

A Survey on Query Recommendation Techniques and Evaluation of Snippet based Query Recommendation

Megha R. Sisode
PG Student

Department of Computer Engg.,
R.C.Patel Institute of Technology, Shirpur, 425405,
India

Ujwala M. Patil
Associate Professor

Department of Computer Engg.,
R.C.Patel Institute of Technology, Shirpur, 425405,
India

ABSTRACT

Recently web has been widely used for getting different kinds of information. Web mining is considered to store the information in a specific format known as weblog. This valuable mined information can be used in many applications such as query log analysis, query recommendation, query reformulation and many more for improved performance of search engine. Search engine provide the platform for users to describe their information need more clearly by using query recommendation. Previously there has been lot of work done for retrieving more relevant data to users in order to meet their information need thus improving performance of search engines. This paper reviews and compares different available methods in query log processing for information retrieval. Moreover the approach based on clicked snippets is better to understand users interaction process with search engines to find the appropriate information need.

Keywords

Web mining, recommendations, knowledge extraction, query log processing, user behaviour analysis, query intent identification, search engines.

1. INTRODUCTION

Nowadays the growth of World Wide Web is increasing so the size and popularity is also increased with the assembly of large scale volumes of web data. Thus it is difficult to extract the relevant information that have been used in wide range of application. Many novice users face the difficulty to get the desired information although they use most efficient search engines such as yahoo, Google.

There is a difference between search engine and recommendation system such as user uses search engine if they know their query in proper wording i.e. exactly what query must be fired to satisfy the information need. In contrast user prefers to use the recommendation system if they do not know from where exactly they can get their query solution and the appropriate wording to fire the query. So the query recommendation system helps users to describe their information needs more clearly so that search engine can return appropriate answers and meet their information needs.

Thus improving user satisfaction is a key challenge for web search engines. For this purpose search engine provide way to the users to specify their information need more clearly in the form of queries simply as list of keywords or phrases. Many search engine companies apply significant efforts to develop means that correctly guess by which hidden intent the user has submitted the given query. In the recent years, web search engines have started to provide users with query recommendation to help them reformulate queries and quickly satisfy their needs. Query recommendation suggests related

queries for search engine users when they are not satisfied with the results of an initial input query, thus assisting users in improving search quality [8].

There is problem of improving search engine result and obtaining the desired information from web because of following common factors 1. Rapid growth in the number of pages indexed in a search engine. 2. Short and ambiguous queries submitted by the web users. 3. Ineffective organization of the search results 4. User's different goals and expectations from the web, etc. [10]

The problem is not in obtaining the large scale results but its in the keywords used in searching are not strongly related thus this keyword-based user interface causes lots of troubles in search process because it is found that the average length of user queries is 2.35 terms also most user queries are short (around two terms per query). A short list of keywords is not always a good descriptor of the information needs of search users because it may have ambiguities either in content or in information need [12].

Many search engines provide query suggestions to users to formulate more effective queries such as Yahoo's "Also Try", Google's "Searches related to" etc. These suggestions are semantically related, but mostly start with the terms which users have used in their queries. Thereby, ambiguity which was present in the original query still remains in it [10]. Also query reformulation is the frequent insertion, deletion or modification of query terms to a seed query made by users in the hope of retrieving results efficiently. Also In order to help users to reorganize their short, ill-formed, and possibly ambiguous queries, search engines develop query recommendation function.

Query recommendation as shown in the figure 1 has been widely adopted by search users as an important way in finding information effectively. According to search behavior survey report, 78.2% users will change their queries (mostly by adopting search engine's recommendation function) if they cannot obtain satisfactory results with the current query and Users adopt query recommendation function to clarify their information needs without taking efforts in inputting new queries. Therefore, it is important for search engines to provide high-quality recommendations which can represent users exact information needs [12].

Query log can be a source of additional evidence to help future users. While searching the web, user formulates a query to represent the information need that is usually short and ambiguous; this affects the search results by retrieving irrelevant documents. This can be reduced by reformulating the query with the help of query recommendations or rewriting. Also, the query log can be analyzed to organize it in

well accessible manner by using different classification and clustering mechanisms. The search engine user submits a query with keeping some underlying goal into mind and interest to seek required information. Thus, the user's search intentions can be determined by analyzing their search behavior from query log. The identified search intents can be further used to improve user satisfaction. Along with all these methods, implicit feedback of the user can be utilized for efficient retrieval. The implicit feedback can be obtained from search engine log by finding sequence of subsequent queries by the same user.

