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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Spatial Index Keyword Search in Multidimensional Database

Sushma Ahirrao, Rahul Chinchore

M. E Student, Department of Computer Engineering, GHRIEM, Jalgaon, India

Assistant Professor, Department of Computer Engineering, GHRIEM, Jalgaon, India

ABSTRACT: Nearest neighbor search in multimedia databases needs more support from similarity search in query processing. Range search and nearest neighbor search depends mostly on the geometric properties of the objects satisfying both spatial predicate and a predicate on their associated texts. We do have many mobile applications that can locate desired objects by conventional spatial queries. Current best solution for the nearest neighbor search is IR2 trees which have many performance bottlenecks and deficiencies. So, a novel method is introduced in this paper in order to increase the efficiency of the search called as Spatial InverterKeyword. This new SI index method enhances the conventional inverted index scheme to cope up with high multidimensional data and along with algorithms that's compatible with the real time keyword search.

KEYWORDS: Querying, multi-dimensional data, indexing, hashing

I. INTRODUCTION

Multidimensional objects such as points, rectangles managed by spatial databases provide fast access to those objects based on different selection criteria. For example, user search the keywords like "Social Networks" if the Data Set hasvalues they are mining and result are display .Some of may have few applications which finds the objects in a huge multidimensional data along with its associated texts. There are easy ways to support queries that combine spatial and text features. For example, if we want to search a café name whose all list contains keywords {Mocha, Espresso, Cappuccino} stored in data base it would fetch all the particular restaurants with the keywords and from that list gives the nearest one to related search restaurants keyword . This approach can also be in another way but this straight forward approach has a drawback, which they will fail to provide real time answers on difficult inputs. A typical example, while all the closer neighbors are missing at least one of the query keywords, that the real nearest neighbor lies quite far away from the query point. The introduction of internet has given rise to an ever increasing amount of text data associated with multiple dimensions (attributes), for example customer feedbacks in online shopping website like flipkart as they are always associated with the price, specifications and product model. Keyword query, one of the most popular and easy-to-use ways retrieves useful data from plain text documents. Given a set of keywords, existing methods aim to find joins or all the relevant items that contains a few or all the keywords. Spatial queries with keywords have not been explored. Recently, attention was diverted to multimedia databases. The integration of two well-known concepts: R-tree , a popular spatial index, and signature file , an effective method for keyword-based document retrieval. This makes to develop a structure called IR2 trees, which has strengths of both signature files and R-Trees. Like R-Trees, IR2 -Tree has object spatial proximity that solves spatial queries efficiently. On the other side, the IR2 -tree is able to filter a considerable portion of the objects that do not contain all the query keywords, like signature files.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

II. LITERATURE SURVEY

A. Keyword search on spatial databases

This method focuses on searching top-k nearest neighbor query. They present an efficient method to answer top-k spatial keyword queries, it introduce an indexing structure called IR2-Tree (Information Retrieval R-Tree). IR2-tree is the most related work which is similar to our work. Here incremental algorithm is used to answer Top-k spatial keyword queries using the IR2 -Tree. Nothing but a hashing-based framework that is based on the concept of superimposed coding as explored in this method. The problem with this technique is that as the number of words grows in size, scanning the entire list become tedious. When the list is not scanned it may result in false hits. As a solution to this problem inverted indexes were introduced in [5].

B. Spatial Keyword Query

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 m-closest 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 are 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.[2]

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

Location based information is stored in GIS database. These information entities of such databases have both spatial and textual descriptions. This method introduces a framework for GIR system and focus is on indexing strategies that can process spatial keyword query. It introduces two index structures to store spatial and textual information.1) Separate index for spatial and text attributes 2) Hybrid index. But by using first structure that is separate index for spatial and text attributes, if 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. And by using second structure that is hybrid index there are high overhead in subsequent merging process. Idea of geographical web search was illustrated in [2], [4] [5] and [6].

