A Survey on Spatial Keyword Queries: Top-K Spatial Keyword Search (TOPK-SK)

Priti Panmand[1], Pradnya Kasture[2]

M. E Student, Dept. of Comp Engg, RMD Sinhgadh, Pune, Maharashtra, India[1]

Professor, Dept. of Comp Engg, RMD Sinhgadh, Pune, Maharashtra, India[2]

ABSTRACT: With advances in geo-positioning technologies and geo-location services, there area unit a apace growing quantity of spatio-textual objects collected in several applications like location based mostly services and social networks, during which associate degree object is represented by its spacial location and a group of keywords (terms). Consequently, the study of spacial keyword search that explores each location and matter description of the objects has attracted nice attention from the business organizations and analysis communities. Within the paper, we have a tendency to study 2 basic issues within the spacial keyword queries: prime k spacial keyword search (TOPK-SK), and batch prime k spacial keyword search (BTOPK-SK). Given a group of spatio-textual objects, question a question a question location and a group of query keywords, the TOPK-SK retrieves the highest k objects every of that contains all keywords within the question. BTOPK-SK is that the execution of sets of TOPK-SK queries. supported the inverted index and also the linear quad tree, we have a tendency to propose a unique index structure, referred to as inverted linear quad tree (IL- Quadtree), that is fastidiously designed to use each spacial and keyword based mostly pruning techniques to effectively scale back the search area. Associate degree economical formula is then developed to tackle prime k spacial keyword search. To additional enhance the filtering capability of the signature of linear quad tree, we have a tendency to propose a partition based mostly methodology. Additionally, to agitate BTOPK-SK, we have a tendency to style a replacement computing paradigm that partition the queries into teams supported each spacial proximity and also the matter connection between queries. We have a tendency to show that the IL-Quadtree technique can even with efficiency support BTOPK-SK. Comprehensive experiments on real and artificial information clearly demonstrate the potency of our strategies.

KEYWORDS: Spatial, Keyword, Batch
.

I. INTRODUCTION

With the increasing generality of the geo-positioning technologies and geo-location services, there area unit a huge quantity of spatio-textual objects obtainable in several applications. for example, within the native search service, on-line business directory (e.g., yellow pages) provides the placement info further as short descriptions of the companies (e.g., hotels, restaurants). Within the GPS navigation system, a dish (point of interest) may be a geographically anchored tack that somebody might notice helpful or attention-grabbing, that is typically annotated with texture info (e.g., descriptions and users' reviews).[21] Moreover, in several social network services (e.g., Facebook, Flickr), a large range of geo-tagged pictures area unit accumulated everyday, which might be geo-tagged by users, GPS-enabled smartphones or cameras with a constitutional GPS receiver (e.g., Panasonic Lumix DMC-TZ10). These uploaded photographs area unit typically related to multiple text labels.[22] As a result, in recent years varied abstraction keyword question models and techniques have emerged specified users will effectively exploit each abstraction and matter info of those spatiotextual objects.

In the paper, we tend to investigate the matter of conducting prime k abstraction keyword search (TOPK-SK) ; that's, given a group of spatiotextual objects, Query a question location q and a group of keywords, we tend to aim to retrieve the k nearest objects every of that contains all keywords within the question. The highest k abstraction keyword search is key in abstraction keyword queries and includes a wide spectrum of applications.[23]

II. LITRATURE SURVEY

1. Hybrid Index Structures for Location-based net Search
AUTHORS: Yinghua Chow dynasty, Xing Xie,

There is a lot of and a lot of business and analysis interest in location-based net search, i.e. finding online page whose topic is said to a specific place or region. During this style of search, location data ought to be indexed additionally as text data. However, the index of typical text computer program is set-oriented, whereas location data is two-dimensional and in metric space.

