ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

Adjacent Neighbor Assessment Through Keywords

[1]M.Syamala Mani, [2]S.Radhika
[1,2]Dept. of Computer Science Engineering, Raghu Institute of Technology, Visakhapatnam AP, India

Abstract
Customary spatial inquiries, closest neighbor recovery and reach pursuit comprises just conditions on items geometric property just. In any case, today, numerous cutting edge applications bolster new type of questions that expect to discover articles that fulfills both spatial information and their related content. For instance as opposed to considering every one of the lodgings, a closest neighbor inquiry would rather request the inn that is nearest to among the individuals who give administrations, for example, pool, web in the meantime. For this kind of inquiry a variation of altered file is utilized that is compelling for multidimensional focuses and accompanies a R-tree which is based on each upset rundown, and uses the calculation of least jumping system that can answer the closest neighbor inquiries with catchphrases progressively. In this paper, we propose a novel convention for area based inquiries that has significant execution enhancements as for the methodology by Ghinita at el. What's more, like such convention, our convention is composed by stages. In the first stage, the client secretly decides his/her area inside of an open framework, utilizing unaware exchange. This information contains both the ID and related symmetric key for the piece of information in the private framework. In the second stage, the client executes a communicational proficient PIR, to recover the proper piece in the private framework. This piece is unscrambled utilizing the symmetric key acquired as a part of the past stage. Our convention along these lines gives insurance to both the client and the server. The client is ensured in light of the fact that the server is not able to decide his/her area. Likewise, the server's information is ensured following a noxious client can just decode the piece of information got by PIR with the encryption key gained in the past stage. At the end of the day, clients can't increase any more information than what they have paid for. We comment this paper is an upgrade of a past work.

Keywords
Nearest Neighbor Search, IR2-tree, Nearest, Range Seek, Spatial modified List

I. Introduction

Nearest Neighbor Search (NNS), otherwise called nearest point seek, closeness look. It is a streamlining issue for discovering nearest (or most comparative) focuses. Closest neighbor seek which gives back the closest neighbor of a question point in an arrangement of focuses, is an imperative and generally concentrated on issue in numerous fields, and it has extensive variety of utilizations. We can look nearest point by giving watchwords as info; it can be spatial or literary. A spatial database use to oversee multidimensional items i.e. focuses, rectangles, and so forth. Some spatial databases handle more perplexing structures, for example, 3D objects, topological coverage's, straight systems. While commonplace databases are intended to oversee different NUMERIC'S and character sorts of information, extra usefulness should be added for databases to handle spatial information sort's productively and it gives quick access to those items in light of diverse choice criteria. Watchword hunt is the most prominent data disclosure technique in light of the fact that the client does not have to know either a question dialect or the fundamental structure of the information. The web search tools accessible today give catchphrase seek on top of sets of records. At the point when an arrangement of question watchwords is given by the client, the internet searcher returns all records that are connected with these inquiry catchphrases. Answer for such questions depends on the IR2-tree, however IR2-tree having a few downsides. Proficiency of IR2-tree gravely is affected on account of a few disadvantages in it. The answer for defeating this issue ought to be sought. Spatial upset file is the method which will be the answer for this issue. Spatial database oversees multidimensional information that is focuses, rectangles. Inquiries in spatial database have turned out to be progressively critical as of late with the expanding ubiquity of a few administrations, for example, Google Earth and Yahoo Maps, and additionally other geographic applications. Today, generally utilized of web indexes has made it reasonable to compose spatial questions newly. Generally, questions concentrate on articles just geometric properties, for instance whether a point is in rectangle or how two focuses are close from one another. Some new application permits clients to scan articles in light of both of their geometric directions and their related writings. Such sort of questions called as spatial catchphrase inquiry. For instance, if a web crawler can be utilized to discover closest lodging that offer offices, for example, pool and web in the meantime. From this question, we could first acquire the whole inn whose administrations contain the arrangement of catchphrases, and after that discover the closest one from the recovered eater

