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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

A Survey on Best Keyword Cover Search

Ashlesh S. Patole[1;] Shripadrao Biradar[2]

M.E., Dept. of Computer, RMD Sinhgad School of Engineering, Pune, India[1]

Assistant Professor, Dept. of Computer, RMD Sinhgad School of Engineering, Pune, India[2]

ABSTRACT: It is common that the objects in a spatial database (e.g., restaurants/hotels) are associated with keyword(s) to indicate their businesses/services/features. An interesting problem known 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 years, we 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 attack 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, and keyword cover, Inverted Index, Inverted index variants, search engine indexing, postings list.

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.

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

DOI: 10.15680/IJIRCCE.2015. 0311170

needs. It is desirable that all these needs can be satisfied without long distance traveling. 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 retrievesdocuments that are the most textually and spatially relevant to the query keywords and the location, respectively, and ranks theretrieved documents according to their joint textual and spatial relevance's to the query. The lack of an efficient index that cansimultaneously handle both the textual and spatial aspects of the documents makes existing geographic search engines inefficient inanswering geographic queries. In this paper, we propose an efficient index, called IR-tree, that together with a top-k document searchalgorithm 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 experimentsover a wide range of scenarios has been conducted and the experiment results demonstrate that IR-tree outperforms the state-of-theart 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 previousapproach that does not consider the effects of nearby objects; andthey show that the proposed algorithms are scalable and outperforma 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 geo-spatial 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 top-k query that takes into account both location proximity and text relevancy. To our knowledge, only naive 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 account 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 rst work on location-aware type-ahead search.

5. Locating mapped resources in web 2.0," in Proc. IEEE 26th Int. Conf. Data [5] From This Paper we Referred-

Mapping mashups are emerging Web 2.0 applications in which data objects such as blogs, photos and videos fromdifferent sources are combined and marked in a map using APIsthat are released by online mapping solutions such as Googleand Yahoo Maps. These objects are typically associated witha set of tags capturing the embedded semantic and a set ofcoordinates indicating their geographical locations. Traditionalweb resource searching strategies are not effective in such anenvironment due to the lack of the gazetteer context in the tags.Instead, a better alternative approach is to locate an object by tag matching. However, the number of tags associated with each object is typically small, making it difficult for an object to capture the complete semantics in the query objects. In this paper, we focus on the fundamental application of locating geographical resources and propose an efficient tag centric query processing strategy. In particular, we aim to finds a set of nearest co-located objects which together match the query tags. Given the fact that there could be large number of data objects and tags, we develop an efficient search algorithm that can scale up in terms of the number of objects and tags.Further, to ensure that the results are relevant, we also propose a geographical context sensitive geo-tf-idf ranking mechanism. Our experiments on synthetic data sets demonstrate its scalability while the experiments using the real life data set confirm its practicality.

III.GOALS AND OBJECTIVE

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 ALGORITHM

1. KEYWORD-NNE ALGORITHM

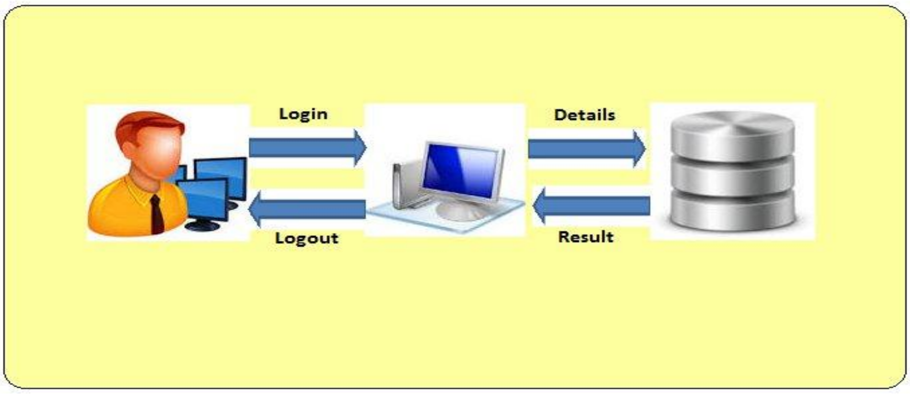
STEP 1. ONE QUERY KEYWORD K 2 T IS SELECTED AS THE PRINCIPAL QUERY KEYWORD;

STEP 2. FOR EACH PRINCIPAL OBJECT OK 2 OK, LBKCOK IS COMPUTED

STEP 3. IN OK, GBKCK IS IDENTIFIED;

STEP 4. RETURN GBKCK.

V.ARCHITECTURE



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. Ke Deng; Xin Li; Jiaheng Lu; Xiaofang Zhou," Best Keyword Cover Search" Knowledge and Data Engineering,IEEETransactions on Year: 2015,

2. 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.

3. 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.

4. S. B. Roy and K. Chakrabarti, "Location-aware type ahead search on spatial databases: Emantics and efficiency," in Proc.ACM SIGMOD Int. Conf. Manage. Data, 2011, pp. 361–37..

5. D. Zhang, B. Ooi, and A. Tung, "Locating mapped resources in web 2.0," in Proc. IEEE 26th Int. Conf. Data Eng., 2010, pp. 521–532.

6. 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.