This brings new analysis issues on a way to expeditiously represent the situation attributes of web content and the way to mix 2 kinds of indexes. During this paper, we tend to propose to use a hybrid index structure, that integrates inverted files and R*-trees, to handle each matter and placement aware queries. 3 totally different combining schemes square measure studied: (1) inverted file and R*-tree double index, (2) 1st inverted file then R*-tree, (3) 1st R*-tree then inverted file. To validate the performance of projected index structures, we tend to style and implement a whole location-based net computer program that chiefly consists of 4 parts: (1) AN extractor that detects geographical scopes of web content and represents geographical scopes as multiple MBRs supported geographical coordinates; (2) an trained worker that builds hybrid index structures to integrate text and placement information; (3) a ranker that ranks results by geographical connection additionally as non-geographical relevance; (4) AN interface that is friendly for users to input location-based search queries and to get geographical and matter relevant results. Experiments on giant realworld net dataset show that each the second and also the third structures are superior in question time and also the second is slightly higher than the third. In addition, indexes supported R*-trees square measure verified to be a lot of economical than indexes supported grid structures.

2. Keyword Search on spatial Databases
AUTHORS: Ian American state Felipe, VagelisHristidis.

Many applications need finding objects nearest to a specific location that contains a collection of keywords. for instance, on-line telephone book enable users to specify an address and a collection of keywords. In return, the user obtains an inventory of companies whose description contains these keywords, ordered by their distance from the required address. The issues of nearest neighbor search on spatial information and keyword search on text information are extensively studied individually. However, to the simplest of our information there's no economical methodology to answer spatial keyword queries, that is, queries that specify each a location and a collection of keywords.

3. Batch question process for net Search Engines
AUTHORS: Shuai ring, and tantalise Attenberg

Large net search engines square measure currently process billions of queries per day. Most of those queries square measure interactive in nature, requiring a response in fractions of a second. However, there are also also square measure are variety of vital situations wherever giant batches of queries are submitted for varied net mining and system optimisation tasks that don't need an instantaneous response. Given the numerous price of corporal punishment search queries over billions of web content, it's a natural question to raise if such batches of queries are often a lot of expeditiously dead than interactive queries.

4. Joint Top-K spatial Keyword question process
AUTHORS: Dingming Shanghai dialect, Man respiratory organ Yiu, Gao Cong,

A range of technologies mix to afford the online and its users a geographical dimension. Geo-positioning technologies like GPS and Wi-Fi and cellular geo-location services, e.g., as offered by Skyhook, Google, and Spotigo, square measure getting usedincreasingly; and totally different geo-coding technologies alter the tagging of online page with positions. Studies counsel that some two hundredth of all net queries from desktop users exhibit native intent, i.e., question for native content. the proportion is likely to be higher for mobile users.

5. Distance Browsing in spatial Databases
AUTHORS: G´ısli R. Hjaltason and HananSamet

Two totally different techniques of browsing through a group of spatial objects hold on in an R-tree spatial system on the idea of their distances from AN arbitrary spatial question object square measure compared. the traditional

approach is one that produces use of a k-nearest neighbor algorithmic program wherever k is thought before the invocation of the algorithmic program. therefore ifm> k neighbors square measure required, the k-nearest neighbor algorithmic program has to be reinvoked for mneighbors, thereby presumably performing arts some redundant computations. The second approach is progressive within the sense that having obtained the k nearest neighbors, the k + first neighbor are often obtained while not having to calculate the k+1 nearest neighbors from scratch.

