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. 4, Issue 5, May 2016

Efficient Location Finding with Text Based Keyword Searching System

Ashitosh N. Mete[1], Hari Pansare[2], Vinayak Satam[3], Ramnath Banerjee[4]

B. E Student, Department of CE, Dr. D. Y. Patil Institute of Engineering and Technology Ambi, Pune, India Pune University, India

B. E Student, Department of CE, Dr. D. Y. Patil Institute of Engineering and Technology Ambi, Pune, India Pune University, India

B. E Student, Department of CE, Dr. D. Y. Patil Institute of Engineering and Technology Ambi, Pune, India Pune University, India

Asst. Professor, Department of CE, Dr. D. Y. Patil Institute of Engineering and Technology Ambi, Pune, India Pune University, India

ABSTRACT: Many search modules area unit won't to search something from anywhere; this method is employed to quick nearest neighbour search exploitation keyword. Existing work usually specialise in finding top-k Nearest Neighbours.It doesn't take into account the density of knowledge objects within the spatial area. Conjointly these strategies area unit low efficient for progressive question. however in meant system, for instance once there's seek for nearest eating place, rather than considering all the hotels, a nearest neighbour question would elicit the eating place that's, nearest among those whose menus contain spicy, hard liquor all at constant time, resolution to such queries is predicated on the IR2-tree, however IR2-tree having some drawbacks. IR2-tree contains Signature file due to its conservative nature may lead to some search result that only contains single keyword result Efficiency of IR2-tree badly is wedged attributable to some drawbacks in it. Thespatial inverted index is that the technique which is able to be the answer for this drawback.

KEYWORDS: Nearest Neighbor Search, Keyword Search, Spatial Index.

I. INTRODUCTION

A spatial info manages 3D objects (such as points, rectangles, etc.), and provides quick access to those objects choice criteria supported completely different. The importance of spatial databases is reflected by the convenience of modelling a geometric manner entities of reality. As an example, locations of restaurants, hotels, hospitals then on typically delineate as points during a map, whereas larger extents like parks, lakes, and landscapes typically as a mixture of rectangles. Several functionalities of a spatial info are helpful in numerous ways in which in specific contexts. For example, during a Geographic's data system, vary search are often deployed to find all restaurants during a sure space, whereas nearest neighbour retrieval will discover the edifice to a given address. Today, the widespread use of search engines has created it realistic to jot down spatial queries during a spanking new manner. Conventionally, queries specialize in objects' geometric properties solely, like whether or not a degree is during a parallelogram, or however shut 2 points from one another. We've seen some trendy applications that decision for the power to pick out objects supported each of their geometric coordinates and their associated texts.

For example, it'd be fairly helpful if a groundwork engine are often wont to find the closest edifice that gives "steak, spaghetti, and brandy" all at a similar time. Note that this can be not the "globally" nearest edifice (which wouldcame back by a standard nearest neighbour query), however the closest edifice among solely those providing all the demanded foods and drinks. During this paper, we tend to style a variant of inverted index that's optimized for 3D points, and is so named the spatial inverted index (SI-index). This access methodology with success incorporates purpose coordinates into a standard inverted index with tiny additional house, because of a fragile compact storage

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. 4, Issue 5, May 2016

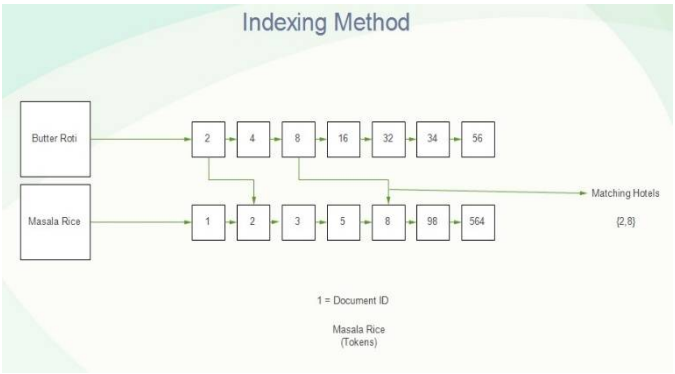
theme. Meanwhile, associate SI-index preserves the spatial vicinity of information points, associated comes with an R-tree engineered on each inverted list at very little house overhead. As a result, it offers 2 competitors ways in which for question process. We will (sequentially) merge multiple lists significantly like merging ancient inverted lists by ids. Instead, we will additionally leverage the R-trees to browse the purposes of all relevant lists in ascending order of their distances to the question point. As incontestable by experiments, the SI-index significantly outperforms the IR two -tree in question potency, typically by an element of orders of magnitude.