D. 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 technique 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 method 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 index 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) An indexer which builds hybrid index structures to integrate text and location information. (3) A ranker which ranks 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. [2]

E. Collective Spatial Keyword Querying

This technique use new type of queries, called spatial group keyword queries, that find groups of objects that collectively satisfy a query. Here two instantiations of the problem and show that both are NP-complete. This method introduce algorithms that offer approximate solutions to the two sub problems with provable approximation bounds.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

And also present exact algorithms for the two sub problems. All algorithms exploit a spatial keyword index to prune the search space. The results demonstrate that the proposals offer scalability and are capable of excellent performance.[8]

V. EXISTING SYSTEM APPROACH

Present system gives the real nearest neighbor that lies quite far away from the query location, while all the closer objects missing one or any of the keywords. This system mainly focuses on finding the nearest neighbor where each node satisfies all the query keywords. This leads to low efficiency for incremental query. The problem is Implement nearest neighbor search algorithm using for given data set and to find out closest point from give query also analyze the result fetch time and accuracy.Implement the Inverted index algorithm by extending point the nearest neighbor and forming R tree to find closest point from given set of query and also analyze the result against time and result accuracy.

Disadvantages:-

1. The IR2-tree is the first access method for answering NN queries with keywords. Although IR2 Trees gives pioneering solutions, it also has few drawbacks that affects its efficiency. The most important drawback is the result set may be empty or the number of false hits can be very large when the object of the final result is far away from query point. The query algorithm would need to load the documents of many objects, incurring expensive overhead.
2. The R-trees allow us to remedy an awkwardness in the way NN queries are processed with an I-index. Recall that, to answer a query, currently we have to first get all the points carrying all the query words. Although the distance browsing is easy with R-Trees, the best-first algorithm is exactly designed to output data points in ascending order of their distances. A serious drawback of R-Trees are its not space efficient, as the point needed to be duplicated once for every word in its description.

VI. PROPOSED SYSTEM APPROACH

The drawbacks of R-Trees and inverted index can be overcome by designing a variant of inverted index that supports compressed coordinate embedding. This system deals with searching and nearer keywordissues and database manage multidimensional objects which resulted in failure of previous systems. To deal with spatial index as searching the entered keyword and from that find the nearest results having that keyword available and showing the result of Related to particular keyword .So easier to find the location of nearer restaurant in map having the available keyword. Spatial databases manage multidimensional objects and provide quick access to those objects. The importance of spatial databases is mirrored by the convenience of modeling entities of reality in an exceedingly geometric manner.The Inverted Index is compressed by coordinate encoding which makes Spatial Inverted Index (SI-Index). Query processing with an SI-index can be done either by together or merging with R-Trees in distance browsing manner. The inverted index compress eliminates the defect of a conventional index such that an SI-index consumes much less space.

Advantages:-

Compression is already wide used technology to reduce the space of an inverted index where each inverted list contains only ids. So, the effective approach is used to record gaps with consecutive ids. Compressing an SI index is less straightforward than other approaches. For example, if we decide to sort the list by ids, gap-keeping on ids may lead to good space saving, but its application on the x- and y-coordinates would not have much effect.