III. PROPOSED SYSTEM

We propose a completely unique index structure, specifically IL-Quadtree, to prepare the spatio-textual objects. associate degree economical algorithmic program is developed to support the highest k abstraction keyword search by taking advantage of the IL-Quadtree. We have a tendency to any propose a partition primarily based technique to reinforce the effectiveness of the signature of linear quadtree. To facilitate an oversized quantity of abstraction keyword queries, we have a tendency to propose a BTOPK-SK algorithmic program moreover as a question cluster algorithmic program to reinforce the performance of the system.
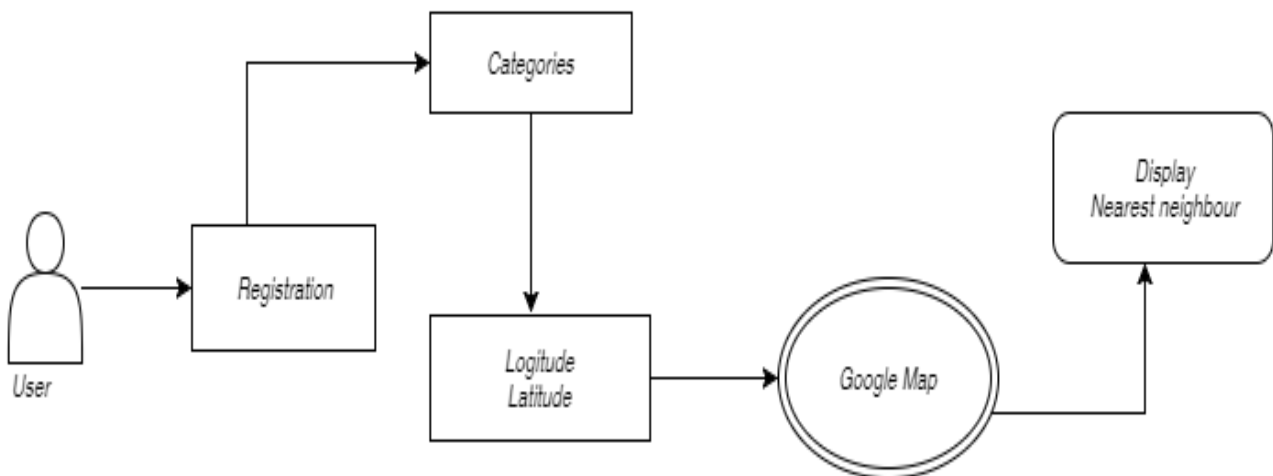

Fig: System Architecture

Advantages of Proposed System:
1. We have a tendency to conjointly investigate the matter of batch spatial keyword question (BTOPK-SK) that aims to expeditiously support an outsized variety of spatial keyword queries at an equivalent time.
2. We have a tendency to propose the inverted linear quadtree (IL-Quadtree) assortment technique that naturally combines the spatial and matter options of the objects.

IV. MATHEMATICAL MODE

INPUT:-

Let S is the Whole System Consist of
S= {I, P, O}
I = Input.
I = {U, Q, D}
U = User
U = {u1,u2….un}
Q = Query Entered by user
Q = {q1, q2, q3…qn}
D = Dataset

P = Process:

Step1: User will enter the query.

Step2: After entering query the following operations will be performed.

Step3: Spatial keyword search which explores both location and textual description of the objects to be search.

Step4: Based on the inverted index and the linear quadtree, we propose a novel index structure, called inverted linear quadtree (IL- Quadtree), which is carefully designed to exploit both spatial and keyword based pruning techniques to effectively reduce the search space.

Step5: An efficient algorithm is then developed to tackle top k spatial keyword search. To further enhance the filtering capability of the signature of linear quadtree, we propose a partition based method. In addition, to deal with BTOPK-SK, we design a new computing paradigm which partition the queries into groups based on both spatial proximity and the textual relevance between queries.

Step6: We show that the IL-Quadtree technique can also efficiently support BTOPK-SK.

Step7: Classification.

Step8: Final output optimized classifier and its performance indicator.

V. SCOPE OF PROJECT

We additional propose a partition primarily based method to boost the effectiveness of the signature of linear quadtree. To facilitate an oversized quantity of abstraction keyword queries, we tend to propose a BTOPK-SK algorithmic rule similarly as a question group algorithmic rule to boost the performance of the system. Our comprehensive experiments convincingly demonstrate the efficiency of our techniques. These techniques are used for the additional improvement.

VI. CONCLUSION

The problem of prime k spatial keyword search is vital due to the increasing quantity of spatio textual objects collected in a wide spectrum of applications. Within the paper, we propose a novel index structure, specifically IL-Quadtree, to arrange the spatio-textual objects. Associate economical rule is developed to support the highest k spatial keyword search by taking advantage of the IL-Quadtree. We have a tendency to more propose a partition based mostly method to boost the effectiveness of the signature of linear quadtree. To facilitate an outsized quantity of spatial keywordqueries, we have a tendency to propose a BTOPK-SK rule similarly as a question group rule to boost the performance of the system. Our comprehensive experiments convincingly demonstrate the efficiency of our techniques.

