Spatial Keyword Queries: Top k-Spatial Keyword Search (TOPK-SK)

Ashish Ranjan[1], Ajay Kumar[2], Akshay Sunil Dixit[3], Deepak Ranjan[4]

[1,2,3,4]Dept.Of Comp.Engg.D.Y Patil College Of Engineering,Ambi-Pune.

Abstract — *With advances in geo-positioning technologies and geo-location services, there area unit a chop-chop growing quantity of spatiotextual objects collected in several applications like location primarily based services and social networks, within which an object is delineate by its spacial location and a collection 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 industrial organizations and analysis communities. Within the paper, we tend to study 2 basic issues within the spacial keyword queries: high k spacial keyword search (TOPK-SK), and batch high k spacial keyword search (BTOPK-SK). Given a collection of spatio-textual objects, question a question location and a collection of query keywords, the TOPK-SK retrieves the nighest k objects every of that contains all keywords within the question. BTOPK-SK is that the instruction execution of sets of TOPK-SK queries[1]. Supported the inverted index and therefore the linear quadtree, we tend to propose a completely unique index structure, known as inverted linear quadtree (IL-Quadtree), that is rigorously designed to take advantage of each spacial and keyword primarily based pruning techniques to effectively cut back the search area[2][3]. An economical algorithmic program is then developed to tackle high k spacial keyword search. To any enhance the filtering capability of the signature of linear quadtree[4], we tend to propose a partition primarily based methodology. Additionally, to alter BTOPK-SK, we tend to style a brand new computing paradigm that partition the queries into teams supported each spacial proximity and therefore the matter connectedness between queries[5][6]. We tend to show that the IL-Quadtree technique may 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 are a vast quantity of spatio-textual objects offered in several applications. as an example, within the native search service, on-line business directory (e.g., yellow pages) provides the situation data further as short descriptions of the companies (e.g., hotels, restaurants). within the GPS navigation system[7], a dish (point of interest) may be a geographically anchored drawing pin that somebody could notice helpful or attention-grabbing, that is sometimes annotated with texture data (e.g., descriptions and users' reviews). Moreover, in several social network services (e.g., Facebook, Flickr), a large range of geo-tagged pictures are accumulated everyday, which may be geo-tagged by users, [8]GPS-enabled smartphones or cameras with a constitutional GPS receiver (e.g., Panasonic Lumix DMC-TZ10). These uploaded pho- to graphs are typically related to multiple text labels. As a result, in recent years varied abstraction keyword question models and techniques have emerged such users will effectively exploit each abstraction and matter data of those spatio-textual objects[9][10].

In the paper, we tend to investigate the matter of conducting prime k abstraction keyword search (TOPK-SK); that's, given a group of spatio-textual objects, letteruery a question location q and a group of keywords, we tend to aim to retrieve the k nighest objects every of that contains all keywords within the question[11]. The highest k abstraction keyword search is key in abstraction keyword queries and encompasses a wide spectrum of applications[12][13].

II.   LITERATURE SURVEY

1] Hybrid Index Structures for Location-based Web Search
AUTHORS: Yinghua Zhou, Xing Xie,
Description: There is a lot of and a lot of business and analysis interest in location-based internet search, i.e. finding web page whose topic is said to a selected place or region. During this variety of search, location data ought to be indexed

furthermore as text data. However, the index of standard text computer program is set-oriented, whereas location data is two-dimensional and in metric space.

This brings new analysis issues on the way to expeditiously represent the situation attributes of web content and the way to mix two forms 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. Three totally different combining schemes area unit studied: (1) inverted file and R*-tree double index, (2) initial inverted file then R*-tree, (3) initial R*-tree then inverted file. To validate the performance of planned index structures, we tend to style associated implement an entire location-based internet computer program that primarily consists of 4 parts: (1) associate extractor that detects geographical scopes of web content and represents geographical scopes as multiple MBRs supported geographical coordinates; (2) an skilled worker that builds hybrid index structures to integrate text and placement information; (3) a ranker that ranks results by geographical relevancy furthermore as non-geographical relevance; (4) associate interface that is friendly for users to input location-based search queries and to get geographical and matter relevant results. Experiments on giant real world internet dataset show that each the second and therefore the third structures are superior in question time and therefore the second is slightly higher than the third. To boot, indexes supported R*-trees area unit well-tried to be a lot of economical than indexes supported grid structures.