Currently, many available commercial search engines and lots of research work focus on how to recommend queries based on users previous query and click behaviors. The idea behind their work is to locate popular queries which are similar with the current query either in content or in click context i.e. click context refers to the documents ever clicked by users for their query. According to research if two queries share similar click context, it is supposed that they are similar and relevant. This method suggests user to adopt a similar and frequently-adopted queries to finish his search task but the major problem with this kind of recommendation methods is that it lacks understanding of users actual information needs also it does not take current users search intent into consideration. Instead it supposes that he shares similar interests with other users who propose similar queries. This assumption is correct for most hot queries. However, it sometimes fails to give proper recommendations, especially for low-frequency queries because there are not so many candidate queries for them [12].

Even after the huge research in the area of query recommendation the increasing need of information at a single click encourages researchers to take efforts in improving the results of search engines. The study and review of recent work related to different methods has inspired us to focus on their relative performance and try to improve the search results.

The rest of the paper is arranged as follows. Section 2 gives literature survey on work done in query log processing methods; the classification of query log processing techniques and the evaluation of snippet based query recommendation is illustrated in section 3. Section 4 describes conclusion and the future work.

2. LITERATURE SURVEY

As queries are one of the most important interface for users to access the information over web, also it affects the performance of search engines. Therefore the query recommendation approach is proposed to suggest a list of possible query choices to users, so that users can able to stat their information need by clicking recommendation query links rather than inputting new queries. For defining the relatedness among the queries the previous recommendations methods are categories as, content based, click-context based, clustering methods and so on to find the similar queries.

Although the query log data of search engine is a huge source of information on users search behaviour, the knowledge extraction from large scale query log with efficiency and effectiveness is still a challenging task. In general, two types of information can be extracted from search logs, i.e., click-through information and session information. Both have been used for query understanding. Approaches based on click-through information assume two queries to be related if they share many clicked URLs. The related queries are usually grouped into clusters and used for recommendations for each other. A literature survey is done to identify different

approaches proposed by researchers in order to mine essential features from query log data of search engine.

J. Wen et al. [1] presented a content based similarity measure to cluster similar queries to recommend URLs to frequently asked queries of a search engine by using four notions according to: first, the context of the query; second, common clicked URL's between queries; third, string matching of keywords, and fourth, the distance of the clicked documents in some pre-defined hierarchy. But result of this method generates very sparse distance matrices but this sparsity is diminished using large query logs. Thus string matching features are used to locate similar queries.

O. Zaiane et al. [2] have used content similarity to recommend similar queries using Query Memory, which is a data structure that holds the collective query trace and also extra information pertaining to the queries that would help in measuring similarities between queries. Query trace is a log containing previously submitted queries. The major advantage of this method is that it suggests the queries when user is not satisfied by current search result but sometimes produces irrelevant result and leaves the choice up to user.

S. Cucerzan et al. [5] have presented a click context based method that suggests queries based on mining into post-query browsing behaviors referred as search trails. They utilized user landing pages which are the ending pages of search trails to generate query suggestions. For each landing page of a user submitted query they identify queries from query logs that have these landing pages as one of their top 10 results ad these queries are used for suggestions.

C. Sumathi et al. [9] also proposed a session based approach where the proposed system is based on the users navigational patterns and provide recommendations to satisfy the current users information need. In this method they classify and match an online user based on his browsing interests.

Q. He et al. [7] used a session based novel sequential query prediction approach to grasp a users search intent based on users past query sequence and its resemblance to historical query sequence models mined from massive search engine logs. Differently from previous work where only single preceding query is used for prediction, this work considers variable number of preceding query and effectively captures more complex context information for recommendation. Results show that the sequence-wise approaches significantly outperform the conventional pair-wise ones in terms of prediction accuracy. Thus the work has one fundamental difference from all previous session-based approaches. As all previous work focuses on pair-wise query relations and uses only a single preceding query for query prediction, presented method consider a variable number of preceding queries and effectively capture more complex context information for query recommendation. Moreover, this approach can automatically determine the optimal context length to be used for query prediction.

R. Baeza-Yates et al. [4] explained a method to suggest a list of related queries to user based on a query clustering process. The method not only discovers the related queries, but also ranks them according to a relevance criterion. This notion of query similarity has several advantages that it is simple and easy to compute. On the other hand, it allows relating queries that are worded differently but stem from the same topic, hence capturing semantic relationships among queries.

H. Zahera et al. [8] proposed a method for suggesting a list of queries that are related to the user input query. The related

queries are based on previously issued queries by the users. The proposed method is based on clustering process in which groups of semantically similar queries are detected. This facility provides some queries which are related to the queries submitted by users in order direct them toward their required information. This method not only discovered the related queries but also rank them according to a similarity measure.

