International Conference on Data Engineering (ICDE), pp 1230 – 1238, 2016.
- [3] Di Jiang, Kenneth Wai-Ting Leung, Lingxiao Yang, Wilfred Ng, "Query suggestion with diversification and personalization", Knowledge-Based Systems. Volume 89, Pages 553–568, 2015.
- [4] Bo Zhang, Bin Zhang, Shubo Zhang, Chao Ma, "Query Recommendation Based on Irrelevant Feedback Analysis", 8th International Conference on Bio Medical Engineering and Informatics, pp 644-648, 2015.
- [5] Magdalini Eirinaki, "QueRIE reloaded: Using matrix factorization to improve database query recommendations", IEEE international conference on Big Data, pp 1500 – 1508, 2015.
- [6] T. Miyanishi and T. Sakai, "Time-aware structured query suggestion," in Proc. 36th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, pp. 809–812, 2013.
- [7] R. Bhushan and R. Nath. 2012. Recommendation of optimized web pages to users using Web Log mining techniques. Advance Computing Conference (IACC), 2013 IEEE 3rd International. IEEE, 2012.
- [8] Goyal, Poonam, and N. Mehala. 2011. Concept based query recommendation. Proceedings of the Ninth Australasian Data Mining Conference-Volume 121. Australian Computer Society. Inc.
- [9] Y. Liu, Junwei Miao, Min Zhang, Shaoping Ma, and Liyun Ru. 2011. How do users describe their information need: Query recommendation based on snippet click model. Expert Systems with Applications 38, no. 11 (2011): 13847-13856.
- [10] D. Broccolo, O. Frieder, F. Nardini, R. Perego and F. Silvestri. 2010. Incremental Algorithms for Effective and Efficient Query Recommendation. SPIRE 2010, LNCS 6393. pp. 13-24. Springer-Verlag Berlin Heidelberg.
- [11] Zahera, H., G. El Hady, and W. El-Wahed. 2010. Query Recommendation for Improving Search Engine Results. World Congress on Engineering and Computer Science (WCECS), San Francisco, USA. Vol. 1. 2010.
- [12] Sumathi, C. P., R. Padmaja Valli, and T. Santhanam. 2010. Automatic recommendation of web pages in web usage mining. International Journal on Computer science and Engineering (IJCSE). Vol 2 (2010): 3046-3052
- [13] He, Qi, Daxin Jiang, Zhen Liao, S. Hoi, Kuiyu Chang, Ee-Peng Lim, and Hang Li. 2009. Web query recommendation via sequential query prediction. IEEE 25th International Conference on Data Engineering, 2009. ICDE'09, IEEE, 2009. pp. 1443-1454.
- [14] Song and L.-W He, "Optimal rare query suggestion with implicit user feedback," in Proc. 19th Int. Conf. World Wide Web, pp. 901–910, 2010.