2] Keyword Search on Spatial Databases

AUTHORS: Ian De Felipe, Vagelis Hristidis.

Description: Many applications need finding objects nearest to a specified location that contains a group of keywords. as an example, on-line telephone book enable users to specify Associate in Nursing address and a group of keywords. In return, the user obtains an inventory of companies whose description contains these keywords, ordered by their distance from the desired address. The issues of nearest neighbor search on spatial information and keyword search on text information are extensively studied individually. However, to the most effective of our data there's no economical methodology to answer spatial keyword queries, that is, queries that specify each a location and a group of keywords.

3] Batch Query Processing for Web Search Engines

AUTHORS: Shuai Ding, and Josh Attenberg

Description: 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 square measure variety of vital eventualities wherever massive batches of queries are submitted for numerous net mining and system optimisation tasks that don't need a direct response. Given the numerous price of death penalty search queries over billions of web content, it's a natural question to raise if such batches of queries will be a lot of expeditiously dead than interactive queries.

4] Joint Top-K Spatial Keyword Query Processing

**AUTHORS:** Dingming Wu, Man Lung Yiu, Gao Cong,

Description: A range of technologies mix to afford the net 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, area unit being employed increasingly; and completely different geo-coding technologies alter the tagging of web 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 share is likely to be higher for mobile users.

5] Distance Browsing in Spatial Databases

AUTHORS: G´ısli R. Hjaltason and Hanan Samet

Description: Two totally different techniques of browsing through a set of spacial objects keep in associate R-tree spacial system on the premise of their distances from associate discretionary spacial question object area unit compared. The standard approach is one that produces use of a k-nearest neighbor algorithmic program wherever k is understood before the invocation of the algorithmic program. So if m & gt; k neighbors area unit required, the k-nearest neighbor algorithmic program must be reinvoked for m neighbors, thereby presumably performing arts some redundant computations. The second approach is progressive within the sense that having obtained the k nearest neighbors, the k + 1 neighbour is obtained while not having to calculate the k+1 nearest neighbors from scratch.

