

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

Spatial Index Keyword Search Mechanism in Multi-dimensional Database Management System

Pradeep Panchal, Prof. Shweta Joshi

M. E Student, Department of Computer Engineering, Savitribai Phule Pune University, Maharashtra, India, Flora
Institute of Technology, Pune, India

Professor, Department of Computer Engineering, Savitribai Phule Pune University, Maharashtra, India, Flora Institute
of Technology, Pune, India

ABSTRACT: Nearest neighbor look in media databases needs more help from similitude seek in question handling. Range seek and closest neighbor look depends for the most part on the geometric properties of the items fulfilling both spatial predicate and a predicate on their related writings. We do have numerous portable applications that can find craved questions by regular spatial inquiries. Ebb and flow best answer for the closest neighbor look are IR2 trees which have numerous execution bottlenecks and lacks. In this way, a novel strategy is presented in this paper with a specific end goal to expand the effectiveness of the hunt called as Spatial Inverter Index. This new SI list strategy upgrades the customary altered file plan to adapt up to high multidimensional information and alongside calculations that is perfect with the continuous catchphrase seek.

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

I. INTRODUCTION

Multi dimensional objects such as points, rectangles managed by spatial databases provides fast access to those objects based on different selection criteria. For example, location of hospitals, hotels and theatres are represented as points whereas parks, lakes and shopping malls are represented as rectangles.

a. Background

For instance, GIS range search gives all the cafes in certain area and nearest neighbor gives location of café near to our geometrical location. Today, the search engine optimization has made a realistic approach to write a spatial query in a brand new style. Some of may have few applications which finds the objects in a huge multidimensional data along with its geometrical locations and associated texts. There are easy ways to support queries that combine spatial and text features. For example, if we want to search a café whose menu contains keywords {Mocha, Espresso, Cappuccino} it would fetch all the restaurants with the keywords and from that list gives the nearest one. 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 flip kart 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.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

b. Motivation

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

II. LITERATURE SURVEY

Unfortunately there is no efficient support for top-k spatial keyword queries, where a prefix of the results list is required. Instead, current systems use ad-hoc combinations of nearest neighbor (NN) and keyword search techniques to tackle the problem. For instance, an R-Tree is used to find the nearest neighbors and for each neighbor an inverted index is used to check if the query keywords are contained. We show that such two-phase approaches are inefficient. [5]

Today, the widespread use of search engines has made it realistic to write spatial queries in a brand new way. Conventionally, queries focus on objects' geometric properties only, such as whether a point is in a rectangle, or how close two points are from each other. We have seen some modern applications that call for the ability to select objects based on both of their geometric coordinates and their associated texts. For example, it would be fairly useful if a search engine can be used to find the nearest restaurant that offers "steak, spaghetti, and brandy" all at the same time. Note that this is not the "globally" nearest restaurant (which would have been returned by a traditional nearest neighbor query), but the nearest restaurant among only those providing all the demanded foods and drinks. [3]

The World-Wide Web has reached a size where it is becoming increasingly challenging to satisfy certain information needs. While search engines are still able to index a reasonable subset of the (surface) web, the pages a user is really looking for are often buried under hundreds of thousands of less interesting results. Thus, search engine users are in danger of drowning in information. Adding additional terms to standard keyword searches often fails to narrow down results in the desired direction. A natural approach is to add advanced features that allow users to express other constraints or preferences in an intuitive manner, resulting in the desired documents to be returned among the first results. In fact, search engines have added a variety of such features, often under a special advanced search interface, but mostly limited to fairly simple conditions on domain, link structure, or modification date. [1]

A spatial keyword query consists of a query area and a set of keywords shown in below figure. The answer is a list of objects ranked according to a combination of their distance to the query area and the relevance of their text description to the query keywords. A simple yet popular variant, which is used in our running example, is the distance-first spatial keyword query, where objects are ranked by distance and keywords are applied as a conjunctive filter to eliminate objects that do not contain them. [7]

III. 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 k nearest neighbor search algorithm using for given data set and to find out closest point from given query also analyzes.

The results fetch time and accuracy. Implement the Inverted index algorithm by extending point the k 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:-

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.

The R-trees allow us to remedy 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.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

IV. PROPOSED SYSTEM APPROACH

1. 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 location issues 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 location having that keyword available and showing the location of restaurant having menus available in map. 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.

