A Survey on Best Keyword Cover Search

Vishal Kolekar[1], Prof. Ajay K. Gupta[2]

M.E. Student, Dept. of Computer Engineering, IOK, Pune, Maharashtra, India[1]

Assistant Professor, Dept. of Computer Engineering, IOK, Pune, Maharashtra, India[2]

ABSTRACT: Spatial databases are stores the information about the spatial objects which are associated with the keywords to show the information such as its business/services/features. Very important problem known as closest keywords search is to query objects, called keyword cover. In nearest keyword search, it covers a set of query keywords and minimum distance between objects. From last few years, keyword rating increases its availability and importance in object evaluation for the decision making. This is the main reason for developing this new algorithm called Best keyword cover which is considers inter distance as well as the rating provided by the customers through the online business review sites. Closest keyword search algorithm combines the objects from various query keywords toa generate candidate keyword cover. Two algorithms k-means clustering and keyword nearest-neighbor expansion algorithms are used to finding best keyword cover. K-means clustering algorithms are used to find out the similarity of different classes. The performance of the closest keyword algorithm drops dramatically, when the number of query keyword increases.

KEYWORDS: Spatial -database, point of the interests, query keywords, keyword ratings, key-word cover.

I. INTRODUCTION

Now a days, use of mobile computing increases. Inspired by the mobile computing, the spatial keywords search problem has attracted much attention recently because of location-based services and wide availability of extensive digital maps and satellite imagery.So the number of users using the location based services has been also increased to large extend. Spatial objects indicates the information such as its business/services/features which are associated to keyword(s). In spatial database, each tuple represents a spatial object. The main idea behind the spatial keywords search is to identify spatial object(s) which are associated with keywords relevant to a set of query keywords which are close to each other and/or close to the query location. This problem has unique value in various applications because users" requirements are often expressed as multiple keywords. In existing, spatial keyword search problem have been studied because of the value of the special keyword search in practice.. In this project k-means clustering algorithms is used to find the keyword. This paper 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 Damping, etc. For example, a restaurant is rated 65 out of 100 (ZAGAT.com) and a hotel is rated 3.9 out of 5 (hotels.com).

 According to a survey in 2013 conducted by Dimensional Research (dimensionalresearch.com), an overwhelming 90 percent of respondents claimed that buying decisions are influenced by online business review/rating. Due to the consideration of keyword rating, the solution of BKC query can be very different from that of mCK query).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. In the baseline algorithm, the idea is to combine nodes in higher hierarchical levels of KRR*-trees to generate candidate keyword covers. Then, the most promising candidate is assessed in priority by combining their child nodes to generate new candidates. Even though BKC query can be effectively resolved, when the number of query keywords increases, the performance drops dramatically as a result of massive candidate keyword covers generated. To overcome this critical drawback, we developed much scalable keyword nearest neighbour 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.The aim is always to find out multiple objects which altogether cover the, query keyword. this is because no one object individually satisfies all query keywords. The approach studied in is similar to the mKC queries. The m closest keyword cover (mKC)aims to find the objects which are close to each other or with minimum distance. In proposed system we are going to consider not only the inner object distance but as well as the keyword rating of the object

II. RELATED WORK

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. Due to the remarkable value in practice, several variants of spatial keyword search problem have been studied. The works aim to find a number of individual objects, each of which is close to a query location and the associated keywords (or called document) are very relevant to a set of query keywords (or called query document).

1. IRTree: An efficient index for geographic document search [1] From This Paper we Referred-

Given a geographic query that is composed of query keywords and a location, a geographic search engine retrieves documents that are the most textually and spatially relevant to the query keywords and the location, respectively, and ranks there trieved documents according to their joint textual and spatial relevance's to the query. The lack of an efficient index that can simultaneously handle both the textual and spatial aspects of the documents makes existing geographic search engines inefficient in answering geographic queries. In this paper, we propose an efficient index, called IR-tree, that together with a top-k document search algorithm facilitates four major tasks in document searches, namely,1) spatial filtering, 2) textual filtering, 3) relevance computation, and 4) document ranking in a fully integrated manner. In addition, IR - tree allows searches to adopt different weights on textual and spatial relevance of documents at the runtime and thus caters for a wide variety of applications. A set of comprehensive experiments over a wide range of scenarios has been conducted and the experiment results demonstrate that IR-tree outperforms the state-of-the art approaches for geographic document searches.

