Fast Nearest Neighbour Search with Keywords using Spatial Inverted Index

Sayyed Muzammil Ali[1], Dr. A. S. Alvi[2]

[1]Department Computer Science & Engineering & S.G.B.A.U, India
[2]Department Computer Science & Engineering & S.G.B.A.U, India
[1] smali.muzammil@gmail.com

Abstract: Spatial queries, such as range search and nearest neighbor retrieval, involve only conditions on objects geometric properties. A spatial database manages multidimensional objects (such as points, rectangles, etc.), and provides fast access to those objects based on different selection criteria. Now-a-days many applications call a new form of queries to find the objects that satisfying both a spatial predicate, and a predicate on their associated texts. For example, instead of considering all the restaurants, a nearest neighbor query would instead ask for the restaurant that is the closest among those whose menus contain the specified keywords all at the same time. Currently the best solution to such queries is based on the IR2-tree, which, as shown in this paper, has a few deficiencies that seriously impact its efficiency. Motivated by this, we develop a new access method called the spatial inverted index that extends the conventional inverted index to cope with multidimensional data, and comes with algorithms that can answer nearest neighbor queries with keywords in real time. As verified by experiments, the proposed techniques outperform the IR2-tree in query response time significantly, often by a factor of orders of magnitude.

Keywords— Information Retrieval Tree, Keyword Search, Spatial Inverted Index.

I. INTRODUCTION

A. Introduction to area:

Data mining is a new powerful technology with great potential to help companies focus on the most important Information in their data warehouses. It has been defined as the fast analysis of large or complex data sets in order to discover significant patterns or trends that would otherwise go unrecognized.

The term data mining is appropriately named as 'Knowledge mining from data' or "Knowledge mining". It is the nontrivial extraction of implicit, previously unknown and potentially useful information from data in database. Data mining is the process of analysing data from different perspectives and summarizing it into useful information –information that can be used to increase revenue, cuts costs, or both. Data mining software is allows users to analyse data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. As the data are available in the different formats so that the proper action to be taken. Not only to analyse these data but also take a good decision and maintain the data .As and when the customer will require the data should be retrieved from the database and make the better decision. The important reason that attracted a great deal of attention in information technology the discovery of useful information from

large collections of data industry towards field of "Data mining" is due to the perception of *"we are data rich but information poor"*.

B. Basic of topics:

A spatial database is used to store large amount of space related data such as maps, medical imaging data etc. and manages multidimensional objects and provides quick access to those objects based on different choice criteria. The importance of spatial databases is it provides a convenient way to model the entities of reality in a geometric manner. As an example, locations of restaurants, hotels, hospitals and then on are typically shown as points in an exceedingly map, whereas larger extents like parks, lakes, and landscapes typically as a mixture of rectanglesThe 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. as an example, in an exceedingly geographic system, vary search will be deployed to search out all restaurants in an exceedingly sure space, whereas nearest neighbour 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.

C. Aim and Objectives:

To design a variant of inverted index that is optimized for multidimensional points. This access method successfully incorporates point coordinates into a conventional inverted index with small extra space, owing to a delicate compact storage scheme. To provide efficiently a support for novel forms of spatial queries that are integrated with keyword search.

II. LITERATURE SURVEY

1) Previous Research work:

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 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. Spatial queries with keywords have not been extensively explored. In the past years, the community has sparked enthusiasm in studying keyword search in relational databases. It is until recently that attention was diverted to multidimensional data [12], [13].

The best method to date for nearest neighbour search with keywords is due to Felipe et al. [12]. They nicely integrate two well-known concepts: *R-tree* [2], a popular spatial index, and *signature file* [11], an effective method for keyword-based document retrieval. By doing so they develop a structure called the *IR2-tree* [12], which has the strengths of both R-trees and signature files. Like R-trees, the IR2-tree preserves objects' spatial proximity, which is the key to solving spatial queries efficiently. On the other hand, like signature files, the IR2-tree is able to filter a considerable portion of the objects that do not contain all the query keywords, thus significantly reducing the number of objects to be examined.

DISADVANTAGES OF EXISTING SYSTEM:
• Fail to provide real time answers on difficult inputs.
• The real nearest neighbour lies quite far away from the query point, while all the closer neighbours are missing at least one of the query keywords.

2) Proposed System and Its Advantages:

In this project, we design a variant of inverted index that is optimized for multidimensional points, and is thus named the spatial inverted index (SI-index). This access method successfully incorporates point coordinates into a conventional inverted index with small extra space, owing to a delicate compact storage scheme. Meanwhile, an SI-index preserves the spatial locality of data points, and comes with an R-tree built on every inverted list at little space overhead.