Nearest neighbour search (NNS), additionally called nearest purpose search, similarity search. It's an improvement drawback for finding nearest (or most similar) points. Nearest neighbour search that returns the closest neighbour of a question purpose during a set of points, is a crucial and wide studied drawback in several fields, and its wide selection of applications. We are able to search nearest purpose by giving keywords as input; it may be abstraction or matter. An abstraction info use to manage multidimensional objects i.e. points, rectangles, etc. Some abstraction databases handle additional advanced structures like 3D objects, topological coverage's, linear networks. whereas typical knowledge bases are designed to manage varied NUMERIC'S and character sorts of data, further practicality has to be additional for knowledge bases to method abstraction data type's efficiency and it provides quick access to those objects supported different choice criteria.

Keyword search is that the hottest info discovery technique as a result of the user doesn't have to apprehend either a question language or the underlying structure of the information. The search engines out there nowadays give keyword search on high of sets of documents. Once a group of question keywords is provided by the user, the programme returns all documents that are related to these question keywords. Answer to such queries relies on the IR2-tree, however IR2- tree having some drawbacks. Potency of IR2-tree badly is compact thanks to some drawbacks in it. The answer for overcoming this drawback ought to be searched. Abstraction inverted index is that the technique which is able to be the answer for this drawback. Abstraction info manages multidimensional knowledge that's points, rectangles.

This paper provides importance abstraction queries with keywords. Abstraction queries with keywords take arguments like location and such keywords and supply net objects that are organized relying upon abstraction proximity and text connection. Another approaches take keywords as mathematician predicates, looking for net objects that contain keywords and rearranging objects supported their abstraction proximity. Some approaches use a linear ranking perform to mix abstraction proximity and matter relevancy. Earlier study of keyword search in relative databases is gaining importance. Recently this attention is diverted to multidimensional data.For keyword-based retrieval, they have integrated R-tree with spatial index and signature file .By combining R-tree and signature they have developed a structure called the IR2-tree. IR2-tree has merits of both R-trees and signature files. The IR2-tree preserves object's spatial proximity which important for solving spatial queries.

II. RELATED WORK



Profitable search engines like Google and Yahoo! have introduced native search services that seem to target the retrieval of native content, e.g., associated with stores and restaurants. However, the algorithms used aren't published. a lot of attention has been given to the matter of extracting geographic data from websites. The extracted data are often employed by search engines. Using the notion of geo-coding, describes geographic indicators found in pages, like nada codes and placement names. Recent studies that think about location-aware text retrieval represent the work most

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. 4, Issue 5, May 2016

closely associated with this study. The matter of retrieving internet documents relevant to a keyword question among a pre-specified abstraction region. They propose 3 approaches supported a loose combination of associate inverted file associated an R*- tree. The simplest approach consistent with their experiments is to make associate R*-tree for every distinct keyword on the online pages containing the keyword. As a result, queries with multiple keywords got to access multiple R*-trees and to come across the results. Building a separate R*-tree for every keyword additionally needs substantial storage. Ian Delaware Felipe.et.al, the matter of high k abstraction keyword search is outlined. The IR2-Tree is projected as associate economical classification structure to store abstraction and matter data for a collection of objects. Economical algorithms are bestowed to take care of the IR2-Tree, that is, insert and delete objects. associate economical progressive algorithmic rule is bestowed to answer top-k Profitable search engines like Google and Yahoo! have introduced native search services that seem to target the retrieval of native content, e.g., associated with stores and restaurants. However, the algorithms used aren't published. A lot of attention has been given to the matter of extracting geographic data from websites. The extracted data are often employed by search engines. Using the notion of geo-coding, describes geographic indicators found in pages, like nada codes and placement names. Recent studies that think about location-aware text retrieval represent the work most closely associated with this study. The matter of retrieving internet documents relevant to a keyword question among a pre-specified abstraction region. They propose 3 approaches supported a loose combination of associate inverted file associated an R*- tree. The simplest approach consistent with their experiments is to make associate R*-tree for every distinct keyword on the online pages containing the keyword. As a result, queries with multiple keywords got to access multiple R*-trees and to come across the results. Building a separate R*-tree for every keyword additionally needs substantial storage. Ian Delaware Felipe.et.al, the matter of high k abstraction keyword search is outlined. The IR2-Tree is projected as associate economical classification structure to store abstraction and matter data for a collection of objects. Economical algorithms are bestowed to take care of the IR2-Tree, that is, insert and delete objects. Associate economical progressive algorithmic rule is bestowed to answer top-k abstraction keyword queries victimization the IR2-Tree. We have a tendency to gift a way to expeditiously answer top-k abstraction keyword queries, that relies on the tight integration of knowledge structures and algorithms employed in abstraction info search and data Retrieval (IR). Particularly, this methodology consists of building associate data Retrieval R-Tree (IR2- Tree), That could be a structure supported the R-Tree.

