

A Query Based Algorithm for Optimal Location Search

Rucha Deshmukh, Pradnya Kasture

Department of Computer Engineering, RMD Sinhgad School of Engineering, Waraje, Pune, Savitribai Phule Pune University, Maharashtra, India

ABSTRACT: We design a location-aware keyword query suggestion framework. We propose a weighted keyword-document graph, which captures both the semantic relevance 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 our framework scalable, we propose a partition-based approach that outperforms the baseline algorithm by up to an order of magnitude. The appropriateness of our framework and the performance of the algorithms are evaluated using real data.

KEYWORDS: Keyword Document Graph, Keyword queries, Spatial distance, Baseline Algorithm

I. INTRODUCTION

Data mining is the information of domain we are mining like concept hierarchies, to organize attributes onto various levels of abstraction. A Spatial Keyword query is an approach of searching qualified spatial objects by considering both the query requester's location and user specified keywords. Taking both spatial and keyword requirements into account, the goal of a spatial keyword query is to efficiently find results that satisfy all the conditions of a search. Searching is a common activity happening in data mining.

a. Motivation

This motivated to develop methods to retrieve spatial objects. A spatial objects consists of objects associated with spatial features. In other words, spatial objects involve spatial data along with longitude and latitude of location. The importance of spatial databases is reflected by the convenience of modeling entities of reality in a geometric manner. For example, locations of restaurants, hotels, hospitals and so on are often represented as points in a map, while larger extents such as parks, lakes, and landscapes often as a combination of rectangles. Many functionalities of a spatial database are useful in various ways in specific contexts. For instance, in a geography information system, range search can be deployed to find all restaurants in a certain area, while nearest neighbor retrieval can discover the restaurant closest to a given address. However, existing keyword suggestion techniques do not consider the locations of the users and the query results. Users often have difficulties in expressing their web search needs they may not know the keywords. After submitting a keyword query, the user may not be satisfied with the results. Writing the queries is never easy because usually queries are short and words are ambiguous because user may not know how to use query in web search so that we suggest a user to use a single word query it makes the user to feel comfortable when they enter a keyword query. However, none of the existing methods provide location aware keyword query suggestion, such that the suggested keyword queries can retrieve documents not only related to the user information needs but also located near the user location. This requirement emerges due to the popularity of spatial keyword search that takes a user location and user-supplied keyword query as arguments and returns objects that are spatially close and textually relevant to these arguments. For example the tourist may also search for all the hotels which are within 10 miles of the airport and provide the two amenities in order to compare the hotels' reviews and prices. For retrieving the qualified hotels, the tourist will launch a Fast Nearest Search query with ranking parameters for the first search; the query results

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

are hotels. Last, we test our query suggestion approach on the search log. The experimental results clearly show that our approach outperforms keyword document graph and fast nearest search in both coverage and quality of suggestions.

b. Background

Keyword suggestion (also known as query suggestion) has become one of the most fundamental features of commercial web search engines. After submitting a keyword query, the user may not be satisfied with the results, so the keyword suggestion module of the search engine recommends a set of m keyword queries that are most likely to refine the user's search in the right direction. Effective keyword suggestion methods are based on click information from query logs and query session data, or query topic models. New keyword suggestions can be determined according to their semantic relevance to the original keyword query. However, to our knowledge, none of the existing methods provide location-aware keyword query suggestion (LKS), such that the suggested queries retrieve documents not only related to the user information needs but also located near the user location. This requirement merges due to the popularity of spatial keyword search. Google processed a daily average of 4.7 billion queries in 2011, a substantial fraction of which have local intent and target spatial web objects (i.e., points of interest with a web presence having locations as well as text). Data mining is the information of domain we are mining like concept hierarchies, to organize attributes onto various levels of abstraction. A Spatial Keyword query is an approach of searching qualified spatial objects by considering both the query requester's location and user specified keywords. Taking both spatial and keyword requirements into account, the goal of a spatial keyword query is to efficiently find results that satisfy all the conditions of a search. Searching is a common activity happening in data mining. This motivated to develop methods to retrieve spatial objects. A spatial object consists of objects associated with spatial features. In other words, spatial objects involve spatial data along with longitude and latitude of location. The importance of spatial databases is reflected by the convenience of modeling entities of reality in a geometric manner. For example, locations of restaurants, hotels, hospitals and so on are often represented as points in a map, while larger extents such as parks, lakes, and landscapes often as a combination of rectangles. Many functionalities of a spatial database are useful in various ways in specific contexts. For instance, in a geography information system, range search can be deployed to find all restaurants in a certain area, while nearest neighbor retrieval can discover the restaurant closest to a given address. However, existing keyword suggestion techniques do not consider the locations of the users and the query results. Users often have difficulties in expressing their web search needs they may not know the keywords. After submitting a keyword query, the user may not be satisfied with the results.

