www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 2 February 2015, Page No. 10520-10523

Survey on Efficient Retrieval of Nearest Neighbor over Relational Database by using Keywords

Miss. Nikita R. Barge, Prof. Pravin R. Barapatre

ME Computer Engineering Student
SKN SITS , Lonavala
SavitriBai Phule Pune University, India
HOD,Department of Information Technology.
SKN SITS , Lonavala
SavitriBai Phule Pune University, India
Email: niki_barge27@rediffmail.com

Abstract- Many applications have need of finding objects nearby to a specified location that contain a set of keywords. Predictable spatial queries, consider for the object geometric include nearest neighbor retrieval and range search. From this find out the predicate of spatial and predicate of associate texts. Academic research in this area has attention mainly on techniques for extracting geographic knowledge from the web . At this moment in time the finest way out to such queries is based going on IR2-tree , which showing seriously impact its efficiency. IR2-Tree is combination of R-Tree with superimposed text signatures. For retrieving more efficient data we used the relevance feedback methods. This method makes this approach more efficient, robust as well as reliable. In real time applications such as medical, banking et. Geographic search engine query processing is different as it require a arrangement of text and spatial data processing techniques. In support of case, consider all the restaurants, a nearest neighbor query would as an alternative put for the restaurant that is the closest along with those contain steak, spaghetti, brandy all at the same time. Also motivate the by support and develop new method that is spatial inverted. These proposed techniques best the IR2-tree in query response time.

1.Introduction

In this paper we focus on geographic web search engines. IR2 tree is the combination of R-tree and signature files. Application number increasing have need of execution of nearest neighbor (NN) queries reserved by the properties of the spatial objects. For this method we think about spatial keyword queries. Spatial keyword query is find the nearest hotel. A typical such query takes a location and a set of keywords and proceeds the single spatial web object that best matches these keywords. Which is used in our running example, is the distance-first spatial keyword query, where objects are ranked by distance and keywords are useful as a conjunctive pass through a filter to remove objects that do not contain them. The Incremental Nearest Neighbor algorithm presented by using the construction of an R-Tree to access a smallest number of R-Tree nodes and objects to retrieve the objects nearest to a specified point or else area in an incremental approach. The advancements in database and speedily growing popularity of location-based services results in massive amounts of data being collected in databases.

Today, the general use of search engines has made it sensible to write spatial queries in a unused way. Predictably, queries center on objects geometric properties , such as whether a point is in a rectangle, or how close two points are each other. We have see some current applications that describe for the capacity to select objects based on both of their geometric coordinates and their associated texts. For example, it would be reasonably useful if a search engine can be used to find the nearest restaurant that offers “steak, spaghetti, and brandy” all at the same time. They satisfactorily integrate two well-known concepts: R-tree, a popular spatial index, and signature file, an effective method for keyword-based document retrieval.

A. Spatial Inverted List

The spatial inverted list (SI-index) is basically a compacted

version of an I-index with implanted coordinate as described. Query processing with an SI-index can be done each by merging, or together with R-trees in a distance browsing method. Moreover, the compression eliminate the fault of a predictable index such that an SI-index consume much less space.

B. Query Keywords

To answer a query, form of query descriptor for each of the block descriptor files. For simplicity, it will be implicit that a single regular word has been abounding in the query. Queries which identify common words or multiple terms. The single term queries, it is accurately one bit set in each of the query descriptors. The equivalent slices are retrieved from the descriptor files and also stored in memory. It is required to determine the record matches using the information about the block matches for organizations.

C. Nearest Neighbor Queries

K-nearest neighbor queries on a spatial database is a standard database difficulty. Many of the methods consider for help out the k-NN search. The largest part of algorithm is the branch and bound algorithm which goes to R-tree while maintaining a documentation of k potential nearest neighbors in a precedence queue. There include also be attempts to use sort queries to solve the k-NN search problem. The necessary idea is to use a choice query to retrieve the likely k-NNs. This algorithm is added wholesale by improving the section belief and during a better-quality search technique of the k-NN in the fragment.

D. Signature file

Signature files think as a manner to resourcefully search a collection of text documents. On signature files current method built. In this effort, document describing a spatial object as a text structure block in their document and build related structures on top of this set of objects. These are superimposed codes achieve from the phrasing block. A signature is a bit string formed from term values that are used to index a record. Indexing using signature files assigns a signature or descriptor to every record in the data file. To perform a query, the descriptors are examined to identify potential matches A signature is a bit string formed from word values that are used to index a record. Indexing with signature files assigns a signature or descriptor to every record in the data file. To get a query, the descriptors are examined to classify possible matches. For example, to form a descriptor for a record using the method of superimposed coding, each term in the record to be indexed is identified, and a descriptor is formed for that term.

