PAPER ON BEST KEYWORD COVER SEARCH USING QUERY PROCESSING

Ghodake Shamli [1], Gurav Dipali [2], Gonge Anuja [3], Prof. A.N.Nawathe [4]

[1]*BE Student, Department of Computer Engineering, Amrutvahini COE, Sangamner, and Maharashtra, India.*
[2]*BE Student, Department of Computer Engineering, Amrutvahini COE, Sangamner, and Maharashtra, India.*
[3]*BE Student, Department of Computer Engineering, Amrutvahini COE, Sangamner, and Maharashtra, India.*
[4]*Assistant Professor, Department of Computer Engineering, AVCOE, Sangamner, and Maharashtra, India.*

ABSTRACT

It is very common that the objects in a structural database (e.g., restaurants/hotels) are linked with keyword(s) to show their businesses/services/features. An interesting problem called as Closest Keywords search is to query objects, called keyword cover, which together cover a set of query keywords and have the minimum inter-objects distance. In recent few years, we observe the increasing availability and importance of keyword rating in object evaluation for the better decision making. This motivates and show us to detailed investigate a generic version of Closest Keywords search called Best Keyword Cover search which considers inter-objects distance as well as the keyword rating of objects. The baseline algorithm is I truly inspired by the methods of Closest Keywords search which is based on exhaustively combining objects from different query keywords to generate and indicate candidate keyword covers. When the number of query keywords increases rapidly, the performance of the baseline algorithm drops dramatically as a result of massive candidate keyword covers generated. To attack this drawback, this work proposes a much more scalable algorithm called keyword nearest neighbour expansion (keyword-NNE). Compared to the base line algorithm, keyword-NNE algorithm significantly reduces the number of candidate keyword covers generated. The in-depth analysis and extensive experiments on real data sets have justified the superiority of our keyword-NNE algorithm.

Keywords — *Spatial database, point of interests, keywords, keyword rating, keyword cover*

1. INTRODUCTION

In a spatial database, each tuple represents a spatial object which is associated with keyword(s) to indicate the information such as its businesses/services/features. Given a set of query keywords, an essential task of spatial keywords search is to identify spatial object(s) which are associated with keywords relevant to a set of query keywords, and have desirable spatial relationships (e.g., close to each other and/or close to a query location). This problem has unique value in various applications because users' requirements are often expressed as multiple keywords. For example, a tourist who plans to visit a city may have particular shopping, dining and accommodation needs. It is desirable that all these needs can be satisfied without long distance travelling.

2. EXISTING SYSTEM

- Some existing works focus on retrieving individual objects by specifying a query consisting of a query location and a set of query keywords (or known as document in some context). Each retrieved object is associated with keywords relevant to the query keywords and is close to the query location.
- The approaches proposed by Cong et al. and Li et al. employ a hybrid index that augments nodes in non-leaf nodes of an R/R*-tree with inverted indexes.
- In virtual bR*-tree based method, an R*-tree is used to index locations of objects and an inverted index is used to label the leaf nodes in the R*-tree associated with each keyword. Since only leaf nodes have keyword information the mCK query is processed by browsing index bottom-up.

3. PROPOSED SYSTEM

- Research investigates a generic version of mCK query, called Best Keyword Cover (BKC) query, which considers inter-objects distance as well as keyword rating. It is motivated by the observation of increasing availability and importance of keyword rating in decision making. Millions of businesses/services/features around the world have been rated by customers through online business review sites such as Yelp, City search, ZAGAT and Dianping, etc.
- Some work develops two BKC query processing algorithms, baseline and keyword-NNE. The baseline algorithm is inspired by the mCK query processing methods. Both the baseline algorithm and keyword-NNE algorithm are supported by indexing the objects with an R*-tree like index, called KRR*-tree.
- We developed much scalable keyword nearest neighbor expansion (keyword-NNE) algorithm which applies a different strategy. Keyword-NNE selects one query keyword as principal query keyword. The objects associated with the principal query keyword are principal objects. For each principal object, the local best solution (known as local best keyword cover lbkc) is computed. Among them, the lbkc with the highest evaluation is the solution of BKC query. Given a principal object, its lbkc can be identified by simply retrieving a few nearby and highly rated objects in each non-principal query keyword (two-four objects in average as illustrated in experiments).

