

Location Aware Keyword Query Suggestion Based On Document Proximity

Mansi Satpute, Shreyans Gandhi, Shital Raykar ,Mayur Kokate, Prof. Suryavanshi P.M.

BE Students, Department of Computer Engineering, Computer Engineering, H.S.B.P.V.T's COE, Kashti, Ahmadnagar,
Savitribai Phule Pune University, Pune India

Professor, Department of Computer Engineering, Computer Engineering, H.S.B.P.V.T's COE, Kashti, Ahmadnagar,
Savitribai Phule Pune University, Pune India.

ABSTRACT: Keyword suggestion in web search helps user to access relevant information without having to know how to precisely express their queries. Existing keyword suggestion techniques do not consider the location of user and the query result. The spatial proximity of user to the retrieved result is not taken as a factor in the recommendation. However, the relevance of search result in many application location based services is known to be correlated with proximity to the query issuer. Each query is related to one of topics identified in the conversation fragments preceding the recommendation and is submitted to a search engine over the English. We propose in this paper an algorithm for diverse merging of these lists using a sub modular reward function that rewards the topical similarity of documents to the conversation words as well as their diversity. We evaluate the proposed method through crowd sourcing the result superiority of the diverse merging technique over several other which enforce the diversity of topics.

KEYWORDS: Query suggestion, Spatial Databases, Document Proximity.

I. INTRODUCTION

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

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 3, March 2017

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

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 constructs 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.

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 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 anonym zed Internet browsing logs.

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

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 3, March 2017

negative effect on query similarity computation, and increase accuracy of query similarity computation, therefore increase accuracy of query recommendation, especially for informational queries.

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 correlation. 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:-

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

IV. PROPOSED SYSTEM APPROACH

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:-

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

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 3, March 2017

V. SYSTEM ARCHITECTURE

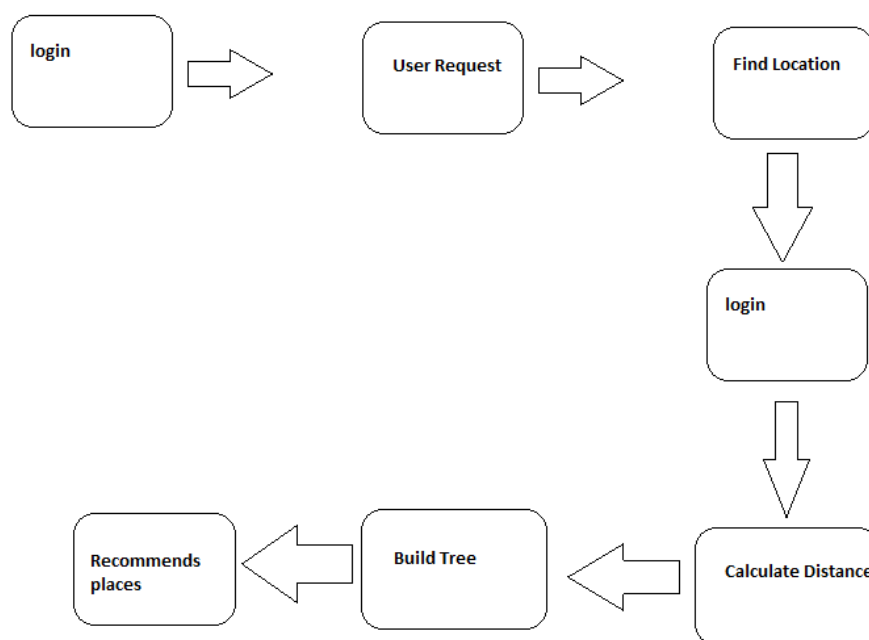


Fig No 01 System Architecture

VI. MODULE DESCRIPTION

Keyword-Document (KD) Graph Construction:

In Location-aware Keyword query Suggestion (LKS) framework constructs an initial keyword-document graph (KD-graph). This directed weighted bipartite graph between Documents and Keyword queries captures the semantics and textual relevance between the keyword query and document nodes; i.e., the first criterion of location-aware suggestion.

Partition Algorithm:

In this partition algorithm, it will divide the keyword queries and documents in the KD-Graph into groups. By doing this, we can improve the performance of the Baseline algorithm.

Selecting keyword Query Suggestion:

In this module, we have to select the suggestions i.e., after adjusting the weights for KD-graph based on the query location we have two selection suggestions those are relevance to the keyword query and closeness to the query location. The suggestions means here, which nodes having highest scores in the query graph those nodes are the suggestions.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 3, March 2017

Advantages of proposed system

- The proposed framework can offer useful suggestions and that PA outperforms the baseline algorithm significantly.
- Reduce the Computational cost by using Partition-based algorithm
- Keyword suggestion techniques consider the locations of the users and the query results.
- This approach is very useful to find the nearest location of the user.
- After submitting a keyword query, the user may satisfy with the results

VII. MATHEMATICAL MODEL

- Let S, be a system such that,
- $S = \{s, \text{Subi}, \text{Pubi}, \text{Si}, E\}$
- Where,
- S- Proposed System
- s- Initial state at
- Subi = the Registered Subscriber with attributes $\{\text{Sid}, \text{pwd}, \text{Li}\}$ where
- Sid = subscriber_id
- Li = Location_id
- Pubi = the Registered Subscriber with attributes $\{\text{Pid}, \text{pwd}, \text{Li}\}$ where
- Li = Location_id
- E – Event
- N – Node
- S – Subscription
- UB – Upper Bound
- Sid – Identifier an subscription

X- Input of System.