Y. Liu et al. [12] presented an approach that is based on the users search behavior. 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.[3] 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 significantly guesses the real hidden intent of user behind a submitted query and proves its efficiency by reducing 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.

R. Bhusan et al. [11] have explained a web recommender approach based on learning from web logs and recommends user a list of pages which 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 thus 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. [10] have proposed a snippet based method to facilitate users with query recommendations in which the concepts related to the users information need are suggested to the users to satisfy their current information need, extracted the concepts from the web snippets and 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.

S. Xiaoyan et al. [6] used an effective method for query suggestions. This method accepts Chinese web query as input and the approach not only identify related queries already existed in the log of previously submitted queries of search engine but also use synonyms that are extracted from web based corpora to construct new related queries. Also rank the queries according to degree of relatedness, freshness and effectiveness. This method proves its effectiveness in recommending related queries for high frequency queries that the low frequency query.

Differently with these methods there are many approaches which recommend the queries to the user by using keywords of the phrases, by extracting the snippets of related queries. However the performance of recommendation in the existing systems sometimes fails to give proper recommendations, mainly for low frequency queries because very few candidates have queried for them.

3. SURVEY

The approach proposed by Y. Liu et al. [12] have the understanding of users actual information need and also used this information need efficiently to organize queries with more exact meaning. The snippet based approach is based on the consideration on that users information needs are better described in their interaction with search engine more specifically, in the snippets of the results which ever clicked by users. This assumption is based on the fact that when user clicks a certain search result, it does not mean that the user is interested with the result this is because the user has not viewed the resulted document yet. It is probably that the user might be interested in the snippets of the corresponding result because these snippets are actually shown to and read by the users. So on the basis of this assumption the snippet click based query at recommendation framework proves its effectiveness over others. So in future we expect to elaborate this approach for better suggestions with improvement in performance.

Table 1 explains comparative study of different query recommendation techniques discussed in section 2.