II. Related Work

In the paper 'quick closest neighbor seek with watchwords', there are systems like spatial file, altered list, closest neighbor look. The principal strategy spatial file is utilized for making records in light of the fact that there is immense measure of information should be put away for looking that information put away as xml archives. In the event that the information stockpiling made as records then space needed is less additionally time required for looking the decisive word is less. Second technique is reversed list. The rearranged file information structure is a focal part of a run of the mill web index indexing calculation. An objective of an internet searcher execution is to advance the rate of the question: discover the archives where word happens. When a file is produced, which procurements arrangements of words per archive; it is next rearranged to build up an upset record. Questioning the list would require consecutive cycle through every report and to every word to check a coordinating record. The time memory and preparing property to execute such an inquiry are not generally hypothetically reasonable. As opposed to posting the words per article in the record, the transformed list information structure is created which records the archives per word. The rearranged record delivered, the question can now be dictated by bouncing to the word id in the altered list. These were successfully transformed lists with a little measure of supplementary clarification that obliged an impossible measure of endeavor to deliver. Third strategy is closest neighbor look. Closest neighbor seek (NNS), additionally recognized as closeness inquiry, parallel quest is an advancement issue for discovering nearest focuses in metric spaces. In the paper 'Productive Keyword-Based Search for Top-K Cells in Text Cube' strategies utilized are rearranged record one-sweep, archive sorted-output, base up element programming, and pursuit space requesting. In the top k cells, there is a seeking of closest key to the question. Blocks structures bunches of single interesting gathering which demonstrates its personality. Strategy like reversed record utilized for giving list as opposed to giving entire information which can be space devouring.

A spatial database oversees multidimensional articles, (for example, focuses, rectangles, and so forth.), and gives quick access to those items taking into account diverse choice criteria. The significance of spatial databases is reflected by the comfort of demonstrating elements of reality in a geometric way [3]. Case in point, areas of eateries, inns, healing centers thus on are regularly spoken to as focuses in a guide, while bigger degrees, for example, stops, lakes, and scenes frequently as a mix of rectangles [4]. Numerous functionalities of a spatial database are valuable in different routes in particular connections [5].

III. Nearest Neighbor Search of Multidimensional Data

A nearest neighbor search in a multidimensional space is the problem of finding the nearest vector to a given vector (query vector) q among N data vectors (candidate vectors) xi(i = 1, 2, . . . , N) placed in n-dimentional space. There are two typical varieties of nearest neighbor search: (i) k-nearest neighbor search (search restricted by number) The search attempts to find the k vectors closest to the given query vector q, (ii) ε-nearest neighbor search (search restricted by range) The search attempts to find vectors within a distance ε from the given query vector q; that is, vector xi satisfying d(q, xi) are found. In a linear search wherein a given vector is compared sequentially to all vectors in a database, the computational complexity increases in direct proportion to the database size. Therefore, the development of multidimensional indexing techniques for efficient nearest neighbor search has been attracting much attention recently.5 There are various algorithms for multidimensional indexing in a Euclidean space, such as R-tree,6 R+-tree,7 R*-tree,8 SS-tree,9 SS+-tree,10 CSS+-tree,11 X-tree,12 and SR-tree,13 as well as more general indexing methods for metric spaces, for example, VP-tree,14 MVP-tree,15 M-tree,16 etc. Such indexing techniques are based on restriction of the search range by hierarchical partitioning of multidimensional search space, and they limit the scope of the basic search.

A. Problems With Multidimensional Indexing Technology in High Dimensions

Content searches of images and other multimedia content employ multidimensional feature vectors that may exceed 100 dimensions. Phenomena of the kind that cannot even be imaged in two-dimensional or three-dimensional space are known to occur in such high-dimensional space. Because the degree of spatial freedom is extremely high in higher-dimensional space, solving various problems in computational geometry and multivariate analysis involves an enormous amount of calculation and is hence notoriously difficult. These difficulties are collectively referred to as the "curse of dimensionality." In nearest neighbor searches in high-dimensional space, a phenomenon occurs whereby the search becomes more and more difficult as the dimensionality becomes higher. For example, when points are uniformly distributed in n-dimensional space, the ratio of the distance of the k-th nearest and the (k + 1)-th nearest point to a given point can be approximated by the following formula:17 E{d(k+1)NN } E{dkNN } ≈ 1 + 1 kn (2) As you can see from the above, as n becomes larger, the ratio of the distance of the k-th nearest point and the (k+ 1)-th nearest point asymptotically approaches 1. Moreover, when the points are uniformly distributed, the ratio of the distance to the nearest point to the distance to the most distant point asymptotically approaches 1 as the dimensionality becomes higher. Therefore, methods for dividing the space hierarchically entail problems in that the difference due to distance is small, making it impossible to limit the area explored, and an amount of calculation that approaches that of a linear search is required.

