ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

A Survey on Location Based Nearest Keyword Search

Rachana V. Kurhekar, Prof. R. R. Shelke

ME (CSE), Dept. of CSE, HVPM's College of Engineering & Technology, SGBAU University, Amravati, India

Asst. Professor, Dept. of CSE, HVPM's College of Engineering & Technology, SGBAU University, Amravati, India

ABSTRACT : It is common that the objects in a spatial database are associated with keywordto indicate their businesses/services/features. An interesting problem known as Closest Keywords search is to query objects, called nearest keyword search , which together cover a set of query keywords and have the minimum inter-objects distance. In recent years, I observe the increasing availability and importance of keyword rating in object evaluation for the better decision making. This motivates us to investigate a generic version of Closest Keywords search called Best Keyword Cover which considers inter-objects distance as well as the keyword rating of objects. The baseline algorithm is inspired by the methods of Closest Keywords search which is based on exhaustively combining objects from different query keywords to generate candidate keyword covers. When the number of query keywords increases, the performance of the baseline algorithm drops dramatically as a result of massive candidate keyword covers generated. To recover this drawback, this work proposes a much more scalable algorithm called keyword nearest neighbor expansion (keyword-NNE). Compared to the baseline 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

I. INTRODUCTION

An increasing number of applications require the efficient execution of nearest neighbor (NN) queries constrained by the properties of the spatial objects. Due to the popularity of keyword search, particularly on the Internet, many of these applications allow the user to provide a list of keywords that the spatial objects (henceforth referred to simply as objects) should contain, in their description or other attribute. For example, online yellow pages allow users to specify an address and a set of keywords, and return businesses whose description contains these keywords, ordered by their distance to the specified address location. As another example, real estate web sites allow users to search for properties with specific keywords in their description and rank them according to their distance from a specified location. We call such queries spatial keyword queries. A spatial keyword query consists of a query area and a set of keywords. The answer is a list of objects ranked according to a combination of their distance to the query area and the relevance of their text description to the query keywords. A simple yet popular variant, which is used in our running example, is the distance-first spatial keyword query, where objects are ranked by distance and keywords are applied as a conjunctive filter to eliminate objects that do not contain them. Which is our running example, displays a dataset of fictitious hotels with their spatial coordinates and a set of descriptive attributes (name, amenities)? An example of a spatial keyword query is "find the nearest hotels to point that contain keywords internet and pool". The top result of this query is the hotel object. Unfortunately there is no efficient support for top-k spatial keyword queries, where a prefix of the results list is required. Instead, current systems use ad-hoc combinations of nearest neighbor (NN) and keyword search techniques to tackle the problem. For instance, an R-Tree is used to find the nearest neighbors and for each neighbor an inverted index is used to check if the query keywords are contained. We show that such two-phase approaches are inefficient.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

II. LITERATURE REVIEW & RELATED WORK

The Baseline Algorithm

According to Dongxiang Zhang, The baseline algorithm is inspired by the mCK query processing methods . For mCK query processing, the method in browses index in top-down manner while the method in does bottom-up. Given the same hierarchical index structure, the top-down browsing manner typically performs better than the bottom-up since the search in lower hierarchical levels is always guided by the search result in the higher hierarchical levels. However, the significant advantage of the method in over the method in has been reported. This is because of the different index structures applied. Both of them use a single tree structure to index data objects of different keywords. But the number of nodes of the index in has been greatly reduced to save I/O cost by keeping keyword information with inverted index separately. Since only leaf nodes and their keyword information are maintained in the inverted index, the bottom-up index browsing manner is used. When designing the baseline algorithm for BKC query processing, we take the advantages of both methods. First, we apply multiple KRR*-trees which contain no keyword information in nodes such that the number of nodes of the index is not more than that of the index in ; second, the top-down index browsing method can be applied since each keyword has own index.[5,6].

Keyword Nearest Neighbor Expansion
(Keyword-NNE)

Using the baseline algorithm, BKC query can be effectively resolved. However, it is based on exhaustively techniques have been explored, it has been observed that the performance drops dramatically, when the number of query keywords increases, because of the fast increase of candidate keyword covers generated. This motivates us to develop a different algorithm called keyword nearest neighbor expansion (keyword-NNE). We focus on a particular query keyword, called principal query keyword. The objects associated with the principal query keyword are called principal objects.[7]

III. GOALS AND SCOPE

Goals: The goal is to rank the methods, so we only report here on the binary comparisons that allowed us to determine the ordering of the four methods (excluding redundant comparisons).Our current goals are to allow explicit queries, and to rank document results with the objective of maximizing the coverage of all the in the spatial database, while minimizing redundancy in a short list of the best keyword search. A keyword cover of keyword that is the word related to that keyword, and cover keyword is called to be the best keyword for the search find's valuable search and ranking, without interrupting the conversation flow, thus ensuring the usability of our system. In the future, this will be tested with human users of the system within real-life meetings.

