International Journal of Scientific Engineering and Research (IJSER)
www.ijser.in
ISSN (Online): 2347-3878, Impact Factor (2014): 3.05

A Review: Fast nearest Neighbour Search with Keywords

Sphurti S. Sao[1], Dr. Rahila Shiekh[2]

[1]M. Tech Student III Sem, Dept of CSE, RCERT Chandrapur, MH, India RCERT Chandrapur, MH, India
[2]Assistant Professor, Dept of CSE, RCERT Chandrapur, MH, India RCERT Chandrapur, MH, India

Abstract: ***Spatial queries with conventional methods like range search or nearest neighbor search, involve only the use of geometric properties of the objects. But nowadays many modern applications select the objects based on both properties: geometric coordinates and texts associated with it. An existing solution for searching difficult queries uses IR-tree method and we have discussed it in the paper. We propose a new method with an objective of finding the nearest neighbor of the query while reducing the delay time incurred in searching and enchasing the of the result of a query.***

Keyword: Nearest Neighbor Search, Spatial Queries, IR-tree

1. Introduction

A spatial database or geodatabase is optimized to store and query data that represents objects defined in a geometric space. A spatial database manages multidimensional objects (such as points, rectangles, etc.), and provides fast access to those objects based on different selection criteria. The importance of spatial databases is reflected by the convenience of modelling entities of reality in a geometric manner. For example, locations of restaurants, hotels, hospitals and so on are often represented as points in a map, while larger extents such as parks, lakes, and landscapes often as a combination of rectangles. Many functionalities of a spatial database are useful in various ways in specific contexts. For instance, in a geography information system, range search can be deployed to find all restaurants in a certain area, while nearest neighbor retrieval can discover the restaurant closest to a given address [1]. Most spatial databases allow representing simple geometric objects such as points, lines and polygons. Some spatial databases handle more complex structures such as 3D objects, topological coverage's, linear networks. While typical databases are designed to manage various numeric and character types of data, additional functionality needs to be added for databases to process spatial data types efficiently. These are typically called geometry or feature.[8] The Open Geospatial Consortium created the Simple Features specification and sets standards for adding spatial functionality to database systems. Today, the widespread use of search engines has made it realistic to write spatial queries in a brand-new way. It provides special functions and indexes for querying and manipulating that data using something like Structured Query Language (SQL). While it is often used as just a storage container for spatial data, it can do much more than that. Although a spatial database need not be relational in nature, most of the well-known ones are.

A spatial database gives you both a storage tool and an analysis tool. A spatial database is not constrained by the need to present data visually. The pizza shop planner can store an infinite number of attributes of the pizza loving household, income level, children in household, pizza ordering history, and even religious preferences and cultural upbringing (as they relate to topping choices on a pizza). More importantly, the analysis need not be limited to the number of variables that can be juggled in the brain. The planner can ask questions like, “Give me a list of neighborhoods ranked by the number of high-income pizza lovers with more than 2 children.” Furthermore, we can add unrelated data such as location of existing pizzerias or even the health-consciousness level of various neighborhoods. Our questions of the database could be as complicated as “Give me a list of locations with the highest number of target households where the average closest distance to any pizza store is greater than 16 kilometres (10 miles). Oh and toss out the health-conscious neighborhoods .

2. Related Work and Literature Survey

Many applications need finding objects that are nearest to a given location that contains a group of keywords. An increasing variety of applications need the economical execution of nearest neighbour (NN) queries affected by the properties of the spatial objects. Owing to the recognition of keyword search, notably on the net, several of those applications enable the user to produce a listing of keywords that the spatial objects ought to contain, in their description or alternative attribute. For example, real estate websites enable users to go looking for properties with specific keywords in their description and rank them in line with their distance from a given location. We tend to decision such queries spatial keyword queries [10][5]. A spatial keyword query consists of a query space and a group of keywords. The solution could be a list of objects hierarchical in line with a mix of their distance to the query space and also the connection of their text description to the query keywords. A simple nevertheless widespread variant that is employed is the distance first spatial keyword query, wherever objects square measure hierarchical by distance and keywords square measure applied as a conjunctive filter to eliminate objects that don't contain them. Unfortunately there's no economical support for top-k abstraction keyword queries, wherever a prefix of the results list is needed. Instead, current systems use ad-hoc combos of nearest neighbour (NN) and keyword search techniques to tackle the matter. For an example, associate points R-Tree is employed to seek out the closest neighbours associate points for every neighbour an inverted index is employed to envision if the query keywords area unit contained. [7]The economical methodology to answer top-k spatial keyword queries is predicated on the combination of knowledge structures and algorithms utilized in spatial information search and Information Retrieval (IR). [2], [3], [4], [6], Particularly, the strategy consists of building an Information Retrieval R- Tree (IR2- Tree) that could be a structure supported the R- Tree. At query time an incremental algorithm is employed that uses the IR2-Tree to efficiently produce the top results of the query. The IR2-Tree is a R-Tree wherever a signature is supplementary to every node v of the IR2- Tree to denote the matter content of all spatial objects within the sub tree non-moving at „v". The top-k spatial keyword search formula that is impressed by the work of Hjaltason and Samet [11] exploits this data to find the highest query results by accessing a bottom portion of the IR^2-Tree. This work has the subsequent contributions:

The matter of top-k spatial keyword search is outlined.

The IR^2-Tree is as an economical categorization structure to store spatial and matter data for a group of objects. Economical algorithms also are bestowed to take care of the IR^2-Tree, that is, insert and delete objects.

An economical progressive formula is bestowed to answer top-k spatial keyword queries mistreatment the IR^2-Tree.