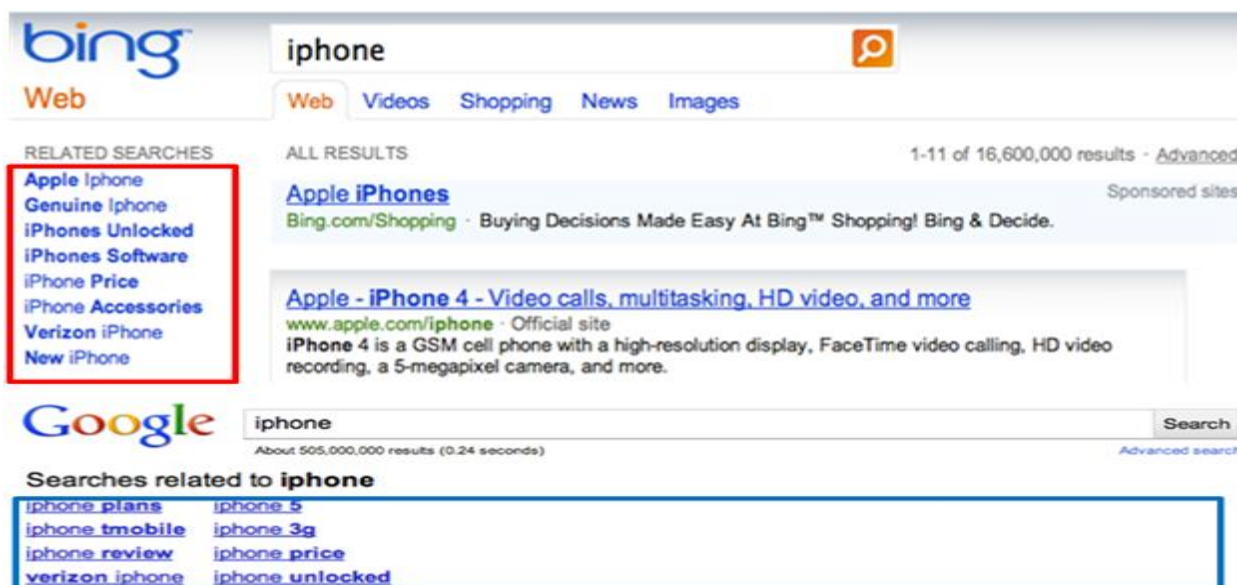


Fig 1: Query Recommendation function provided by commercial search engines

Table 1. Survey on query recommendation techniques

Sr. No.	Author	Title	Publication year	Remark
1	R. Bhushan et al. [11]	Recommendation of optimized Web Pages to Users Using Web Log Mining Techniques	2012	Recommends a list of relevant pages by comparing query with historic pattern. Mainly the recommendation is based on users feedback and analysis of web log. It reduces the search time of the user.
2	P. Goyal et al. [10]	Concept Based Query Recommendation	2011	Recommends the concepts extracted from the web snippets without using query logs and measure the relevance between query and concept.
3	Y. Liu et al. [12]	How do users describe their information need: Query recommendation based on snippet click model	2011	Uses users behaviour information for query recommendation. Recommendation is based on the snippets clicked by user and also generates good results for low frequency queries.
4	D. Broccolo et al. [3]	Incremental Algorithms for Effective and Efficient Query Recommendation	2010	A dynamic knowledge based approach which gets updated continuously as queries are issued. This model reduces the effect of aging by updating and rebuilding the query recommendation model incrementally.
5	H. Zahera et al. [8]	Query Recommendation for Improving Search Engine Results	2010	Recommendation is based on clustering process in which groups of semantically similar queries are extracted from query log.
6	C. Sumathi et al. [9]	Automatic Recommendation of Web Pages in Web Usage Mining	2010	Uses a session based approach based on the users navigational patterns. Also classify and match an online user based on browsing interests.
7	Q. He et al. [7]	Web Query Recommendation via Sequential Query Prediction	2009	A novel sequential query prediction approach is used to grasp a users search intent based on past query sequence and its resemblance to historical query sequence models. Differently, this work considers variable number of preceding query
8	S. Xiaoyan et al. [6]	An Effective Method for Chinese Related Queries Recommendation	2008	Identify already existed related queries in the log of previously submitted queries of search engine also use synonyms that are extracted from web based corpuses to construct new related queries
9	S. Cucerzan et al. [5]	Query Suggestion based on User Landing Pages	2007	Suggests queries based on mining into post-query browsing behaviours i.e. search trails. They utilized user landing pages (the ending pages of search trails).
10	R. Baeza-Yates et al. [4]	Query Recommendation using Query Logs in Search Engines	2004	Suggest a list of related queries to user based on a query clustering process by capturing semantic relationships among queries.
11	O. Zaiane et al. [2]	Finding Similar Queries to Satisfy Searches based on Query Traces	2002	Uses content similarity to recommend similar queries using Query Memory, this suggests the queries when user is not satisfied by current search result but sometimes produces irrelevant results.
12	J. Wen et al. [1]	Clustering user queries of a search engine	2001	Clusters similar queries to recommend URLs to frequently asked queries, but sometimes generate very sparse results.

4. CONCLUSIONS AND FUTURE WORK

As the user satisfaction plays important role in information retrieval. Query recommendation is one of the 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, also by using clicked snippets, and so on. This paper helps to review some of these query recommendation techniques with the limitations and advantages. Moreover according to review the method based on snippet click model is effective for query recommendation even for the low frequency queries and better to have understanding of users actual information need behind the submitted query. So for further research we plan to make use of the snippet based query recommendation method.

In future, this method also can be enhanced by incorporating the synonyms extraction of keywords from original query retrieved from snippets for query recommendation, which can expect to further improve the performance of searching on the web.

5. REFERENCES

- [1] J. Wen, J. Nie, and H. Zhang. 2001. Clustering user queries of a search engine. In Proceedings of the 10th international World Wide Web conference. W3C. pp. 162–168.
- [2] O. Zaiane and A. Strilets. 2002. Finding similar queries to satisfy searches based on query traces. In Proceedings of the International Workshop on Efficient Web-Based Information Systems (EWIS), Montpellier, France.
- [3] 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.
- [4] R. Baeza-Yates, C. Hurtado and M. Mendoza. 2004. Query Recommendation Using Query Logs in Search Engines. LNCS 3268. pp. 588-596. Springer-Verlag Berlin Heidelberg.
- [5] S. Cucerzan and R. White. 2007. Query suggestion based on user landing pages. In Proceedings of SIGIR 2007. Amsterdam, Netherland.
- [6] Shen Xiaoyan, Cheng Bo, Chen Junliang and Meng Xiangwu. 2008. An Effective Method for Chinese Related Queries Recommendation. In Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2008. SNPD '08. 6-8 Aug. 2008. pp. 381 - 386.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] Goyal, Poonam, and N. Mehala. 2011. Concept based query recommendation. Proceedings of the Ninth Australasian Data Mining Conference-Volume 121. Australian Computer Society. Inc.
- [11] 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.
- [12] 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.