ACKNOWLEDGMENT

REFRENCES

[1] Y. Zhou, X. Xie, C. Wang, Y. Gong, and W.-Y. Ma, "Hybrid index structures for location-based web search," in *CIKM*, 2005.

[2] I. D. Felipe, V. Hristidis, and N. Rishe, "Keyword search on spatial databases," in *ICDE*, 2008.

[3] S. Ding, J. Attenberg, R. A. Baeza-Yates, and T. Suel, "Batch query processing for web search engines," in *Proceedings of the ForthInternational Conference on Web Search and Web Data Mining, WSDM2011, Hong Kong, China, February 9-12, 2011*, 2011, pp. 137–146. [Online]. Available: http://doi.acm.org/10.1145/1935826.1935858

[4] D. Wu, M. L. Yiu, G. Cong, and C. S. Jensen, "Joint top k spatial keyword query processing," *TKDE*, 2011.

[5] G. R. Hjaltason and H. Samet, "Distance browsing in spatial databases." *TODS*, vol. 24, no. 2, pp. 265–318, 1999.

[6] M. Christoforaki, J. He, C. Dimopoulos, A. Markowetz, and T. Suel, "Text vs. space: efficient geo-search query processing," in *CIKM*, 2011.

[7] J. B. Rocha-Junior, O. Gkorgkas, S. Jonassen, and K. Nørv°ag, "Efficient processing of top-k spatial keyword queries," in *SSTD*, 2011.

[8] R. Hariharan, B. Hore, C. Li, and S. Mehrotra, "Processing spatial- keyword (sk) queries in geographic information retrieval (gir) systems," in *SSDBM*, 2007.

[9] G. Cong, C. S. Jensen, and D. Wu, "Efficient retrieval of the top-k most relevant spatial web objects," *PVLDB*, vol. 2, no. 1, 2009.

[10] D. Wu, G. Cong, and C. S. Jensen, "A framework for efficient spatial web object retrieval," *VLDB J.*, 2012.

[11] A. Cary, O. Wolfson, and N. Rishe, "Efficient and scalable method for processing top-k spatial boolean queries," in *SSDBM*, 2010, pp. 87–95.

[12] Z. Li, K. C. K. Lee, B. Zheng, W.-C. Lee, D. L. Lee, and X. Wang, "Ir-tree: An efficient index for geographic document search," *IEEE Trans.Knowl. Data Eng.*, vol. 23, no. 4, pp. 585–599, 2011.

[13] D. Zhang, K.-L. Tan, and A. K. H. Tung, "Scalable top-k spatial keyword search," in *EDBT*, 2013, pp. 359–370.

[14] G. Li, J. Feng, and J. Xu, "Desks: Direction-aware spatial keyword search," in *ICDE*, 2012.

[15] S. B. Roy and K. Chakrabarti, "Location-aware type ahead search on spatial databases: semantics and efficiency," in *SIGMOD Conference*, 2011.

[16] J. B. Rocha-Junior and K. Nørv°ag, "Top-k spatial keyword queries on road networks," in *EDBT*, 2012.

[17] C. Zhang, Y. Zhang, W. Zhang, X. Lin, M. A. Cheema, and X. Wang, "Diversified spatial keyword search on road networks," pp. 367–378, 2014.

[18] I. Gargantini, "An effective way to represent quadtrees," *Commun. ACM*, vol. 25, no. 12, pp. 905–910, 1982.

[19] G. M. Morton, "A computer oriented geodetic data base and a new technique in file sequencing," *Technical Report, Ottawa, Canada: IBMLtd*, 1966.

[20] C. Faloutsos, "Multiattribute hashing using gray codes," in *SIGMOD Conference*, 1986.

[21] AnkitLodha, Clinical Analytics – Transforming Clinical Development through Big Data, Vol-2, Issue-10, 2016

[22] AnkitLodha, Agile: Open Innovation to Revolutionize Pharmaceutical Strategy, Vol-2, Issue-12, 2016

[23] AnkitLodha,Analytics: An Intelligent Approach in Clinical Trail Management,Volume 6 ,Issue 5 , 1000e124