IV. Proposed Scheme

A spatial info manages dimensional objects (such as points, rectangles, etc.), and provides quick access to those objects supported totally different choice criteria. The importance of spatial databases is mirrored by the convenience of modelling entities of reality in an exceedingly geometric manner. for instance, locations of restaurants, hotels, hospitals so on square measure typically described as points in an exceedingly map, whereas larger extents like parks, lakes, and landscapes typically as a mix of rectangles. several functionalities of a spatial info square measure helpful in varied ways in which in specific contexts. as an example, in an exceedingly geographics system, vary search will be deployed to search out all restaurants in an exceedingly sure space, whereas nearest neighbor retrieval will discover the eating place nearest to a given address. Furthermore, because the SI-index relies on the traditional technology of inverted index, it's without delay incorporable in an exceedingly business computer programme that applies large similarity, implying its immediate industrial deserves.

In this paper, we propose a novel protocol for location based queries that has major performance improvements with respect to the approach by Ghinita at el. And. Like such protocol, our protocol is organized according to two stages. In the first stage, the user privately determines his/her location within a public grid, using oblivious transfer. This data contains both the ID and associated symmetric key for the block of data in the private grid. In the second stage, the user executes a communicational efficient PIR, to retrieve the appropriate block in the private grid. This block is decrypted using the symmetric key obtained in the previous stage.

Our protocol thus provides protection for both the user and the server. The user is protected because the server is unable to determine his/her location. Similarly, the server's data is protected since a malicious user can only decrypt the block of data obtained by PIR with the encryption key acquired in the previous stage. In other words, users cannot gain any more data than what they have paid for. We remark that this paper is an enhancement of a previous work. It will Redesigned the key structure. And added a formal security model. Implemented the solution on both a mobile device and desktop machine

V. Nearest Neighbor Search Technique

A. IR-Tree, Approximation Algorithm and Exact Algorithm

This method is used to retrieve a group of spatial web objects such that the query's keywords are cover by group's keywords and objects are near to the query location and have the lowest inter object distances. This method addresses the two instantiation of the group keyword query. First is to find the group of objects that cover the keywords such that the sum of their distances to the query is minimized. Second is to find a group of objects that cover the keywords such that sum of the maximum distance among an object in group of objects and query and maximum distance among two objects in group of objects is minimized. Both of these sub problems are NP-complete. Greedy algorithm is used to provide an approximation solution to the problem that utilizes the spatial keyword index IR-tree to reduce the search space. But in some application query does not contain a large number of keywords, for this exact algorithm is used that uses the dynamic programming [1].

B. IUR-tree (Intersection union R-tree)

Geographic objects associated with descriptive texts are becoming common. This gives importance to spatial keyword queries that take both the location and text description of content. This technique is used to analyze the problem of reverse spatial and textual k nearest

ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

neighbor search i.e finding objects that takes the query object as one of their spatial textual similar objects. For this type of search hybrid index structure is used that successfully merge the location proximity with textual similarity. For searching, branch and bound algorithm is used. In addition to increase the speed of query processing a variant of IURtree and two optimization algorithm is used. To enhance the IUR-tree text clustering is used, in this objects of all the data base is group into clusters according to their text similarity. Each node of the tree is extended by the cluster information to create a hybrid tree which is called as cluster IUR-tree. To enhance the search performance of this tree two optimization methods is used, first is based on outlier detection and extraction and second method is based on text entropy [2].

C. BR* -tree

This hybrid index structure is used to search m-closest keywords. This technique finds the closest tuples that matches the keywords provided by the user. This structure combines the R* -tree and bitmap indexing to process the mclosest keyword query that returns the spatially closest objects matching m keywords To reduce the search space a priori based search strategy is used. Two monotone constraints is used as a priori properties to facilitates efficient pruning which is called as distance mutex and keyword mutex. But this approach is not suitable for handling ranking queries and in this number of false hits is large [3].

D. IR2 -tree

The growing number of applications requires the efficient execution of nearest neighbor queries which is constrained by the properties of spatial objects. Keyword search is very popular on the internet so these applications allow users to give list of keywords that spatial objects should contain. Such queries called as a spatial keyword query. This is consisted of query area and set of keywords. The IR2 -tree is developed by the combination of R-tree and signature files, where each node of tree has spatial and keyword information. This method is efficiently answering the top-k spatial keyword queries. In this signature is added to the every node of the tree. An able algorithm is used to answer the queries using the tree. Incremental nearest algorithm is used for the tree traversal and if root node signature does not match the query signature then it prunes the whole subtrees. But IR2 -tree has some drawbacks such as false hits where the object of final result is far away from the query or this is not suitable for handling ranking queries [4].

E. Spatial Inverted Index and Minimum Bounding Method

So, new access method spatial inverted access method is used to remove the drawbacks of previous methods such as false hits. This method is the variant of inverted index using for multidimensional points. This index stores the spatial region of data points and on every inverted list Rtree is built. Minimum bounding method is used for traversing the tree to prune the search space.