2.Literature Survey

In 2008, D. Felipe, V. Hristidis and N. Rishe proposed a user identification for Inverted index concept consider for implementing spatial query search. We divide the related work into two parts. First part covers related work on NN Query although the second section covers work done on performing Probabilistic Nearest Neighbor queries on unsure data. Spatial web objects are fast in occurrence, and several works on geographical retrieval study the problem of extracting geographic information from web pages , which yield spatial web objects that can afterward be queried. Profitable services such as Google and Yahoo! offer local search functionality.

In 1999, G. R. Hjaltason and H. Samet have find out that the spatial keyword query, they go again spatial web objects, e.g., stores and restaurants, close to the query location. The results consist of single objects that each satisfy the query in isolation. In contrast, we aim to find groups of objects such that the objects in a group jointly convince a query. Frequent recently planned mixture index that strongly combine spatial indexing and text indexing. In these index, each entry in a tree node supplies a keyword digest field that quickly review the keywords in the subtree rooted. This enable dissimilar entries to be reduce all through query processing. Our proposed algorithms are not coupled to the IR-tree, but can be used also with the other strongly combined index. Most offered works on spatial keyword queries retrieve single objects that are close up to the query point and are relevant to the query keywords. In gap, we recover groups of objects that are close to the query point and collectively meet the keywords requirement. To the best of our knowledge, the only work that retrieves groups of spatial keyword objects.

Later, in 2009 E. Chu, A. Baid, X. Chai, A. Doan, and J. Naughton find drawback in keyword search and forms for ad hoc querying of databases. Consider about the designing and generating forms in a systematic fashion, handling keyword queries that are a mix of data terms and schema terms. From this consideration we concluded that query rewrite by mapping data values to schema values during keyword search, Also simply displaying the returned forms as a list.developing automated techniques for generating better form descriptions**.**

In 2006, Y.-Y. Chen, T. Suel, and A. Markowetz concluded that Geographic search technique is used for receiving significant interest from major search engine**.** Also extracting geographic knowledge from the web by using this technique. The problem of this paper is efficient query processing in scalable geographic search engines, reason of query

processing is thousands of machines used by the major engines. Naive Algorithms used for this solution: Text-First Baseline and Geo-First Baseline.

In 2009, C. S. Jensen, and D. Wu have find out that the keyword query returns ranked objects that are near a query location and that have textual descriptions that match query keywords. Also relevant result is find out the nearby objects ,also large dataset are created for this implementation. Prestige-based relevance to capture both the textual relevance of an object to a query and the effects of nearby objects by using Location-aware top-k Prestige-based Text retrieval (LkPT)Algorithm

3. IR2-Tree

IR2-Tree is a combination of an R-Tree and signature files, both are well known techniques in spatial database. Also contain the each node of an IR2-Tree include both spatial and keyword information; the previous in the form of a minimum bounding area and the final in the form of a signature. An IR2-Tree facilitates both spatial queries and spatial keyword queries.

3.1 Other relevant work

We consider nearest neighbor search spray in spatial keyword search nearest neighbor queries explicitly similar to our formulation, but differ the how determine the result of query. Relevance consider that the document of an object p and a query q. Also consider the distance of p and q and highest similarity is returned. In geographic web search, all webpage are assigned for geographic region. Some technique gives the development of spatial keyword queries . A classic such query takes a location and a set of keywords as arguments and returns the single spatial web object that greatest matches these arguments.

4.Conclusion

In this paper we introduced the difficulty of spatial keyword search and explained the giving limitations of current approaches. We calculated a result which is widely faster than existing approaches and compare R-Trees with signature files techniques. Relevance feedback method very efficient and reliable. An efficient incremental algorithm was presented that uses the IR2-Tree to answer spatial keyword queries.

Acknowledgment

We would like to thanks to our guide & respected teachers for their constant support and motivation for us. Our sincere thanks to Sinhgad Institute of Technology for providing a strong platform to develop our skill and capabilities.

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),2002.

[2] 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),1990.

[3] 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), 2002.

[4] X. Cao, G. Cong, and C. S. Jensen. Retrieving top-k prestige-based relevant spatial web objects. PVLDB, 2010.

[5] X. Cao, G. Cong, C. S. Jensen, and B. C. Ooi. Collective spatial
keyword querying. In Proc. of ACM Management of Data (SIG-MOD), 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), 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), 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, 2009.

[10] I. D. Felipe, V. Hristidis, and N. Rishe. Keyword search on spatial
databases. In Proc. of International Conference on Data Engineering
(ICDE), 2008.

[11] G. R. Hjaltason and H. Samet. Distance browsing in spatial databases. ACM Transactions on Database Systems (TODS), 1999.

[12] J. S. Vitter. Algorithms and data structures for external memory.
Foundation and Trends in Theoretical Computer Science,2006.