2. Retrieving top-k prestige-based relevant spatial web objects [2] From This Paper we Referred-

The location aware keyword query returns ranked objects that are near a query location and that have textual descriptions that match query keywords. This query occurs inherently in many types of mobile and traditional web services and applications, e.g., Yellow Pages and Maps services. Previous work considers the potential results of such a query as being independent when ranking them. However, a relevant result object with nearby objects that are also relevant to the query is likely to be preferable over a relevant object without relevant nearby objects. The paper proposes the concept of prestige based relevance to capture both the textual relevance of an object to a query and the effects of nearby objects. Based on this, a new type of query, the Location -aware top-k Prestige-based Text retrieval (LkPT) query, is proposed that retrieves the top-k spatial web objects ranked according to both prestige-based relevance and location proximity. We propose two algorithms that compute LkPT queries. Empirical studies with real - world spatial data demonstrate that LkPT queries are more effective in retrieving web objects than a previous approach that does not consider the effects of nearby objects; and they show that the proposed algorithms are scalable and out Perform baseline approach significantly.

3. Efficient retrieval of the top-k most relevant spatial web objects [3] From This Paper we Referred-

The conventional Internet is acquiring a geospatial dimension. Web documents are being geo-tagged, and geo referenced objects such as points of interest are being associated with descriptive text documents. The resulting fusion of geo-location and documents enables a new kind of to p-k query that takes into account both location proximity and text relevancy. To our knowledge, only naïve techniques exist that is capable of computing a general web information retrieval query while also taking location into account. This paper proposes a new indexing framework for location aware top-k text retrieval. The framework leverages the inverted file for text retrieval and the R-tree for spatial proximity querying. Several indexing approaches are explored within the framework. The framework encompasses algorithms that utilize the proposed indexes for computing the top-k query, thus taking into accounts both text

relevancy and location proximity to prune the search space. Results of empirical studies with an implementation of the framework demonstrate that the paper's proposal offers scalability and is capable of excellent performance.

4. Location-aware type ahead search on spatial databases: emetics and efficiency [4] From This Paper we Referred-

Users often search spatial databases like yellow page data using keywords to and businesses near their current location. Such searches are increasingly being performed from mobile devices. Typing the entire query is cumbersome and prone to errors, especially from mobile phones. We address this problem by introducing type ahead search functionality on spatial databases. Like keyword search on spatial data, type-ahead search needs to be location aware, i.e. with every letter being typed, it needs to return spatial objects whose names (or descriptions) are valid completions of the query string typed so far, and which rank highest in terms of proximity to the user's location and other static scores. Existing solutions for type-ahead search cannot be used directly as they are not location aware. We show that a straight- forward combination of existing techniques for performing type-ahead search with those for performing proximity search perform poorly. We propose a formal model for query processing cost and develop novel techniques that optimize that cost. Our empirical evaluations on real and synthetic datasets demonstrate the effectiveness of our techniques. To the best of our knowledge, this is the first work on location aware type ahead search.

Goals and Objective:

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 shortlist 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 these arch 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

III. PROPOSED ALGORITHM

A. *KEYWORD-NNE:*

In previous work, BKC algorithm drops its performance when the number of query keywords is increases. To solve this problem, here developed a more efficient keyword nearest neighbour expansion (keyword-NNE) which uses the different strategy. 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 of the principal object, its lbkc can be simply selects few closest and highly rated objects by the viewer/customer. Compared with the k-means clustering, the keyword covers significantly reduced. These keyword covers a further processes in keyword-NNE-algorithm that will be optimal, and each keyword candidate covers processed generates very less new candidate keyword are covers.