VI. Processing Spatial-Keyword (SK) Queries in Geographic Information Retrieval (GIR) Systems:

Location based information stored in GIS database. These information entities of such databases have both spatial and textual descriptions. This paper proposes a framework for GIR system and focus on indexing strategies that can process spatial keyword query. The following contributions in this paper: 1) It gives framework for query processing in Geo- graphic Information Retrieval (GIR) Systems. 2) Develop a novel indexing structure called KR*-tree that captures the joint distribution of keywords in space and significantly improves performance over existing index structures. 3) This method have conducted experiments on real GIS datasets showing the effectiveness of our techniques compared to the existing solutions. It introduces two index structures to store spatial and textual information.

A. Separate Index for Spatial and Text Attributes

Advantages: -

1. Easy of maintaining two separate indices.
2. Performance bottleneck lies in the number of candidate object generated during the filtering stage.

Disadvantages: -

1. If spatial filtering is done first, many objects may lie within a query is spatial extent, but very few of them are relevant to query keywords. This increases the disk access cost by generating a large number of candidate objects. The subsequent stage of keyword filtering becomes expensive.

B. Hybrid Index

Advantages and limitations: -

1. When query contains keywords that closely correlated in space, this approach suffer from paying extra disk cost accessing R*-tree and high overhead in subsequent merging process.

VII. Hybrid Index Structures for Location-based Web Search:

There is more and more research interest in location-based web search, i.e. searching web content whose topic is related to a particular place or region. This type of search contains location information; it should be indexed as well as text information. text search engine is set-oriented where as location information is two-dimensional and in Euclidean space. In previous paper we see same two indexes for spatial as well as text information. This creates new problem, i.e. how to combine two types of indexes. This paper uses hybrid index structure, to handle textual and location based queries, with help of inverted files and R*-trees. It considered three strategies to combine these indexes namely: 1) inverted file and R*-tree double index.2) first inverted file then R*-tree.3) first R*-tree then inverted file. It implements search engine to check performance of hybrid structure, that contains four parts:(1) an extractor which detects geographical scopes of web pages and represents geographical scopes as multiple MBRs based on geographical coordinates. (2) The work of indexer is use to build hybrid index structures integrate text and location information. (3) The work of ranker is to ranks the results by geographical relevance as well as non-geographical relevance. (4) an interface which is friendly for users to input location-based search queries and to obtain geographical and textual relevant results.

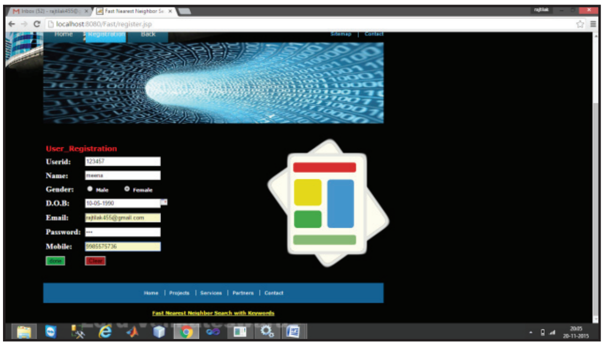
Advantages: -

1. Instead of using two indexes for textual and spatial information. this paper gives hybrid index structures that integrate text indexes and spatial indexes for location based web search.

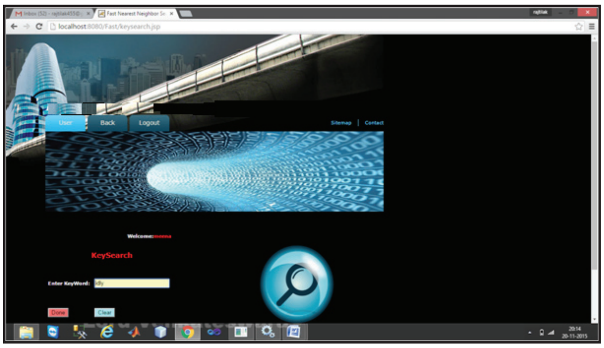
Disadvantages: -

1. Indexer wants to build hybrid index structures to integrate text and location information of web pages. To textually index web pages, inverted files are a good. To spatially index web pages, two-dimensional spatial indexes are used, both include different approaches, this cause to degrading performance of indexer.
2. In ranking phase, it combine geographical ranking and non-

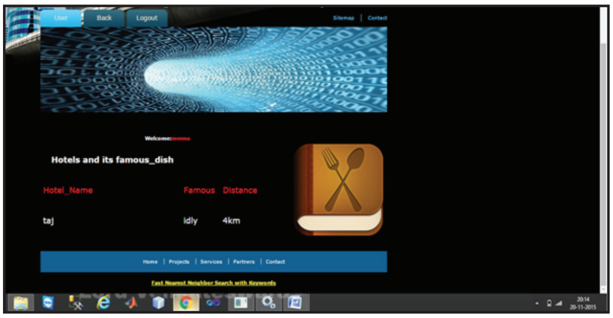
geographical ranking, combination of two rankings and the computation of geographical relevance may affects on performance of ranking.