II. RELATED WORK

Authors in [1] propose a novel context-aware query suggestion approach which is in two steps. In the offline model-learning step, to address data sparseness, click-through bipartite is clustered in order to summarize queries into concepts. In this approach queries are suggested to the user in a context-aware manner.

Authors in [2] propose a novel query suggestion algorithm based on ranking queries with the hitting time on a large scale bipartite graph. This method captures the semantic consistency between the suggested query and the query given by user. Experiments show time is effective to generate semantically consistent query suggestions. The proposed algorithm and its variations can successfully execute huge queries, accommodating query suggestion.

Author [3] 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 modeling 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.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

Author [4] 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.

Author [5] 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.

Author [6] 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 same kind of users. These queries are used to recommend queries which user may find helpful. 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.

Author [7] 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.

Author [8] 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.

Author [9] 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.

Author [10] 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.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

Author [11] 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.

Author [12] 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. Keyword query suggestion approaches can be classified into three main categories: random walk based approaches, learning to rank approaches, and clustering based approaches. We also briefly review alternative methods that do not belong to any of these categories. To the best of our knowledge, no previous work considers user location in query suggestion. The methods in this category use a graph structure to model the information provided by query logs, and then apply a random walk process on the graph to compute the suggestions.

III. EXISTING SYSTEM APPROACH

Keyword suggestion in web search helps users to access relevant information without having to know how to precisely express their queries. Existing keyword suggestion techniques do not consider the locations of the users and the query results; i.e., the spatial proximity of a user to the retrieved results is not taken as a factor in the recommendation. However, the relevance of search results in many applications (e.g., location-based services) they did not give the correct correlance. A baseline algorithm extended from algorithm BCA is introduced to solve the problem. Then, we proposed a partition-based algorithm (PA) which computes the scores of the candidate keyword queries at the partition level and utilizes a lazy mechanism to greatly reduce the computational cost. The performance of the proposed algorithms is low.

Disadvantages:-

1. Generally Google map are not view the current location.
2. They are not provided the shortest location between two locations.

IV PROPOSED SYSTEM ARCHITECTURE

We proposed to providing keyword suggestions that are relevant to the user information needs and at the same time can retrieve relevant documents near ideas, but aims at optimizing different objective functions. The concept of prestige based spatial keyword search. The SI-index comes with two query algorithms based on merging and distance browsing respectively. To design a variant of inverted index that is optimized for multidimensional points, and is thus named the Spatial Inverted index (SI-index). To remedy the situation by developing an access method called the spatial inverted index (SI-index). Not only that the SI-index is fairly space economical, but also it has the ability to perform keyword augmented nearest neighbor search in time that is at the order of dozens of milli-seconds

ADVANTAGES OF PROPOSED SYSTEM

1. Keyword suggestion techniques consider the locations of the users and the query results
2. This approach is very useful to find the nearest location of the user.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

- After submitting a keyword query, the user may satisfy with the results.

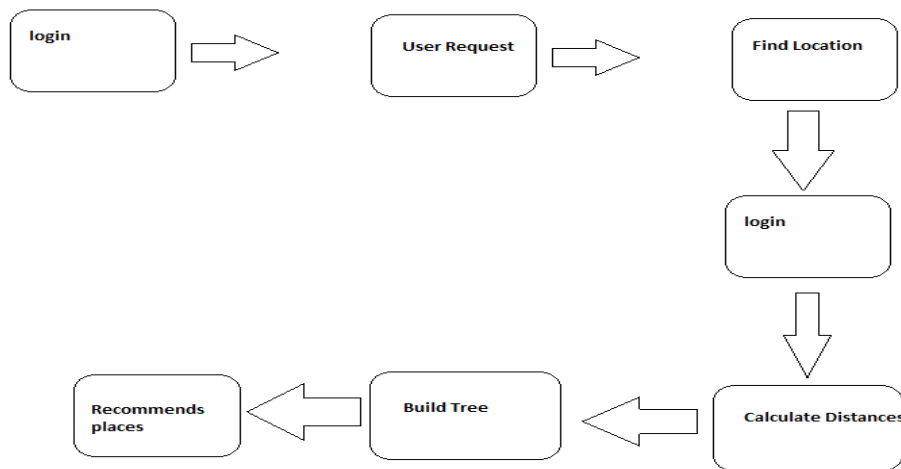


Fig 1. Proposed System Architecture

V. EXPERIMENTAL RESULTS

Keyword query suggestion approaches can be classified into three main categories: random walk based approaches, learning to rank approaches, and clustering based approaches. We also briefly review alternative methods that do not belong to any of these categories.