B. K-MEANS:

Let X = fx1, x2, x3,.., xn g be the set of data points and

V = fv1, v2, vcg be the set of centers.

1) Randomly select 'c' cluster centers.

2) Calculate the distance between each data point and cluster centers.

3) Assign the data point to the cluster canter whose distance from the cluster centre is minimum of all the cluster centres.

4) Recalculate the new cluster canter using:

$V_i = (1/C_i) \sum_{cj=1} X_i$

Where, C_i represents the number of data points in ith cluster.

5) Recalculate the distance between each data point and new obtained cluster centres.

6) If no data point was reassigned then stop, otherwise repeat from step 3).
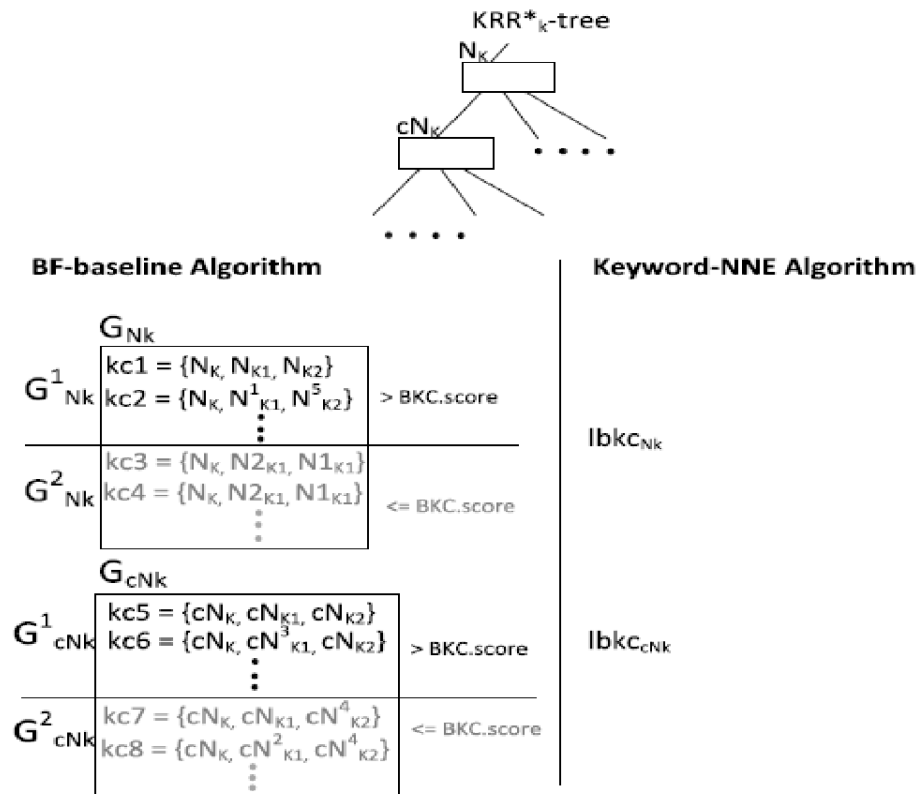
Fig. 1: Baseline & Keyword-(NNE)