4. SYSTEM ARCHITECTURE

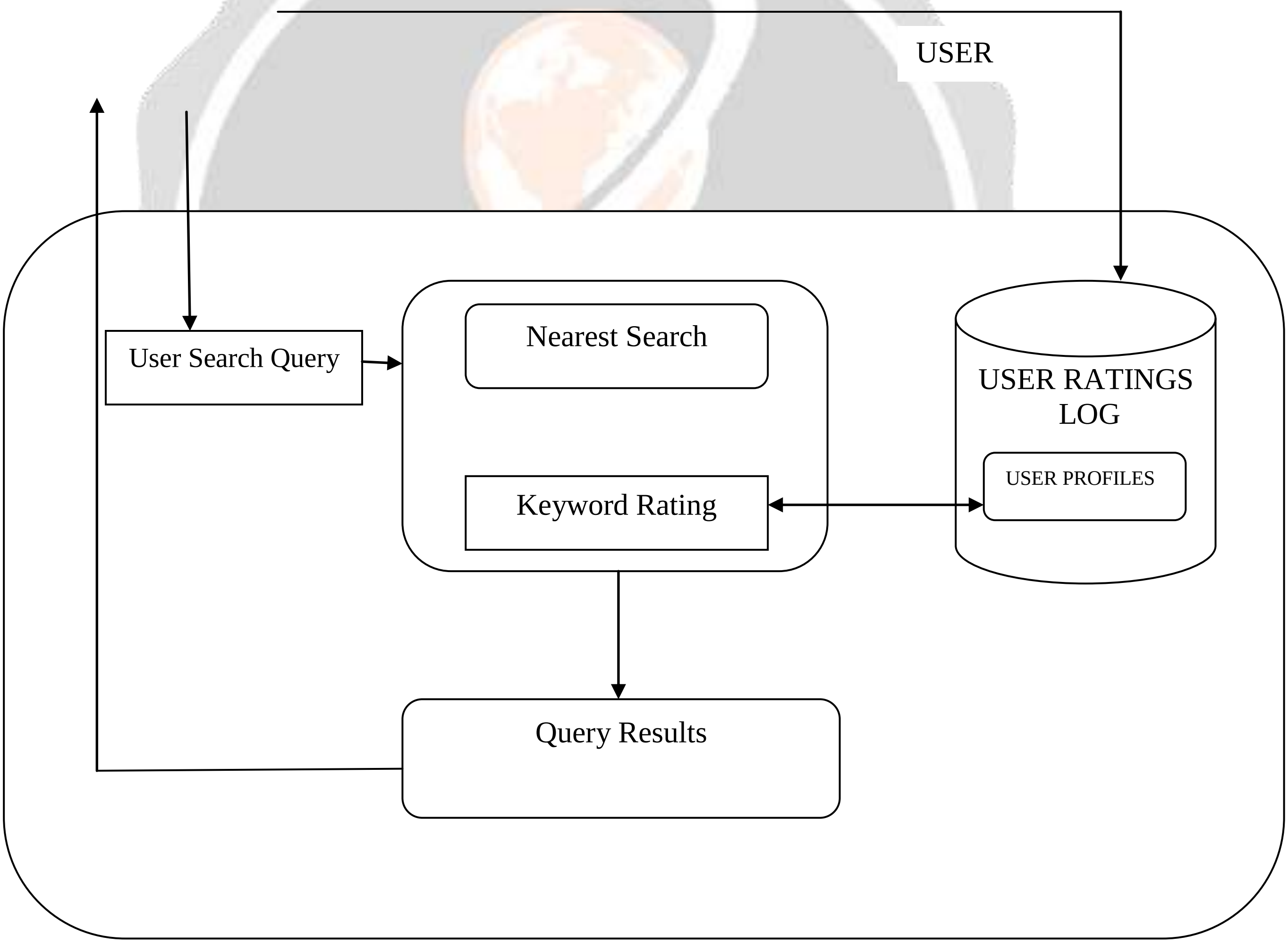

Fig: System Architecture for keyword cover search

In this architecture, a query consisting of a query location and a set of query keywords. Each retrieved object is associated with keywords relevant to the query keywords and is close to the query location.

5. SYSTEM MODULES

1. Spatial database,
2. Point of interests,
3. Keywords,
4. Keyword rating,
5. Keyword cover.

6. PROPOSED ALGORITHM

6.1 Keyword-NNE algorithm

In this algorithm, one query is considered as a principal query keyword. Those objects are associated with principal query keyword are considered as principal objects. Keyword-NNE computes local best solution for each principal object. BKC algorithm returns the lbkc with having highest evaluation. For each principal object, its lbkc can be simply selects few nearby and highly rated objects by the viewer/customer. Compared with the baseline algorithm, the keyword covers significantly reduced. These keyword covers further processed in keyword-NNE algorithm that will be optimal, and each keyword candidate cover processing generates very less new candidate keyword covers.

7. CONCLUSION

Compared to the most relevant mCK query, BKC query provides an additional dimension to support more sensible decision making. The introduced baseline algorithm is inspired by the methods for processing CK query. The baseline algorithm generates a large number of candidate keyword covers which leads to dramatic performance drop when more query keywords are given. The proposed keyword-NNE algorithm applies a different processing strategy, i.e., searching local best solution for each object in a certain query keyword.

REFERENCES

[1]R. Agrawal and R. Srikant, “Fast algorithms for mining association rules in large databases,” in Proc. 20th Int. Conf. Very Large Data Bases, 1994, pp.487–499.

[2] T. Brinkhoff, H. Kriegel, and B. Seeger, “Efficient processing of spatial joins using r-trees,” in Proc. ACM SIGMOD Int. Conf. Manage. Data, 1993, pp.37–246.

[3] X. Cao, G. Cong, and C. Jensen, “Retrieving top-k prestige-based relevant spatial web objects,” Proc. VLDB Endowment, vol. 3, nos. 1/2, pp. 373–384, Sep. 2010.

[4] X. Cao, G. Cong, C. Jensen, and B. Ooi, “Collective spatial keyword querying,” in Proc. ACM SIGMOD Int. Conf. Manage. Data,2011, pp. 373–384.

[5] G. Cong, C. Jensen, and D. Wu, “Efficient retrieval of the top-k most relevant spatial web objects,” Proc. VLDB Endowment, vol. 2, no. 1, pp. 337–348, Aug. 2009.

[6] R. Fagin, A. Lotem, and M. Naor, “Optimal aggregation algorithms for middleware,” J. Comput. Syst. Sci., vol. 66, pp. 614–656, 2003.

[7] I. D. Felipe, V. Hristidis, and N. Rishe, “Keyword search on spatial databases,” in Proc. IEEE 24th Int. Conf. Data Eng.,2008, pp. 656–6