Fig 2. Time vs Document

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

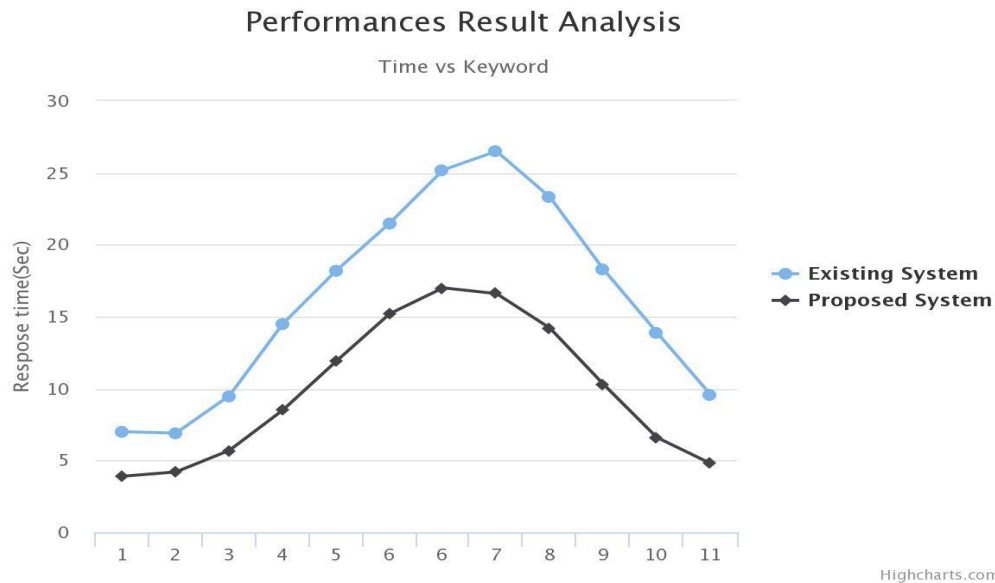


Fig 3. Time vs Keyword

VI. CONCLUSION

In this paper, we proposed an LKS framework providing keyword suggestions that are relevant to the user information needs and at the same time can retrieve relevant documents near the user location. A baseline algorithm extended from algorithm BCA is introduced to solve the problem then, we proposed a partition-based algorithm which computes the scores of the candidate keyword queries at the partition level and utilizes a lazy mechanism to greatly reduce the computational cost. Empirical studies are conducted to study the effectiveness of our LKS framework and the performance of the proposed algorithms. The result shows that the framework can offer useful suggestions and that PA outperforms the baseline algorithm significantly. In the future, we plan to further study the effectiveness of the LKS framework by collecting more data and designing a benchmark. In addition, subject to the availability of data, we will adapt and test LKS for the case where the locations of the query issuers are available in the query log. Finally, We believe that PA can also be applied to accelerate RWR on general graphs with dynamic edge weights; we will investigate this potential in the future.

REFERENCES

- [1] H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li, Context-aware query suggestion by mining click-through and session data, in Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2008, pp. 875–883.
- [2] Mei, Qiaozhu, Dengyong Zhou, and Kenneth Church. Query suggestion using hitting time. Proceedings of the 17th ACM conference on Information and knowledge management. ACM, 2008.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] MagdaliniEirinaki ,QueRIE reloaded: Using matrix factorization to improve database query recommendations, IEEE international conference on Big Data, pp 1500 – 1508, 2015.
- [7] 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.



ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

- [8] 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.
- [9] Goyal, Poonam, and N. Mehala. 2011. Concept based query recommendation. Proceedings of the Ninth Australasian Data Mining Conference- Volume 121.Australian Computer Society.Inc.
- [10] Y. Liu, Junwei Miao, Min Zhang, Shaoping Ma, and LiyunRu. 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.
- [11] 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.
- [12] ShuyaoQi,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.
- [13] N. Craswell and M. Szummer, Random walks on the click graph, in Proc. 30th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2007, pp. 239–246.
- [14] U. Ozertem, O. Chapelle, P. Donmez, and E. Velipasaoglu, Learning to suggest: A machine learning framework for ranking query suggestions, in Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2012, pp. 25–34.
- [15] Y. Liu, R. Song, Y. Chen, J.-Y.Nie, and J.-R. Wen, Adaptive query suggestion for difficult queries, in Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2012, pp. 15–24.