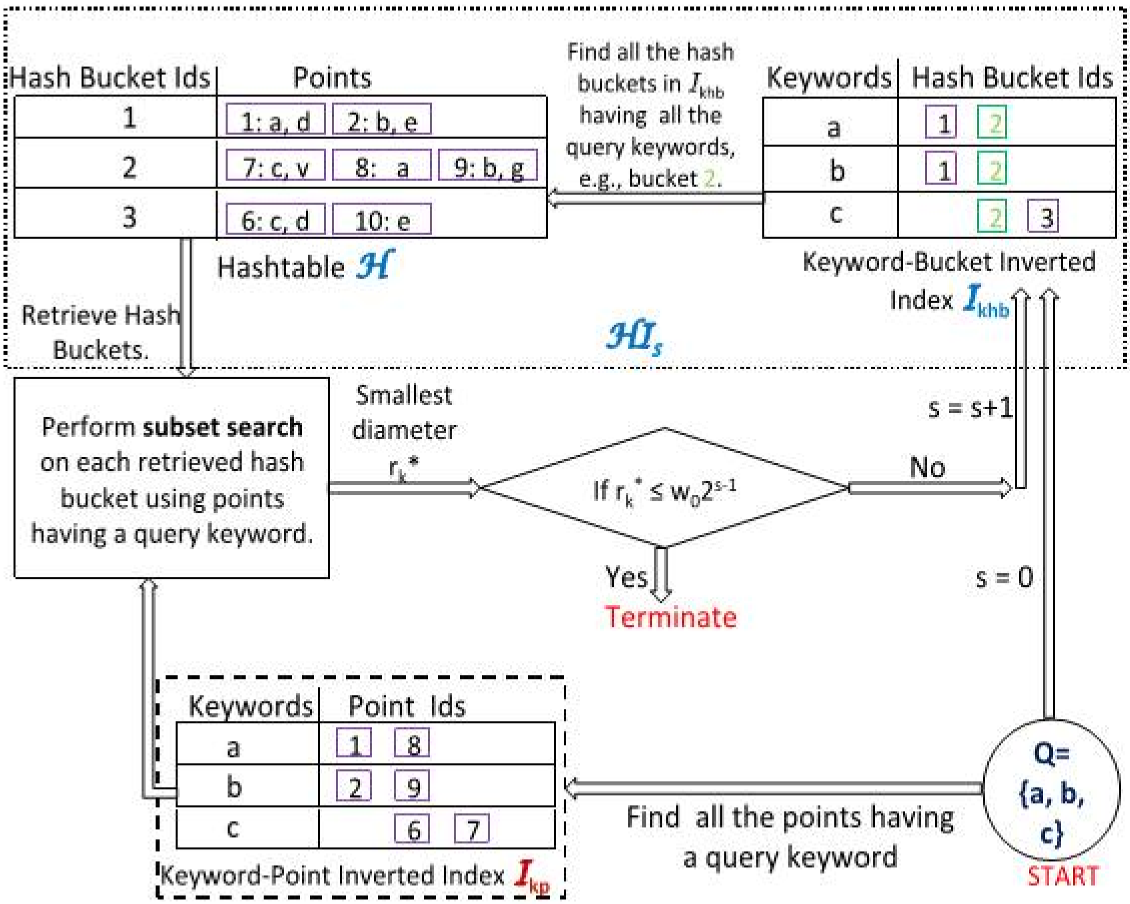

Fig No 01 . Index structure and flow of execution of ProMiSH

VII. CONCLUSIONS

In this paper, we proposed solutions to the problem of top-k nearest keyword set search in multi-dimensional datasets. We proposed a novel index called ProMiSH based on random projections and hashing. Based on this index, we developed ProMiSH-E that finds an optimal subset of points and ProMiSH-A that searches near-optimal results with betterefficiency. Our empirical results show that ProMiSH is faster than state-of-the-art tree-based techniques, with multiple orders of magnitude performance improvement. Moreover, our techniques scale well with both real and synthetic datasets.

REFERENCES

[1] D. Zhang, B. C. Ooi, and A. K. H. Tung, "Locating mapped resources in web 2.0", in Proc. IEEE 26th Int. Conf. Data Eng., 2010, pp. 521–532.
[2] V. Singh, S. Venkatesha, and A. K. Singh, "Geo-clustering of images with missing geotags", in Proc. IEEE Int. Conf. Granular Comput., 2010, pp. 420–425.
[3] V. Singh, A. Bhattacharya, and A. K. Singh, "Querying spatial patterns", in Proc. 13th Int. Conf. Extending Database Technol.: Adv. Database Technol., 2010, pp. 418–429.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

[4] X. Cao, G. Cong, C. S. Jensen, and B. C. Ooi, "Collective spatial keyword querying", in Proc. ACM SIGMOD Int. Conf. Manage Data, 2011, pp. 373–384.
[5] C. Long, R. C.-W. Wong, K. Wang, and A. W.-C. Fu, "Collective spatial keyword queries: A distance owner-driven approach", in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2013, pp. 689–700.
[6] A. Khodaei, C. Shahabi, and C. Li, "Hybrid indexing and seamless ranking of spatial and textual features of web documents", in Proc. 21st Int. Conf. Database Expert Syst. Appl., 2010, pp. 450–466.
[7] V. Singh and A. K. Singh, "SIMP: Accurate and efficient near neighbor search in high dimensional spaces", in Proc. 15th Int. Conf. Extending Database Technol., 2012, pp. 492–503.
[8] I. De Felipe, V. Hristidis, and N. Rishe, "Keyword search on spatial databases", in Proc. IEEE 24th Int. Conf. Data Eng., 2008, pp. 656–665.
[9] D. Zhang, Y. M. Chee, A. Mondal, A. K. H. Tung, and M. Kitsuregawa, "Keyword search in spatial databases: Towards searching by document," in Proc. IEEE 25th Int. Conf. Data Eng., 2009, pp. 688–699.
[10] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, "Locality sensitive hashing scheme based on p-stable distributions," in Proc. 20th Annu.Symp.Comput. Geometry, 2004, pp. 253–262.