User registration with user name and password inorder to use the services that are provided by the admin.



Search with the keyword which are provided by admin and can see the distance and the place where available.



The search is completed and the data is shown in the image.

VII. Conclusion

In this paper, we proposed a new technique for efficient nearest neighbor search in a set of high-dimensional points. Our technique is based on the pre-computation of the solution space of any arbitrary nearest-neighbor search. This corresponds to the computation of the voronoi cells of the data points. Since voronoi cells may become rather complex when going to higher dimensions, we presented a new algorithm for the approximation of high-dimensional voronoi cells using a set of minimum bounding (hyper-) rectangles. Although our technique is based on a pre-computation of the solution space, it is dynamic, i.e. it supports insertions of new data points. We finally showed in an experimental evaluation that our technique is efficient for various kinds of data and clearly outperforms the state of the art nearest-neighbor algorithms.

We have seen plenty of applications calling for a search engine that is able to efficiently support novel forms of spatial queries that are integrated with keyword search. The existing solutions to such queries either incur prohibitive space consumption or are unable to give real time answers. In this paper, we have remedied the situation by developing an access method called the spatial inverted index (SI-index). Not only that the SIindex is fairly space economical, but also it has the ability to perform keyword-augmented nearest neighbor search in time that is at the order of dozens of milliseconds. Furthermore, as the SI-index is based on the conventional technology of inverted index, it is readily incorporable in a commercial search engine that applies massive parallelism, implying its immediate industrial merits.

References

[1] S. Agrawal S. Chaudhuri, G. Das. Dbxplorer,"A System for Keyword-based Search over Relational Databases", Proc. Of International Conference on Data Engineering (ICDE), (2002), pp. 5–16.

[2] N. Beckmann, H. Kriegel, R. Schneider, B. Seeger,"The R*-tree: An Efficient and Robust Access Method for Points and Rectangles", Proc. of ACM Management of Data (SIGMOD), (1990), pp. 322–331.

[3] G. Bhalotia, A. Hulgeri, C. Nakhe, S. Chakrabarti, S. Sudarshan,"Keyword Searching and Browsing in Databases Using Banks", Proc. of International Conference on Data Engineering (ICDE), (2002), pp. 431– 440.

[4] X. Cao, L. Chen, G. Cong, C. S. Jensen, Q. Qu, A. Skovsgaard, D. Wu, M. L. Yiu,"Spatial Keyword Querying. ER, (2012), pp. 16–29. X. Cao, G. Cong, and C. S. Jensen, "Retrieving top-k prestige-based relevant spatial web objects", PVLDB, vol. 3, no. 1, (2010), pp. 373–384.

[5] X. Cao, G. Cong, C. S. Jensen, B. C. Ooi,"Collective Spatial Keyword Querying", Proc. of ACM Management of Data (SIGMOD), (2011), pp. 373– 384.

[6] B. Chazelle, J. Kilian, R. Rubinfeld, A. Tal,"The Bloomier Filter: An Efficient Data Structure for Static Support Lookup Tables", Proc. of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), (2004), pp. 30–39.

[7] Y.-Y. Chen, T. Suel, A. Markowetz,"Efficient Query Processing in Geographic Web Search Engines", Proc. of ACM Management of Data (SIGMOD), (2006), pp. 277–288.

M.Syamala Mani pursuing her M.Tech in the department of Computer Science and Engineering, Raghu Institute of Technology, Dakamarri Village, Bhimunipatnam Mandal, Visakhapatnam, A.P., India. Affiliated to Jawaharlal Nehru Technological University, Kakinada. Approved by AICTE, NEW DELHI. She obtained her B.Tech(CSE) from Avanthi St.Theressa Institute of Engineering and Technology, Garividi, Vizianagaram.

S.Radhika, M.Tech working as associate Professor in the department of Computer Science and Engineering, Raghu Institute of Technology, Dakamarri Village, Bhimunipatnam Mandal, Visakhapatnam, A.P., India. Affiliated to Jawaharlal Nehru Technological University, Kakinada. Approved by AICTE, NEW DELHI.