III. SYSTEM ARCHITECTURE
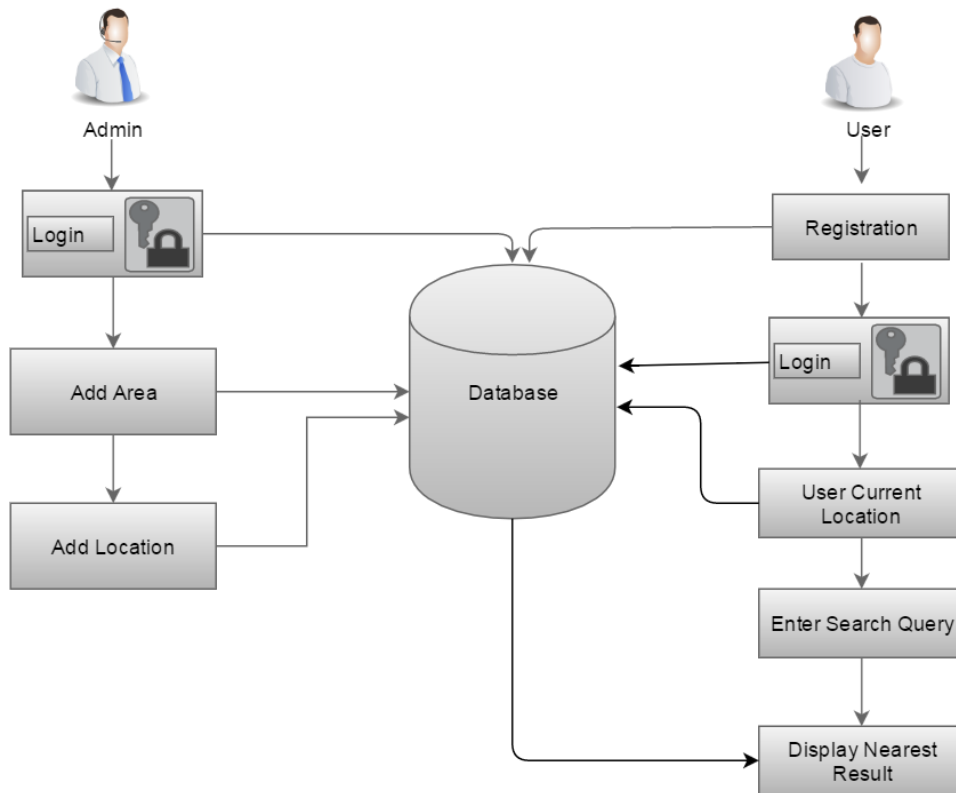
IV. SEQUENCE DIAGRAM
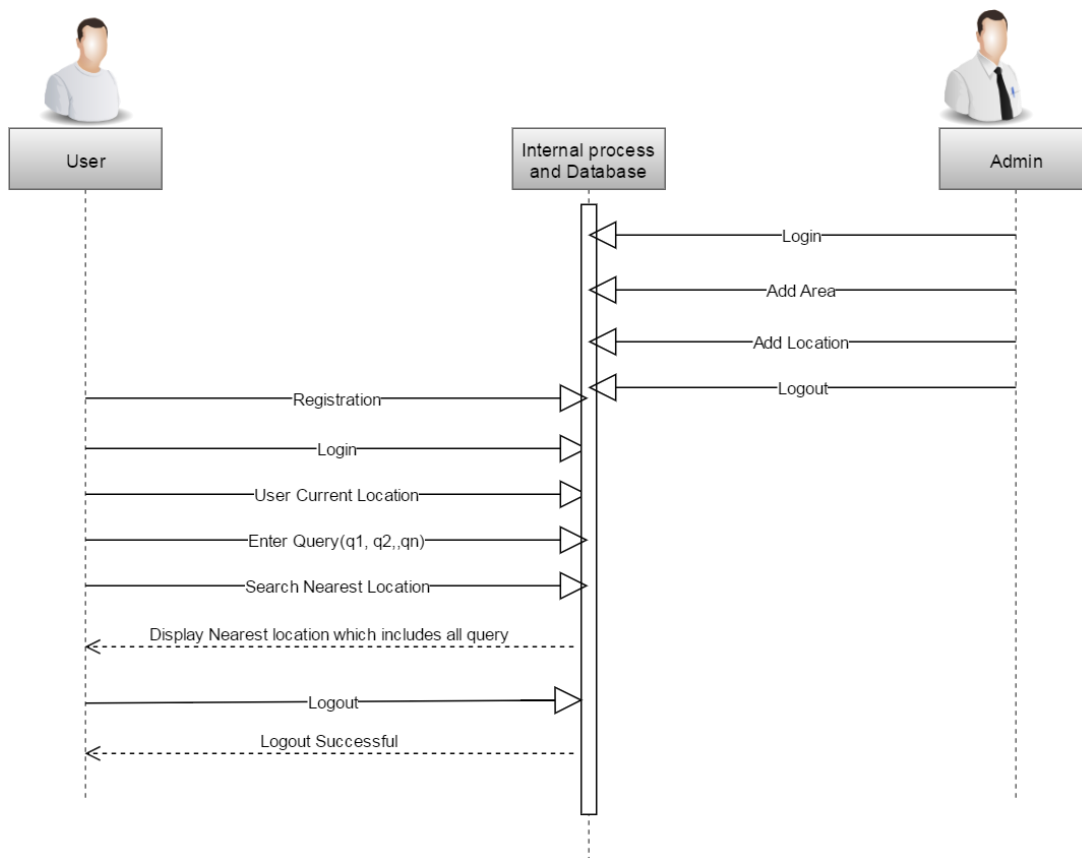
V MATHEMATICAL MODEL

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.

VI. CONCLUSION

The problem of prime k abstraction keyword search is very important thanks to the increasing quantity of spatio-textual objects collected in an exceedingly wide spectrum of applications. Within the paper, we have a tendency to propose a unique index structure, specifically IL-Quadtree, to arrange the spatio-textual objects. Associate in Nursing economical algorithmic rule 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 based mostly technique to boost 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 rule likewise as a question cluster algorithmic rule to boost the performance of the system. Our comprehensive experiments convincingly demonstrate the potency of our techniques.

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 Forth International Conference on Web Search and Web Data Mining, WSDM 2011, 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] Ankit Lodha, Clinical Analytics – Transforming Clinical Development through Big Data, Vol-2, Issue-10, 2016

[17] Ankit Lodha, Agile: Open Innovation to Revolutionize Pharmaceutical Strategy, Vol-2, Issue-12, 2016

[18] Ankit Lodha, Analytics: An Intelligent Approach in Clinical Trail Management, Volume 6 ,Issue 5 , 1000e124