ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE) Vol. 3, Special Issue 7, March 2017

Position Aware Keyword Query Proposal Based on Document Imminence

V.SARANYA [1] M.MOHANAPRIYA [2]

1. PG Student, Dept.of Computer Applications, VSB Engineering College, Karur.
2. HoD, Dept.of Computer Applications, VSB Engineering College, Karur.

Abstract: 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) is known to be correlated with their spatial proximity to the query issuer. In this paper, 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.

1. 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. 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. 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 search in both coverage and quality of suggestions.

2. LITERATURE SURVEY

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.

Search engines should provide better assistance especially when rare or single-term queries are input, and that they should dynamically provide query suggestions according to the searcher's current state.It will further investigate the usage of query suggestion with data sets including user information to propose a query reformulation taxonomy specifically designed for query suggestion classification, and to improve query suggestion functionality based on our insights.

2. Query recommendation using query logs in search engines,R. Baeza-Yates,

C. Hurtado, and M. Mendoza, in EDBT, 2004, pp.588–596.

A given a query submitted to a search engine, suggests a list of related queries. The related queries are based in previously issued queries, and can be issued by the user to the search engine to tune or redirect the search process. It will further improve the notion of interest of the suggested queries and to develop other notions of interest for the query recommender system. For example, finding queries that share words but not clicked URL's.

3. Agglomerative clustering of a searchengine query log,D. Beeferman and A. Berger, in KDD, 2000, pp. 407–416.

It introduces a technique for mining a collection of user transactions with an Internet search engine to discover clusters of similar queries and similar URLs. The information we exploit is "clickthrough data": each record consists of a user's query to a search engine along with the URL which the user selected from among the candidates offered by the search engine. It is not resolved by the work is how best to combaine the complementary strategies of content ignorant and content-aware clustring. Each method has weakness.

4. Location Aware Keyword Query Suggestion based on Document Proximity" Shuyao Qi, Dingming Wu, and Nikos Mamoulis, in IEEE,2015,pp.82-97.

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.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.

2.1 EXISTING SYSTEM

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

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE) Vol. 3, Special Issue 7, March 2017

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.

Disadvantages of Existing System:

1. Existing keyword suggestion techniques do not consider the locations of the keyword and the query results

2.2 PROPOSED SYSTEM

In this paper, we 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. We extend the state-of-the-art Bookmark Coloring Algorithm (BCA) for random walk with restart (RWR) search to compute the location-aware suggestions. In addition, we propose a partition-based algorithm (PA) that greatly reduces the computational cost of BCA. We conduct an empirical study that demonstrates the usefulness of location-aware keyword query suggestion. We also show experimentally that PA is two times to one order of magnitude faster than BCA.

Advantages of Proposed System:

1. The proposed framework can offer useful suggestions and that PA outperforms the baseline algorithm significantly.
2. Reduce the Computational cost by using Partition-based algorithm

3. SYSTEM DESCRIPTION

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 emerges 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 or geo-documents (i.e., documents associated

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. 3, Special Issue 7, March 2017

with geo-locations). Furthermore, 53 percent of Bing's mobile searches in 2011 have a local intent.In this system, the first Location aware Keyword query Suggestion framework is proposed. It illustrates the benefit of LKS using a toy example. Consider five geo-documents d1-d5 as listed in Fig. 1a. Each document di is associated with a location di:as shown in Fig.1b. Assume that a user issues a keyword query kq ¼ "seafood" at location _q, shown in Fig. 1b. Note that the relevant documents d1–d3 (containing "seafood") are far from _q. A location aware suggestion is "lobster", which can retrieve nearby documents d4 and d5 that are also relevant to the user's original search intention. Previous keyword query suggestion models ignore the user location and would suggest "fish", which again fails to retrieve nearby relevant documents. LKS has a different goal and therefore differs from other location-aware recommendation methods.

d_1	Fish and Seafood
d_2	Fish Seafood
d_3	Lobster Seafood
d_4	Lobster Restaurant
d_5	Lobster House
k_1	Fish
k_2	Seafood
k_3	Lobster

