International Journal of Research
Available at https://edupediapublications.org/journals

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 04 Issue 09
August 2017

A New Approach for Keyword Cover Search

Y.Nagendar, & V.Charan
[1]Assistant Professor, Dept.of CSE, SR Engineering College, Warangal, Telangana, India.
[2]PG Scholar, C.S.E., SR Engineering College,Warangal,Telangana, India.

Abstract: *Spatial databases area unit stores the knowledge concerning the spatial objects that area unit related to the keywords to point the knowledge like its business/services/features. Best keyword cover question aims to search out objects related to keywords. The method planned considers keyword rating, keyword connection and spatial connection. This can be the most reason for developing this new algorithmic program referred to as Best keyword cowl that is considers inter-distance still because the rating provided by the shoppers through the web business review sites. Nearest keyword search algorithm combines the objects from completely different question keywords to generate candidate keyword covers. Baseline algorithmic program and keyword nearest neighbor enlargement algorithms area unit wont to notice the best keyword cowl. The performance of the nearest keyword algorithm drops dramatically, once the amount of question keyword will increase. In comparison with the baseline algorithmic program, keyword-NNE algorithmic program immensely reduces the amount of candidate keyword covers generated.*

Keywords: Spatial Database, Point of Interests, Keywords,Keyword Rating, Keyword Cover.

I. INTRODUCTION

An increasing variety of applications need the economical execution of nearest neighbor (NN) queries strained by the properties of the spatial objects. Attributable to the recognition of keyword search, notably on the net, many of these applications permit the user to supply a listing of keywords that the spatial objects (henceforth noted merely as objects) ought to contain, in their description or alternative attribute. For instance, on-line yellow pages permit users to specify AN address and a group of keywords, and come back businesses whose description contains these keywords, ordered by their distance to the desired address location. We have a tendency to decision such queries spatial keyword queries. A spatial keyword question consists of a question space and a group of keywords. The solution may be a list of objects graded per a mix of their distance to the question space and also the connection of their text description to the question keywords. A straightforward however fashionable variant, that is employed in our running example, is that the distance-first spatial keyword question, where objects ar graded by distance and keywords are applied as a conjunctive filter to eliminate objects that don't contain them. That 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 question is “find the closest hotels to purpose that contain keywords net and pool”. The highest results of this query is that the edifice object. Sadly there's no economical 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. There are easy ways to support queries that combine spatial and text features. For example, for the above query, we could first fetch all the restaurants whose menus contain the set of keywords, and then from the retrieved restaurants, find the nearest one. Similarly, one could also do it reversely by targeting first the spatial conditions – browse all the restaurants in ascending order of their distances to the query point until encountering one whose menu has all the keywords. The major drawback of these straightforward approaches is that they will fail to provide real time answers on difficult inputs. A typical example is that the real nearest neighbor lies quite far away from the query point, while all the closer neighbors are missing at least one of the query keywords. For better decision making, concept of keyword rating was introduced along with its features other than distance. For such search, query will take form of feature of objects. It search for nearest neighbor based on a new similarity measure, named weighted average of index rating which combine keyword rating, keyword search and nearest neighbour search.

II. LITERATURE SURVEY

Ke Deng, Xin Li, Jiaheng Lu, and Xiaofang Zhou, Best Keyword Cover Search[1]

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.

Yufei Tao and Cheng Sheng, Fast Nearest Neighbor Search with Keywords[2]
Conventional spatial queries, such as range search and nearest neighbor retrieval, involve only conditions on objects' geometric properties. Today, many modern applications call for novel forms of queries that aim to find objects satisfying both a spatial predicate, and a predicate on their associated texts. For example, instead of considering all the restaurants, a nearest neighbor query would instead ask for the restaurant that is the closest among those whose menus contain “steak, spaghetti, brandy” all at the same time. Currently, the best solution to such queries is based on the IR ^{2}-tree, which, as shown in this paper, has a few deficiencies that seriously impact its efficiency. Motivated by this, we develop a new access method called the spatial inverted index that extends the conventional inverted index to cope with multidimensional data, and comes with algorithms that can answer nearest neighbor queries with keywords in real time. As verified by experiments, the proposed techniques outperform the IR ^{2}-tree in query response time significantly, often by a factor of orders of magnitude.

Ying Zhang, Wenjie Zhang, Qianlu Lin, Xuemin Lin, Heng Tao Shen, Effectively Indexing the Multidimensional Uncertain Objects[3]
As the uncertainty is inherent in a wide spectrum of applications such as radio frequency identification (RFID) networks and location-based services (LBS), it is highly demanded to address the uncertainty of the objects. In this paper, we propose a novel indexing structure, named U-Quadtree, to organize the uncertain objects in the multidimensional space such that the queries can be processed efficiently by taking advantage of U-Quadtree. Particularly, we focus on the range search on multidimensional uncertain objects since it is a fundamental query in a spatial database. We propose a cost model which carefully considers various factors that may impact the performance. Then, an effective and efficient index construction algorithm is proposed to build the optimal U-Quadtree regarding the cost model. We show that U-Quadtree can also efficiently support other types of queries such as uncertain range query and nearest neighbor query. Comprehensive experiments demonstrate that our techniques outperform the existing works on multidimensional uncertain objects.