i. k-Nearest Neighbors Queries

In spatial databases, k nearest neighbor (KNN) and vary queries area unit basic question sorts. These 2 forms of spatial queries are extensively studied and applied in varied location-based service (LBS) applications. The solutions for nearest neighbor queries area unit designed within the context of spatial databases. Additionally, a diagram-based answer for KNN searches in spatial databases is bestowed in by partitioning an oversized regions to little region's and pre-computing distances each inside and across the regions. As a result of most Dijkstra's algorithm-based KNN solutions are shown to be economical just for short distances, for distance computation. Planned associate degree economical index (distance signature)and question process over long distances. Their technique discretizes the distances between objects and network nodes into classes then encodes these classes to execute the KNN search method. So as to hurry up KNN searches, same algorithmDesigned associate degree formula to figure the shortest ways between all the vertices within the network and using a shortest path quadtree to capture spatial coherence. With the formula, the shortest ways between all attainable vertices will be computed to answer varied KNN queries on a given spatial network. Nonetheless, all the higher than mentioned techniques principally targeted on the space metric. They didn't think about text description (keywords) of spatial objects in their question analysis processes

ii. TextRetrival

Text retrieval is another vital topic associated with spatial Keyword queries. There area unit 2 main compartmentalization techniques, inverted files and signature files, wide utilized in text retrieval systems.Signature files need a far larger house to store index structures and area unit costlier to construct and update than inverted files. The inverted files outdo Signature files in most cases. Though these ways performwell in text retrieval applications, none of them will expeditiously method spatial keyword queries. In alternative words, it's impractical to answer spatial keyword queries by merely using approaches introduced within the previous section. A good thanks to handle spatial keyword queries is to mix the 2 teams suggests that nearest neighbor queries and Text retrieval.

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. 4, Issue 5, May 2016

iii. Spatial Keyword Query

Many solutions are developed to spatial keyword queries. Websites associated with a spatial region is seek outby Location-based internet search. They delineated 3 totally different hybrid categorization structures of integration inverted files and R*-trees along. In keeping with their experiments, the simplest theme is to create Associate in Nursing inverted index on the highest of R*-trees. In alternative words, the formula initial sets up Associate in Nursing inverted index for all keywords, so creates Associate in Nursing R*-tree for every keyword. This technique performs well in spatial keyword queries, however its maintenance price is high. once Associate in Nursing object insertion or deletion happens, the answer needs to update the R*-trees of all the keywords of the item.IR-tree a hybrid index structure that could be a combination of Associate in Nursing R-tree and inverted files to method location-aware text retrieval and supply k best candidates in keeping with a rank system. It minimizes areas of introduction rectangles and increasing text into consideration throughout construction procedures. Felipe et al. developed a unique index, IR2-Tree that integrates Associate in Nursing R-tree and signature files along with answer top- k special queries keyword. It will record signature data in every node of R-trees so as to make a decision whether or not there's any object that satisfies each spatial and keyword constraints at the same time. However, the scale of area for storing signatures before IR2-Tree construction in every node is determined. Once the IR2-Tree has been designed, it's not possible to enlarge the area unless the tree is reconstructed. If the quantity of keywords grows quickly, a system can pay lots of your time repeatedly reconstruction the IR2-Tree. Hariharan et al. projected Associate in Nursing categorization mechanism, KR*-tree, which mixes Associate in Nursing R*-tree Associate in Nursing an inverted index. The distinction between their answer associate degreed is that they solely store connected keywords in every node of an R*-tree so as to avoid merging operations to search out candidates containing all keywords. If the quantity of keywords that seem in every node varies. However, such a sophisticated assortment technique features a high maintenance value furthermore. Though there are variety of previous studies on abstraction keyword queries, most of their solutions will solely judge queries in geometer areas. This limitation is thanks to the adoption of the R-tree (or its variants), that cannot index abstraction objects supported network distances, into their hybrid index structures.