V. SYSTEM ARCHITECTURE

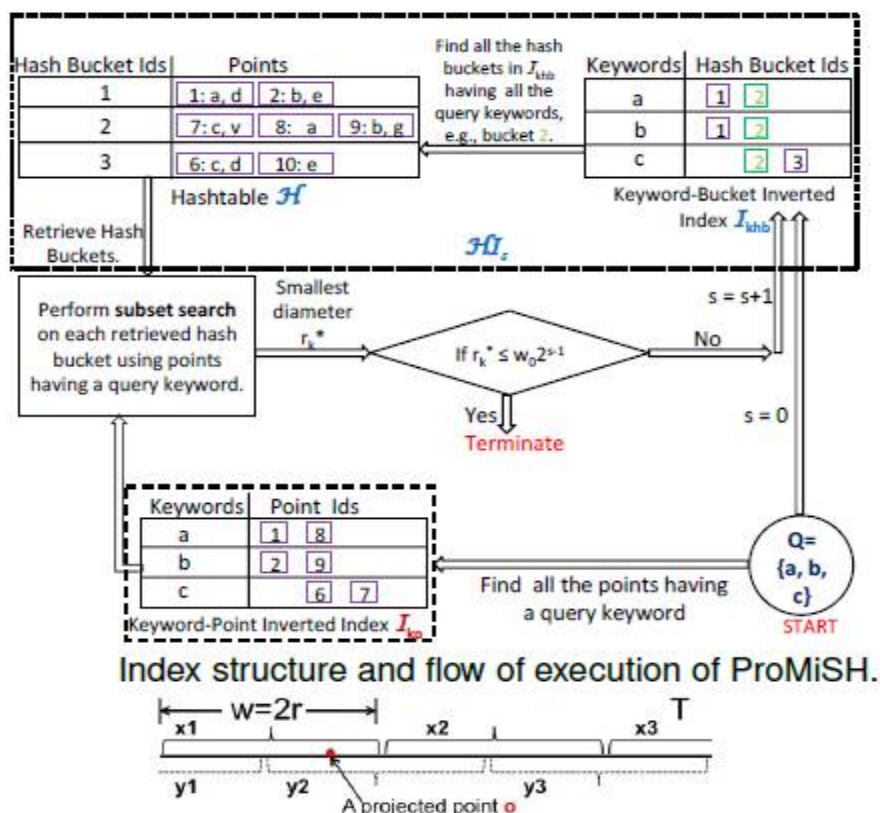


Fig No 01 System Architecture

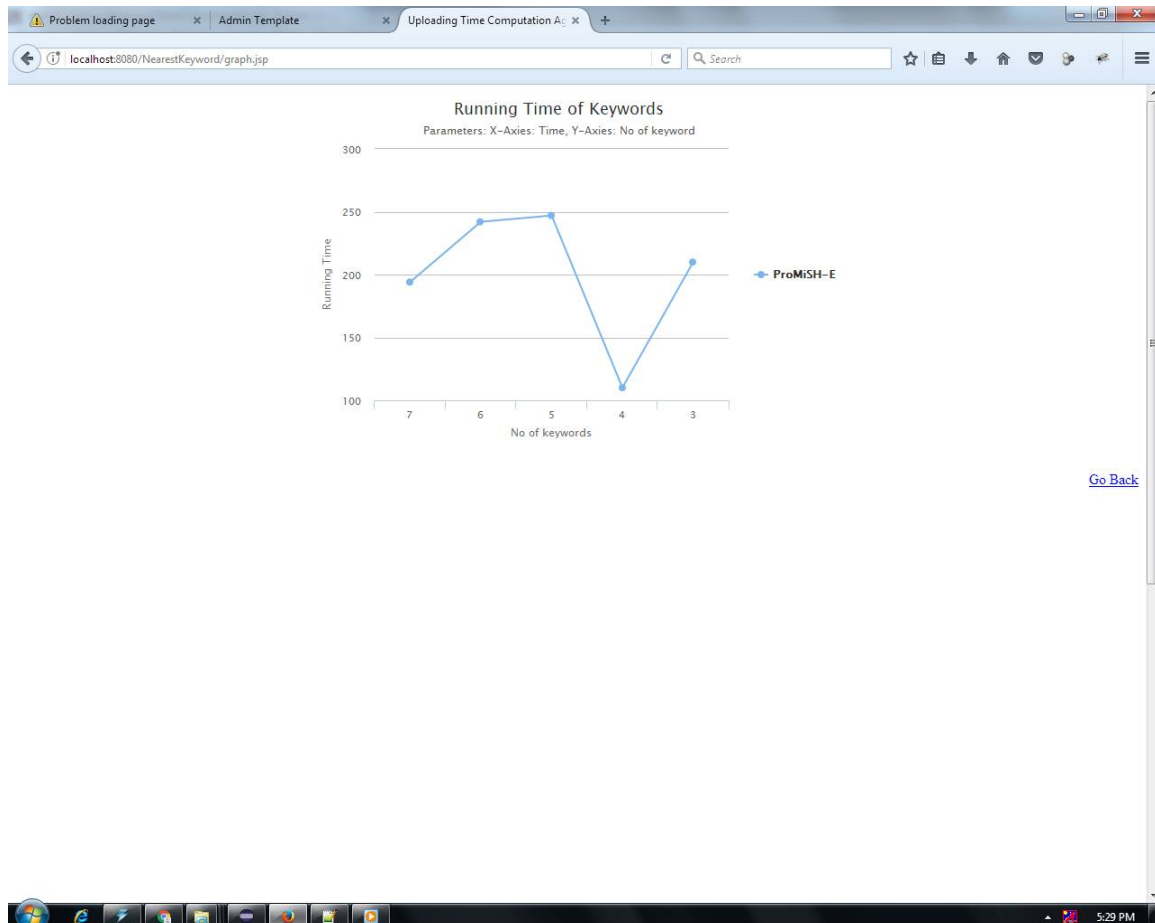
International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

VI. EXPERIMENTAL SET UP



Result Table:-

No. of keyword Ids	Running Time
7	194
6	242
5	247
4	110
3	210

VI. CONCLUSION

They are numerous number of applications of with a search engine which efficiently support novel forms of spatial queries integrated with keyword search. By all the above methods, the main goal is searching a relevant keyword with appropriate info with minimum time and with valid results. In this paper, we come to a conclusion by developing an access method called Spatial Inverted Index (SI-Index). SI Index has high space efficiency and also has the ability to perform keyword augmented NN search in time.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

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.
- [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.