IV. SYSTEM ARCHITECTURE
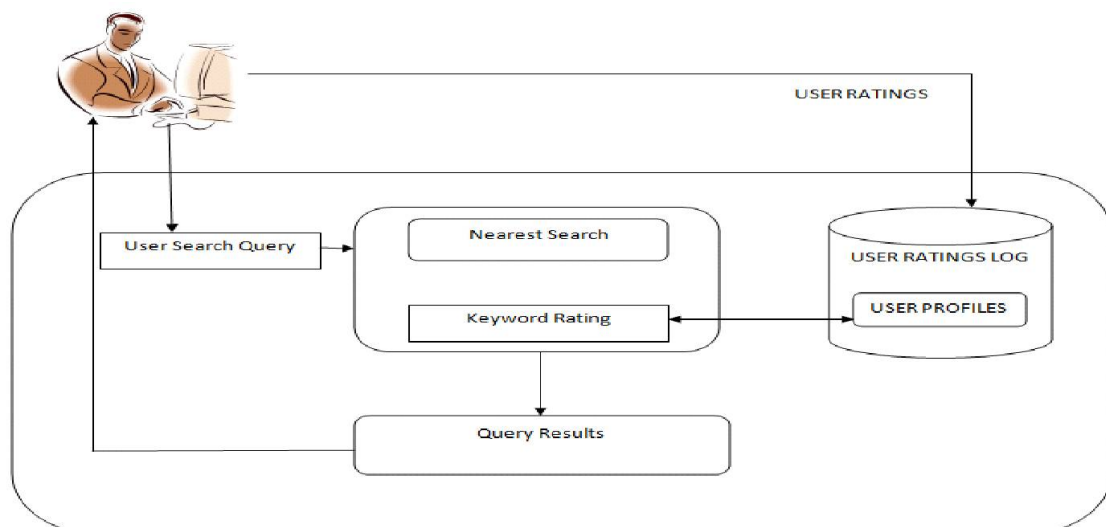
Fig 2. System architecture

- 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.

- This work can be introduced two BKC query processing algorithms, base-line and keyword-(NNE). The base-line algorithm is a inspire by the mCK query processing technique. Both the base-line algorithm and keyword-(NNE) algorithm are supporting by indexing the objects with an R-tree index, called as KRR*-tree.

V. CONCLUSION

Comparing the most of the relevant mCK query, BKC query provides an advanced dimension to support more effective decision making. The introduced base-line algorithm is inspired by the methods for processing mCK query. K-maens clustering algorithms are also used. The baseline algorithm generates a large candidate keyword covers which leads to dramatic performance drop when more query keywords are given. The proposed keyword-(NNE) algorithm applies the various processing schema, i.e., finding local good solution for every object in a certain query keyword. As a consequence, the No. 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 less new candidate keyword covers in keyword-(NNE) algorithm than in the baseline algorithm.

REFERENCES

[1] Ruicheng Zhong, Ju Fan, Guoliang Li, Kian-Lee Tan and Lizhu Zhou, „Location- Aware Instant Search‟ CIKM‟12, October 29 November 2, 2012, Maui, HI, USA

[2] Xin Caoy, Gao Cong, Christian S. Jensenz, ″ Retrieving Top-k Prestige Based Relevant Spatial Web Objects‟, Proceedings of the VLDB Endowment, Vol. 3, No.1 , 2010.

[3] Ian De Felipe, Vagelis Hristidis, Naphtali Rishe, ″ Keyword Search on Spatial Databases‟, in Proc. IEEE 24th Int. Conf. Data Eng., 2008, pp. 656–665.

[4] Z. Li, K. C. Lee, B. Zheng, W.C. Lee, D. Lee, and X. Wang,″ IRTree: An efficient index for geographic document search‟, IEEE Trans. Knowl. Data Eng., vol. 99, no. 4, pp. 585–599, Apr. 2010.

[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] D.Amutha Priya, Dr. T.Manigandan,″ Fast Accurate Mining on Spatial Database Using Keywords‟, International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 3, Special Issue- 1, May 2015.\

[7] Ke Deng, Xin Li, Jiaheng Lu, and Xiaofang Zhou ,″ Best Keyword Cover Search‟, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 1, JANUARY 2015.

[8] X .cao, G Cong, C Jensen "COLLECTIVE SPATIAL KEYWORD QUERYING‟ IN Proc.ACM SIGMOD Int. Conf manage. Data , 2011 ,pp. 373-384.

BIOGRAPHY

Mr. Vishal Kolekar is ME, Student at Department of Computer Engineering in Institute of Knowledge College of Engineering, Pune, and Maharashtra, India. His research interests include Database, Data mining, Data warehousing and Big Data.

Prof. Ajay K. Gupta is an Assistant Professor, at Department of Computer Engineering, Institute of knowledge college of Engineering, Pune, Maharashtra, India. His research interests include Database, Hadoop and Big Data.