III. SYSTEM IMPLEMENTATION

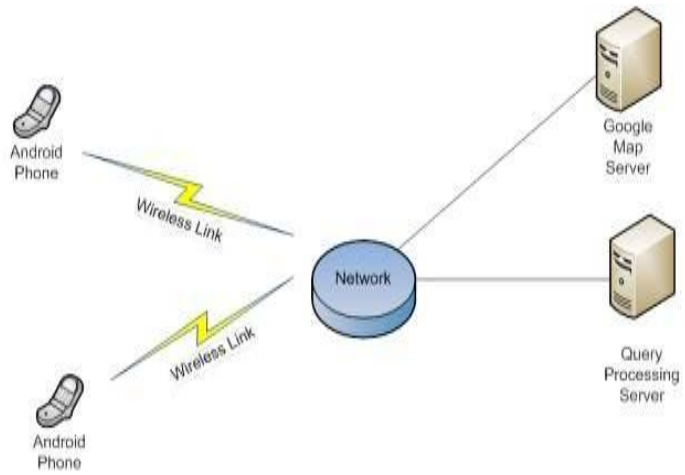


Fig. 1 System Architecture

The Figure shows the System Architecture. The above diagram shows the working of our approach which involves a Query processing's ever, Google map server and an android mobile phone. The query processing server stores the spatial data (Latitude, Longitude)of a place and information about a place such items available at that place. The mobile phone performs the searching the data stored on the server by using the phones current position and keywords and retrievestheresultednearestNeighbour'sanddisplaystheplacesonthemapbyinteractingwiththeGooglemap server.

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

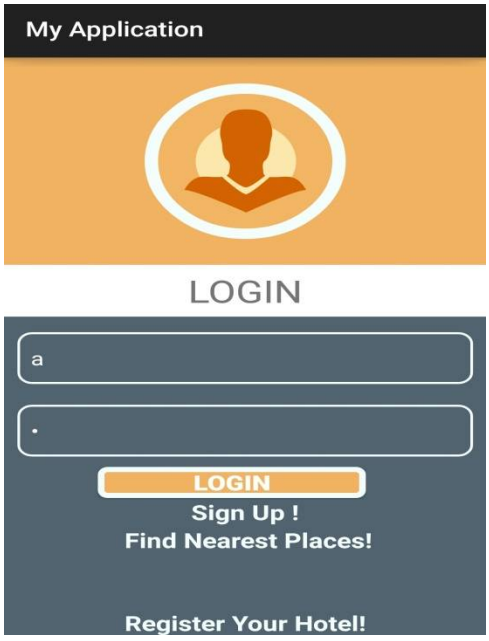
International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

Module 1: client Registration:
In this module, First user need to register. Once the user registration is done, then he/she will access the appliance. For registration user need to enter the fundamental info regarding himself. User even have to line the username and watchword. This all registration info is get keep into table. We used web services to achive this result for security and portability.

Module 2: client Login:
During this module, once the registration client will login through mentioned username and watchword. After login user can see hotel list, which user can search any hotel by name, then user can search hotels by entering multiple keywords, also nearest hotels are display first.



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. 4, Issue 5, May 2016

Module 3: edifice Registration:
During this module, Admin register the edifice with its known dish. Edifice owner need to do the registration then solely the edifice get search through application. conjointly edifice owner need to add the menu that is on the market within the edifice in order that user will search the edifice through keyword. Solely registered hotels are displayed within the application. every edifice owner can get the separate username and watchword for login.

Module 3.1: edifice Login / Admin:
During this model once edifice Owner login into application then he will insert the menu or update the menu.
Module 4: Looking out Keyword:
During this module, the user can enter the keyword looking for menus offered in edifice which can nearer from its position. Whenever user can enter keyword (menu name) it'll match information with the edifice information server and realize the closest edifice with the offered entered menu by client. For nearest hotel search we are using SI indexing technique in which every menu have its own index, now our system will match each index of one keyword with another as a result we will get exact hotels that having searched menus and also shows nearest ones.

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

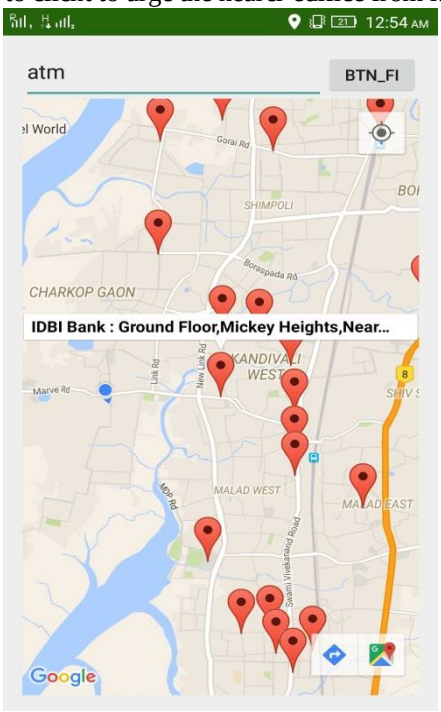
International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

Module 5:

Map read / looking out Location: during this module, all the names of edifice can seem within the list that came from information and realize the position in map (google_play_service_library) is needed for showing position of edifice in map and which can be additional easier to client to urge the nearer edifice from its current position.