X. Cao, G. Cong, C.S. Jensen, and B.C. Ooi, Collective Spatial Keyword Querying[4]
With the proliferation of geo-positioning and geo-tagging, spatial web objects that possess both a geographical location and a textual description are gaining in prevalence, and spatial keyword queries that exploit both location and textual description are gaining in prominence. However, the queries studied so far generally focus on finding individual objects that each satisfy a query rather than finding groups of objects where the objects in a group collectively satisfy a query. We define the problem of retrieving a group of spatial web objects such that the group’s keywords cover the query’s keywords and such that objects are nearest to the query location and have the lowest inter-object distances. Specifically, we study two variants of this problem, both of which are NP-complete. We devise exact solutions as well as approximate solutions with provable approximation bounds to the problems. We present empirical studies that offer insight into the efficiency and accuracy of the solutions.

III. SYSTEM MODEL

This paper investigates a typical variant of mckquery, referred to as Best Keyword cowl (BKC) question, which considers inter-objects distance pretty much as good as keyword ranking. It's inspired through the statement of increasing availableness and importance of key phrase rating in choice creating. Thousands of organizations/offerings/ features round the world have been rated through users by suggests that of

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 04 Issue 09
August 2017

on-line industry assessment sites just like Yelp, City search, ZAGAT and dianping, and so on. This work develops 2 BKC question processing algorithms, baseline and key phrase-NNE. The baseline rule is aroused with the help of the mck query processing ways that. Each the baseline rule and keyword-NNE algorithm area unit supported via indexing the objects with associate degree R*-tree like index,known as KRR*-tree. We have a tendency to developed abundant scalable key word nearestneighbor enlargement (key phrase-NNE) algorithmwhich applies an additional approach. Key word-nne selects one question keyword as predominant question keyword. The objects associated with the important query keyword area unit primary objects. For everyfundamental object, the close nice answer (often referred to as localbest key word quilt lbkc) is computed. Among them, the lbkc with the perfect evaluationis the solution of BKC question. Given a majorobject, its lbkc can even be known by victimisation readilyretrieving a number of native and tremendously rated objects in every and each non-predominant question key phrase (two-4objects in traditional as illustrated in experiments).

A. Indexing Keyword Ratings

A single tree structure is employed to index objects ofextraordinary key terms. The one tree might also be elevated with another dimension to index key phrase rating. One tree constitution fits matters that just about all key words area unit question key words. For the on top of mentioned illustration, all keywords, i.e., “resort”, “restaurant” and “bar”, area unit question keywords. Yet, it's a lot of universal that almost all effective alittle fraction of key terms area unit question keyword phrases. For illustration within the experiments, solely not up to five keywords area unit question key words. During this state of affairs, one tree is dangerous to approximate the spatial relationship between objects of few distinct keywords. Consequently, a few of KRR*- tree area unit used on this work, each for one keyword.1 The KRR*-tree for key word ki is denoted as KRR*kitree. Given associate degree object, the ranking of a connected keyword is sometimes the imply of ratings given by variety of patrons for associate degree interval of your time. The alternate will happen but slowly. Despite the actual fact that dramatic alternate happens, the KRR*-tree is up so far within the general approach of R*-tree replace.

B. Keyword nearest Neighbor enlargement

Using the baseline rule, BKC question will beeffectively resolved. However, it's primarily based onexhaustively combining objects (or their mbrs). Eventhough pruning techniques are explored, it hasbeen discovered that the performance dropsdramatically, once the amount of question keywordsincreases, attributable to the quick increase of candidatekeyword covers generated. This motivates US todevelop a special rule referred to as keyword nearestneighbor enlargement. We have a tendency to specialise in a specific querykeyword, referred to as principal question keyword. The objectsassociated with the principal question keyword area unit calledprincipal objects. The goal of the interface is toprovide purpose of interest info (static anddynamic ones) with, at least, a location, somemandatory’s attributes and elective details(description…). So as to produce that info,the element that implements the interface uses themap info info to find and show pointof interest (POI) or to pick dish as route waypoint andfavorite. This element not solely provides searchfunctionalities for the native info however conjointly the simplest way toconnect external program to the current element andenhance the search criteria and therefore the list of results It alsoproposes an answer to urge custom dishs (not a part of thelocal map database) or to dynamically update contentand description of native POI. Using the baseline rule, BKC question can even be effortlessly resolved. Yet, it's established on thoroughly combining objects (or their mbrs). Though cutting ways were explored, it's been determined that the potency drops dramatically, once the amount of question key terms will increase, considering of the fast increase of candidate keyword covers generated. This motivates US to strengthen an additional rule cited as key word nearest neighbor growth. We have a tendency to specialise in a selected question keyword, called predominant question keyword. The objects related to the foremost question key word area unit called vital objects. The aim of the interface is to furnish issue of interest experience (static and dynamic ones) with, at least, a place, some necessary's attributes and discretionary tiny print (description…). With a read to produce that understanding, the issue that implements the interface uses the map info info to find and show purpose of interest (POI) or to pick dish as route waypoint and favorite. This component not handiest provides search functionalities for the native info however conjointly a way to affix outside program to the current component and increase the search criteria {and the|and therefore the|and conjointly the} list of results It also proposes an answer to urge custom dishs (now not a part of the neighborhood map database) or to dynamically replace content and outline of native POI.