The IR^2-Tree could be a combination of a R-Tree and signature files. Specially, every node of an IR^2-Tree contains each abstraction and keyword information, the previous within the kind of a minimum bounding space and therefore the latter within the kind of a signature. An IR^2-Tree facilitates each top-k abstraction queries and top-k abstraction keyword queries. R-tree, a preferred spatial index, and signature file, a good technique for keyword-based document retrieval.[1] IR^2- tree (Information Retrieval R-Tree) structure developed, that has the strengths of each R-trees and signature files. Like R-trees, the IR^2-tree preserves objects" spatial proximity that is the key to determination spatial queries expeditiously. Like signature files, the IR2-tree is in a position to filter a substantial portion of the objects that does not contain all the query keywords, so considerably reducing the amount of objects to be examined.

3. Proposed System

The proposed work is based on R-tree and performs searching operation on it. There methodologies used in proposed system ar described as below and the flow of the proposed system is depicted in figure 2

A. R-Tree:

R-trees are tree data structures used for spatial access methods, i.e., for indexing multi-dimensional information such as geographical coordinates, rectangles or polygons.

The key idea of the data structure is to group nearby objects and represent them with their minimum bounding rectangle in the next higher level of the tree; the "R" in R-tree is for rectangle. Since all objects lie within this bounding rectangle, a query that does not intersect the bounding rectangle also cannot intersect any of the contained objects. At the leaf level, each rectangle describes a single object; at higher levels the aggregation of an increasing number of objects.

B. Processing spatial queries:

Spatial queries will be processed with the help of group nearest neighbor query technique. In group nearest neighbor query user will give input for particular point. Given two sets of points P and Q, a group nearest neighbor (GNN) query retrieves the point(s) of P with the smallest sum of distances to all points in Q.

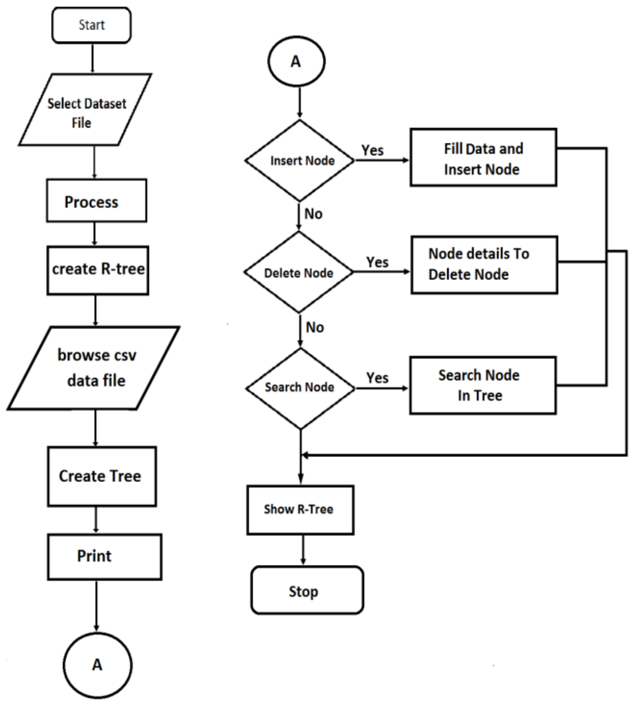

Figure 1: We will create R-Tree from the input spatial data file. The input spatial data file contains Name of place, latitude and longitude and type of the place (hospital, school, college, building, police station etc.).We will parse this file to create R tree, which can be further used in inserting, deleting and searching a node.

4. Conclusion

The objective of our work is to increase the speed of query processing and give the real time answer of spatial query. As we have seen many application which is called for a search engine that is able to support new form of query that are combined with keyword search. The existing system does not give real time answer for such type of query. For this purpose we proposed to use R-tree and spatial inverted index which is readily incorporable into commercial search engine. that applies massive parallelism, implying its immediate industrial merits.

References

[1] Yufei Tao and Cheng Sheng, “Fast Nearest Neighbor Search with Keywords”, IEEE transactions on knowledge and data engineering, VOL. 26, NO. 4, APRIL 2014.
[2] X. Cao, L. Chen, G. Cong, C.S. Jensen, Q. Qu, A. Skovsgaard, D. Wu, and M.L. Yiu, “Spatial Keyword Querying, ” Proc. 31st Int’l Conf. Conceptual Modeling (ER), pp. 16-29, 2012..

[3] 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.
[4] 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
[5] D. Zhang, Y.M. Chee, A. Mondal, A.K.H. Tung, and M. Kitsuregawa, "Keyword Search in Spatial Databases: Towards Searching by Document," Proc. Int'l Conf. Data Eng. (ICDE), pp. 688-699, 2009.
[6] G. Cong, C.S. Jensen, and D. Wu, "Efficient Retrieval of the Top-k Most Relevant Spatial Web Objects, " PVLDB, vol. 2, no. 1, pp. 337- 348, 2009.
[7] I.D. Felipe, V. Hristidis, and N. Rishe, "Keyword Search on Spatial Databases," Proc. Int'l Conf. Data Eng. (ICDE), pp. 656-665, 2008.
[8] R. Hariharan, B. Hore, C. Li, and S. Mehrotra, "Processing Spatial- Keyword (SK) Queries in Geographic Information Retrieval (GIR) Systems," Proc. Scientific and Statistical Database Management (SSDBM), 2007.
[9] Y.-Y. Chen, T. Suel, and A. Markowetz, "Efficient Query Processing in Geographic Web Search Engines," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 277-288, 2006.
[10] V. Hristidis and Y. Papakonstantinou, "Discover: Keyword Search in Relational Databases, " Proc. Very Large Data Bases (VLDB), pp. 670-681, 2002
[11] G.R. Hjaltason and H. Samet, "Distance Browsing in Spatial Databases, "ACM Trans. Database Systems, vol. 24, no. 2, pp. 265-318, 1999