Fig.1a Documents and Keyword Search

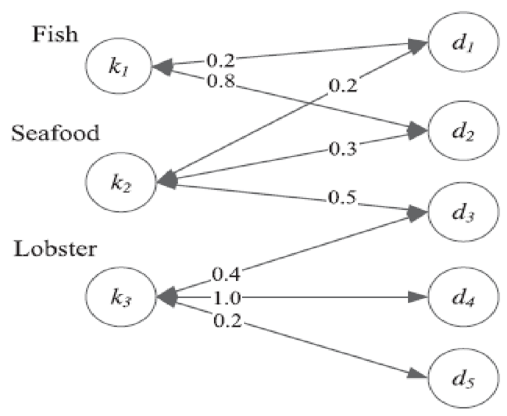


Fig.1b KD Graph

The proposed system extends the popular Bookmark-Coloring Algorithm to compute the random walk with restart (RWR)-based top-m query suggestions as a baseline algorithm. BCA models RWR as a bookmark coloring process. Starting with one unit of active ink injected into node kq, BA processes the nodes in the graph in descending order of their active ink. Different from typical personalized PageRank problems where the graph is homogeneous, KD-graph Gq has two types of nodes: keyword query nodes and document nodes. As opposed to BCA, BA only ranks keyword query nodes; a keyword query node retains a portion of its active ink and distributes 1-λ portion to its neighbor nodes based on its outgoing adjusted edge weights, while a document

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE) Vol. 3, Special Issue 7, March 2017

node distributes all its active ink to its neighbor nodes.

Keyword 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.

String Tokenization

Divide a string of characters into linguistically salient units. In .Net, there is build in string tokenizer class that allows an application to break a string into tokens. For example, string "Anna University of Tamilnadu" can be broken into:

Anna, University, of, Tamilnadu four token, i.e. four single words.

With this string tokenization technique, the pages indexed by web crawler can be partitioned into tokens. This method is essential for the creating inverted index and then for ranking.

Inverted Index

A sequence of (key, pointer) pairs where each pointer points to a record in a database which contains the key value in some particular field. The index is sorted on the key values to allow rapid searching for a particular key value, using e.g. binary search. The index is "inverted" in the sense that the key value is used to find the record rather than the other way round. For databases in which the records may be searched based on more than one field, multiple indices may be created that are sorted on those keys. An index may contain gaps to allow for new entries to be added in the correct sort order without always requiring the following entries to be shifted out of the way.

Ranking

Ranking is the theory that calculates importance of each word in a page. It is the theory that keyword search based on.

Theory of ranking

The following rules will affect the rank number of a keyword in a

(importance of each rule is ordered in descending order):

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE) Vol. 3, Special Issue 7, March 2017

Whether the

1. file name include the keyword.

2. title tag include the keyword.

3. meta tag include the keyword.

4. keyword exist in the first paragraph.

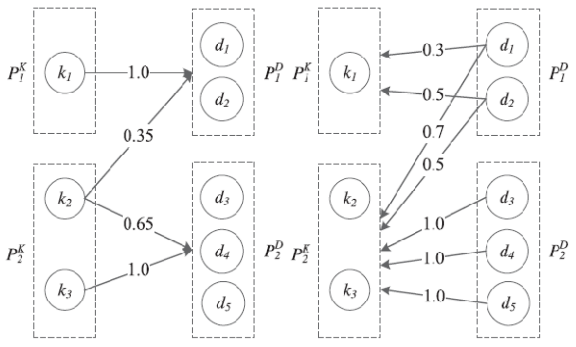
5. keyword exist in the last paragraph.

6. keyword exist in the body paragraph.

The above rule can be expressed in a term *prominence,* that is, how close to the beginning of the sentence the keyword is found. Besides prominence, the *frequency* of the keyword appear in the page is also very important. e.g. a page with the keyword appear 10 times (page A) may get higher rank(rank number) than a page with the keyword only appear 4 times(page B). However, it is not always true when prominence of page B is much better than page A. Frequency of keyword appear in a page is not fair when the size (no of words) of pages a not equal. For example, a page with 10000 words is probably having higher frequency than a page with only 1000 words. In turn, we have a better theory replace just counting the frequency. It is called *weight*.