Scope: Our treatment of nearest neighbour search falls in the general topic of spatial keyword search, which has also given rise to several alternative problems. A complete survey of all those problems goes beyond the scope of this paper. Below we mention several representatives, but interested readers can refer to for a nice survey Specifically, aiming at an IR flavor, the approach of computes the relevance between the documents of an object p and a query q. This relevance score is then integrated with the Euclidean distance between p and q to calculate an overall similarity of p to q. The few objects with the highest similarity are returned. In this way, an object may still be in the query result, even though its document does not contain all the query keywords.

IV. PROPOSED SYSTEM

- This paper investigates a generic version of mCKquery, 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

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

around the world have been rated by customers through online business review sites such as Yelp, Citysearch, ZAGAT and Dianping, etc.

- This 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 aslocal best keyword cover lbkc) is computed. Among them, the lbkc with the highest evaluationis 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).

Advantages Of Proposed System:

- Compared to the baseline algorithm, the number of candidate keyword covers generated in keyword-NNE algorithm is significantly reduced. The indepth analysis reveals that the number of candidate keyword covers further processed in keyword-NNE algorithm is optimal, and each keyword candidate cover processing generates much lessnew candidate keyword covers than that in the baseline algorithm.

- The proposed keyword-NNE algorithm applies a different processing strategy, i.e., searching local best solution for each object in a certain query keyword. As a consequence, the number of candidate keyword covers generated is significantly reduced.

- The analysis reveals that the number of candidate keyword covers which need to be further processed in keyword-NNE algorithm is optimal and processing each keyword candidate covertypically generates much less new candidate keyword covers in keyword-NNE algorithm than in the baseline algorithm.

V. SYSTEM ARCHITECTURE

Step 1. Select a keyword from query as the principal_query_ keyword;
Step 2. Select the objects associated with principal_query_ keyword as principal_objects.
Step 3. For every principal_objects, local_best_solution is computed.
Step 4. Identify the Global_best _solution from step3;
Step 4. Return the solution as Global_best_solution

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

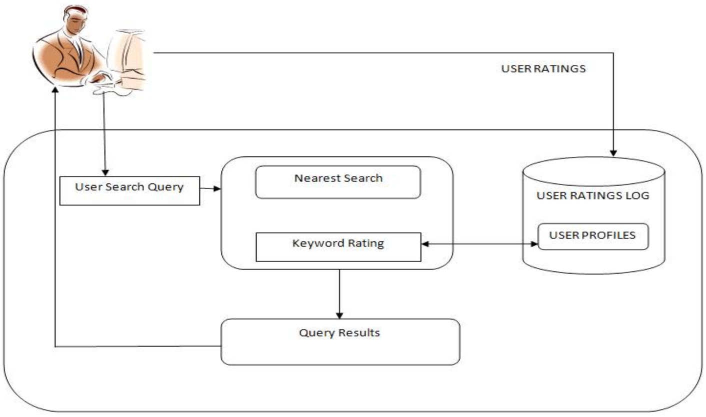

Figure 1: Structre of the proposed system

VI. CONCLUSIONS

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 mCK 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. As a consequence, the number of candidate keyword covers generated is significantly reduced. The analysis reveals that the number of candidate keyword covers which need to be further processed in keyword-NNE algorithm is optimal and processing each keyword candidate cover typically generates much less new candidate keyword covers in keyword-NNE algorithm than in the baseline algorithm.

REFERENCES

[1] G. Cong, C. Jensen, and D. Wu. “Efficient retrieval of the top-k most relevant spatial web objects”. In: Proc. VLDB Endow. 2.1 (2009), pp. 337–348.
[2] Ian De Felipe, Vagelis Hristidis, and Naphtali Rishe. “Keyword Search on Spatial Databases”. In: ICDE. 2008, pp. 656–665.
[3] Ramaswamy Hariharan et al. “Processing Spatial- Keyword (SK) Queries in Geographic Information Retrieval (GIR) Systems”. In: Proceedings of the 19th International Conference on Scientific and Statistical Database Management. 2007, pp. 16–23.
[4] Jo˜ao B. Rocha-Junior et al. “Efficient processing of top-k spatial keyword queries”. In: Proceedings of the 12th international conference on Advances in spatial and temporal databases. 2011, pp. 205–222.
[5] Dongxiang Zhang, Beng Chin Ooi, and Anthony K. H. Tung. “Locating mapped resources in web 2.0”. In: ICDE (2010).
[6] Dongxiang Zhang et al. “Keyword Search in Spatial Databases: Towards Searching by Document”. In: ICDE. 2009, pp. 688–699.
[7] Ian De Felipe, Vagelis Hristidis, and Naphtali Rishe. “Keyword Search on Spatial Databases”. In: ICDE. 2008, pp. 656–665.
[8] Ke Deng, Xin Li, Jiaheng Lu, and Xiaofang Zhou.” Best Keyword Cover Search” TKDE.2014
[9] S. B. Roy and K. Chakrabarti. “Location-Aware Type Ahead Search on Spatial Databases: Semantics and Efficiency”. In: SIGMOD (2011)