This is achieved by specifying and providing interfacesto:

o Select POIs from one among their attributes (e.g.,Category, Name,…)
o Retrieve dish attributes (e.g., Location andDescription)
o Get dynamic content for a given dish.
o Add custom dish to the map show
o Import new POIs and POIs classes fromlocal file.

C. LBKC Computation

Given a spatial info, every object could also be associated with one or a few of key words. while not loss of generality, the article with a few of key terms area unit regenerate to over one objects placed on the identical space, each with a such that single keyword. once further process a candidate key phrase cowl, key word-NNE rule most ordinarily generates abundant less new candidate key phrase covers compared to BFbaseline rule. within the grounds that the amount of candidate key phrase covers further processed in key phrase-NNE rule is perfect the amount of key word covers generated in BF-baseline rule is way over that in keyword- NNE rule. In flip, we have a tendency to conclude that the amount of key word covers generated in baseline rule is quite over that in keyword NNE rule. This conclusion is unbiased of the principal question key word for the explanation that the analysis will not apply any constraint on the selection strategy of foremost question key word.

IV. CONCLUSION

Spatial info accommodates giant spatial objects, the time required for looking objects is a lot of. By combining R-tree technique with minimum bounding technique the performance of system for retrieving an information from info is improved, also access time is reduced. The baseline rule generates a large number of candidate keyword covers that results in dramatic performance drop once a lot of question keywords are given. The planned keyword-NNE rule applies a different process strategy, i.e., looking native best solution for every object in an exceedingly bound question keyword. As a consequence, the quantity of candidate keyword covers generated is considerably reduced. The analysis reveals that the number of candidate keyword covers which require to be further processed in keyword-NNE rule is perfect and processing every keyword candidate cowl generally generates much less new candidate keyword covers in keyword-NNE algorithm than within the baseline rule.

References:

[1] 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
[2] Yufei Tao and Cheng Sheng, "Fast Nearest Neighbor Search with Keywords", IEEE Transactions on Knowledge and Data Engineering, VOL. 26, NO. 4, APRIL 2014.
[3] Ying Zhang, Wenjie Zhang, Qianlu Lin, Xuemin Lin, Heng Tao Shen, " Effectively Indexing the Multidimensional Uncertain Objects", IEEE Transactions on Knowledge and Data Engineeringvol 26 Issue No.03 March 2014
[4] X. Cao, G. Cong, C.S. Jensen, and B.C. Ooi, "Collective Spatial Keyword Querying," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 373- 384, 2011.
[5] J. Lu, Y. Lu, and G. Cong, "Reverse Spatial and Textual k Nearest Neighbor Search," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 349- 360, 2011.
[6] Feifei Li, Member, IEEE, Bin Yao, Student Member, IEEE, and Piyush Kumar, "Group Enclosing Queries", Transactions on Knowledge and Data Engineering, VOL. 23, NO. 10, October 2011.
[7] X. Cao, G. Cong, and C. Jensen, "Retrieving top-k prestige-basedrelevant spatial webobjects,"

Proc. VLDB Endowment, vol. 3, nos.1/2, pp. 373–384, Sep. 2010.

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

[9] J .fan , G Li , L . Zhou, S Chen , and J. Hu Seal : "SPATIO- TEXTUAL SIMILARITY SEARCH ". PVLDB, 5(9): 824-835, 2012.

[10] W Huang , G .Li , K-L. tan and J Feng. "EFFICIENT SAFE-REGION CONSTRUCTION FOR MOVING TOP-K SPATIAL KEYWORD QUERIES" in CIKM 2012.

[11] Ken C.K. Lee, Wang-Chien Lee, Member, and Hong Va Leong," Nearest Surrounder Queries", IEEE Transactions on Knowledge and Data Engineering, VOL. 22, NO. 10, OCTOBER 2010

[12] Kao, B. Lee, S. Lee, F. Cheung, D." Clustering Uncertain Data Using Voronoi Diagrams and R-Tree Index,"IEEE Transactions on Knowledge and data engineering, Vol.22 No.9, 2010.