-L(Location),S(Searching) and R tree node

Y- Output of System. – Top K Subscription matching and show in Map

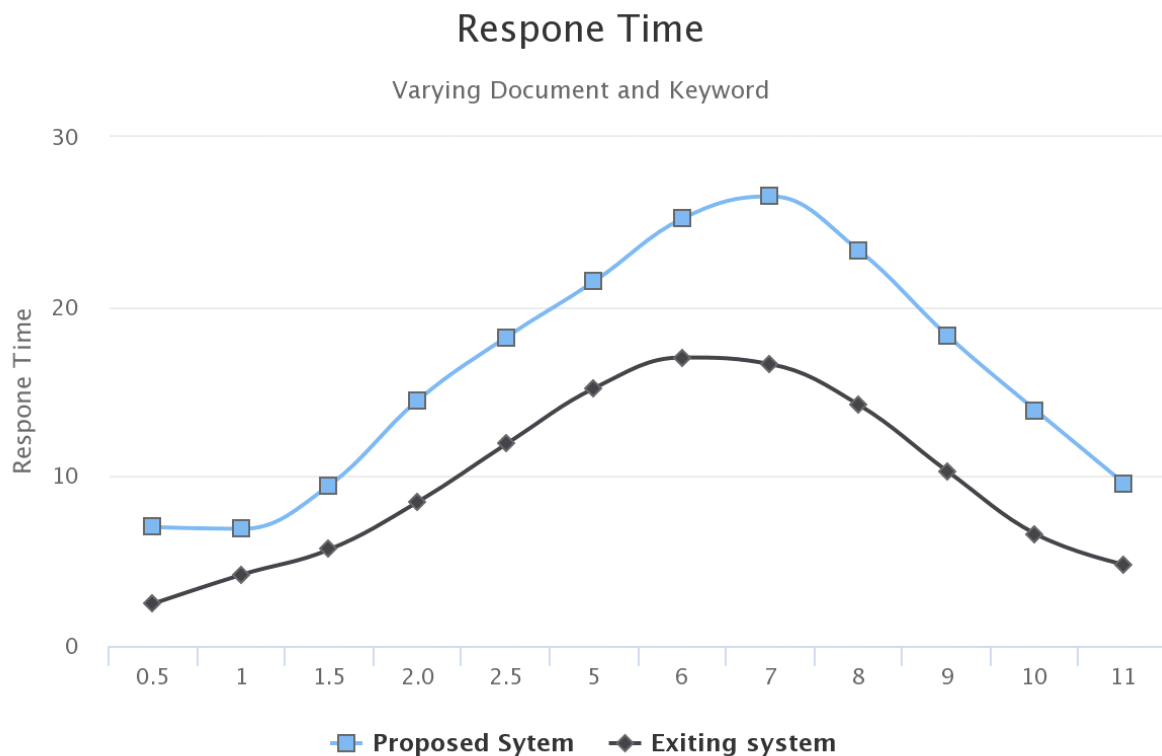
International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 3, March 2017

VIII. EXPERIMENTAL RESULT



Highcharts.com

IX. 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. A query suggestion log analysis, M. P. Kato, T. Sakai, and K. Tanaka, "When do people use query suggestion? Inf. Retr., vol. 16, no. 6, pp. 725–746, 2013.
2. Query recommendation using query logs in search engines, R. Baeza-Yates, C. Hurtado, and M. Mendoza, in EDBT, 2004, pp. 588–596.
3. Agglomerative clustering of a searchengine query log, D. Beeferman and A. Berger, in KDD, 2000, pp. 407–416.



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 3, March 2017

4. Location Aware Keyword Query Suggestion based on Document Proximity” Shuyao Qi, Dingming Wu, and Nikos Mamoulis, in IEEE,2015,pp.82-97.
5. 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.
6. 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.
7. J.-R. Wen, J.-Y. Nie, and H.-J. Zhang, “Clustering user queries of a search engine,” in Proc. 10th Int. Conf. World Wide Web, 2001, pp. 162–168.
8. R. Li, B. Kao, B. Bi, R. Cheng, and E. Lo, “DQR: A probabilistic approach to diversified query recommendation,” in Proc. 21st ACM Conf. Inf. Knowl. Manage., 2012, pp. 16–25.
9. Y. Song, D. Zhou, and L.-w. He, “Post-ranking query suggestion by diversifying search results,” in Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2011, pp. 815–824.
10. X. Zhu, J. Guo, X. Cheng, P. Du, and H.-W. Shen, “A unified framework for recommending diverse and relevant queries,” in Proc. 20th Int. Conf. World Wide Web, 2011, pp. 37–46.