Weight: = No. of keyword (frequency)/ No. of total words in the page.

PA uses two directed graphs GKP and GDP constructed offline from the KD-graph G and partitions PK and PD. In graph GKP , a keyword query node ki connects to a document partition PD if ki connects in G to at least one document in PD. Similarly, in graph GDP, a document node dj connects to a keyword partition PK if dj connects in G to at least one keyword query node ki. As an example, in the below figure, the document partitions are PD1 ¼ fd1; d2g and PD2 ¼ fd3; d4; d5g and the keyword query partitions are PK1 ¼ fk1g and PK2 ¼ fk2; k3g. The edge weights are defined based on graph Gq, computed during the execution of PA. Each edge weight shown in Fig. indicates the portion of the ink to be distributed to a partition P from a node v that is the sum of the adjusted weights of the edges from node v to the nodes in P according to Gq.



(a) Keywords to Partitions (b) Documents to Partitions

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE) Vol. 3, Special Issue 7, March 2017

3.1 MODULE DESCRIPTION

In this project, there are 3 main modules;

- KD-Graph Construction Module
- Partition Algorithm Module
- Selecting keyword Query Suggestion Module

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, it can be improved 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.

4. CONCLUSION

This work proposed an LKS framework providing keyword suggestions that are relevant to the user information needs and at the same time can retrieve relevant documents based on the keyword location. A baseline algorithm extended from algorithm BCA is introduced to solve the problem. Then, it aproposed 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.

FUTURE WORK

In the future, the effectiveness of the LKS framework can be studied by collecting more data and designing a benchmark. In

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. 3, Special Issue 7, March 2017

addition, subject to the availability of data, it will adapt and test LKS for the case where the locations of the query issuers are available in the query log. Finally, PA can also be applied to accelerate RWR on general graphs with dynamic edge weights; this potential can be investigated in the future.

REFERENCES

[1] R. Baeza-Yates, C. Hurtado, and M. Mendoza, "Query recommendation using query logs in search engines," in Proc. Int. Conf. Current Trends Database Technol., 2004, pp. 588–596.

[2] D. Beeferman and A. Berger, "Agglomerative clustering of a search engine query log," in Proc. 6th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2000, pp. 407–416.

[3] 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.

[4] 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.

[5] Q. Mei, D. Zhou, and K. Church, "Query suggestion using hitting time," in Proc. 17th ACM Conf. Inf. Knowl. Manage., 2008, pp. 469–478.

[6] Y. Song and L.-W. He, "Optimal rare query suggestion with implicit user feedback," in Proc. 19th Int. Conf. World Wide Web, 2010, pp. 901–910.

[7] T. Miyanishi and T. Sakai, "Time-aware structured query suggestion," in Proc. 36th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2013, pp. 809–812.

[8] A. Anagnostopoulos, L. Becchetti, C. Castillo, and A. Gionis, "An optimization framework for query recommendation," in Proc. ACM Int. Conf. Web Search Data Mining, 2010, pp. 161–170.

[9] P. Boldi, F. Bonchi, C. Castillo, D. Donato, A. Gionis, and S. Vigna, "The query-flow graph: Model and applications," in Proc. 17th ACM Conf. Inf. Knowl. Manage., 2008, pp. 609–618.

[10] Y. Song, D. Zhou, and L.-w. He, "Query suggestion by constructing term-transition graphs," in Proc. 5th ACM Int. Conf. Web Search Data Mining, 2012, pp. 353–362.

[11] L. Li, G. Xu, Z. Yang, P. Dolog, Y. Zhang, and M. Kitsuregawa, "An efficient approach to suggesting topically related web queries using hidden topic model," World Wide Web, vol. 16, pp. 273–297, 2013.

[12] D. Wu, M. L. Yiu, and C. S. Jensen, "Moving spatial keyword queries: Formulation, methods, and analysis," ACM Trans. Database Syst., vol. 38, no. 1, pp. 7:1–7:47, 2013.