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. 4, Issue 5, May 2016

Module 6:
Distance Search: during this module, client will realize the space from supply to destination. In order that it may be easier to search out the space and reached the destination. It'll offer the space of the edifice from the present location.
Logic for conniving distance between user and hotel:
DoubledLat = scientific discipline.toRadians(lat1-lat2); double dLng = scientific discipline.toRadians(lng1-lng2);
We are using this formula called haversine formula to calculate diastance between two points. It takes latitude and longitude of two points.

IV. IMPLEMENTATION OF NEAR NEIGHBOR

We are using this formula to calculate distance between two places using latitude and longitude.

$$\operatorname{hav}\left(\frac{d}{r}\right) = \operatorname{hav}(\phi_2 - \phi_1) + \cos(\phi_1)\cos(\phi_2)\operatorname{hav}(\lambda_2 - \lambda_1)$$

where

- *hav* is the haversine function:

$$\operatorname{hav}(\theta) = \sin^2\left(\frac{\theta}{2}\right) = \frac{1-\cos(\theta)}{2}$$

- *d* is the distance between the two points (along a great circle of the sphere; see spherical distance),
- *r* is the radius of the sphere,
- ϕ_1, ϕ_2: latitude of point 1 and latitude of point 2
- λ_1, λ_2: longitude of point 1 and longitude of point 2

V. CONCLUSION AND FUTURE WORK

Our Project is very effective for looking out nearest building from userwith expected menus. It will this by spatial inverted index algorithm, Merging and Distance Browsing, and GPS System. During this project we've developed associate access technique known as the spatial Inverted Index (SI-Index). This technique is fairly area economical and its ability to perform keyword increased nearest neighbour searching real time. This technique relies on standard technology of Inverted Index. It's promptly incorporable in an exceedingly industrial computer programme.The major advantage of this technique is it avoids false hits and gives 100% true result.

In future this technique is very useful for hotel vendors as well as costumers to get hotels which provides menus what they like and for hotel it's easy to get new customers and grow their business.

REFERENCES

1. S. Agrawal, S. Chaudhuri, and G. Das. Dbxplorer: A system for keyword-based search over relational databases. In Proc. of International Conference on Data Engineering (ICDE), pages 5–16, 2002.Hong-ryeol Gil1, Joon Yoo1 and Jong-won Lee2,'An On-demand Energy-efficient Routing Algorithm for Wireless Ad hoc Networks', Proceedings of the 2nd International Conference on Human. Society and Internet HSI'03, pp. 302-311, 2003. N. Beckmann, H. Kriegel, R. Schneider, and B. Seeger. The R*-tree: An efficient and robust access method for points and rectangles. In Proc. of ACM Management of Data (SIGMOD), pages 322–331, 1990.
2. G. Bhalotia, A. Hulgeri, C. Nakhe, S. Chakrabarti, and S. Sudarshan. Keyword searching and browsing in databases using banks. In Proc. of International Conference on Data Engineering (ICDE), pages 431–440, 2002.
3. X. Cao, L. Chen, G. Cong, C. S. Jensen, Q. Qu, A. Skovsgaard, D. Wu, and M. L. Yiu. Spatial keyword querying. In ER, pages 16–29, 2012.
4. X. Cao, G. Cong, and C. S. Jensen. Retrieving top-k prestige-based relevant spatial web objects. PVLDB, 3(1):373–384, 2010

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. 4, Issue 5, May 2016

5. X. Cao, G. Cong, C. S. Jensen, and B. C. Ooi. Collective spatial keyword querying. In Proc. of ACM Management of Data (SIGMOD), pages 373–384, 2011
6. B. Chazelle, J. Kilian, R. Rubinfeld, and A. Tal. The bloomier filter: an efficient data structure for static support lookup tables. In Proc. of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pages 30–39, 2004.
7. Y.-Y. Chen, T. Suel, and A. Markowetz. Efficient query processing in geographic web search engines. In Proc. of ACM Management of Data (SIGMOD), pages 277–288, 2006.
8. E. Chu, A. Baid, X. Chai, A. Doan, and J. Naughton. Combining keyword search and forms for ad hoc querying of databases. In Proc. of ACM Management of Data (SIGMOD), 2009.
9. G. Cong, C. S. Jensen, and D. Wu. Efficient retrieval of the top-k most relevant spatial web objects. PVLDB, 2(1):337–348, 2009.