As a result, it offers two competing ways for query processing.
• We can (sequentially) merge multiple lists very much like merging traditional inverted lists by ids.
• Alternatively, we can also leverage the R-trees to browse the points of all relevant lists in ascending order of their distances to the query point.

III.SYSTEM DESIGN

Design is one of the most important phases of software development. The design is a creative process in which a system organization is established that will satisfy the functional and non-functional system requirements. Large Systems are always decomposed into sub-systems that provide some related set of services. The design of the project consists of the data flow between the functionalities, the main objects of the system functions.

3.1Data Flow Diagram

The dataflow diagram shows the flow of data between the system functions and process between users. Here shows the process between the admin, system and enduser.
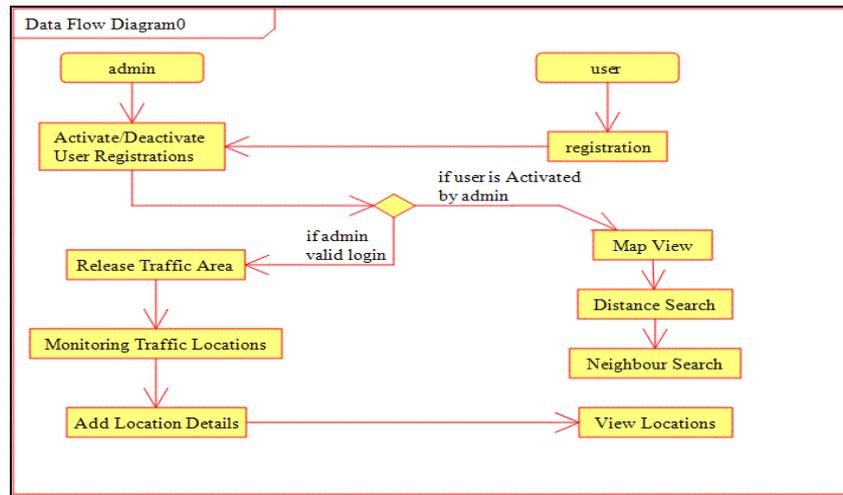

Fig a: data flow diagram

3.2 Use case Diagram

It is a type of behavioral diagram defined and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionalities provided by a system in terms of actors and their goals. The main purpose of use case diagram is to show what system functions are performed for which actors.
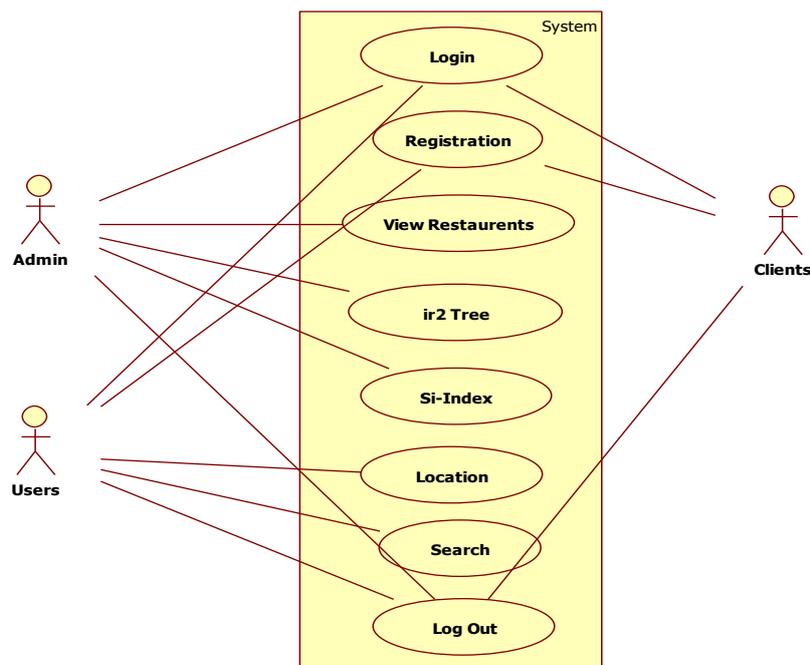

Fig b: use-case diagram

543

3.3Class Diagram: Class diagrams are widely used to describe the types of objects in a system and their relationships. Class diagrams model class structure and contents using design elements such as classes, packages and objects. Class diagrams describe three different perspectives when designing a system, conceptual, specification and implementation.
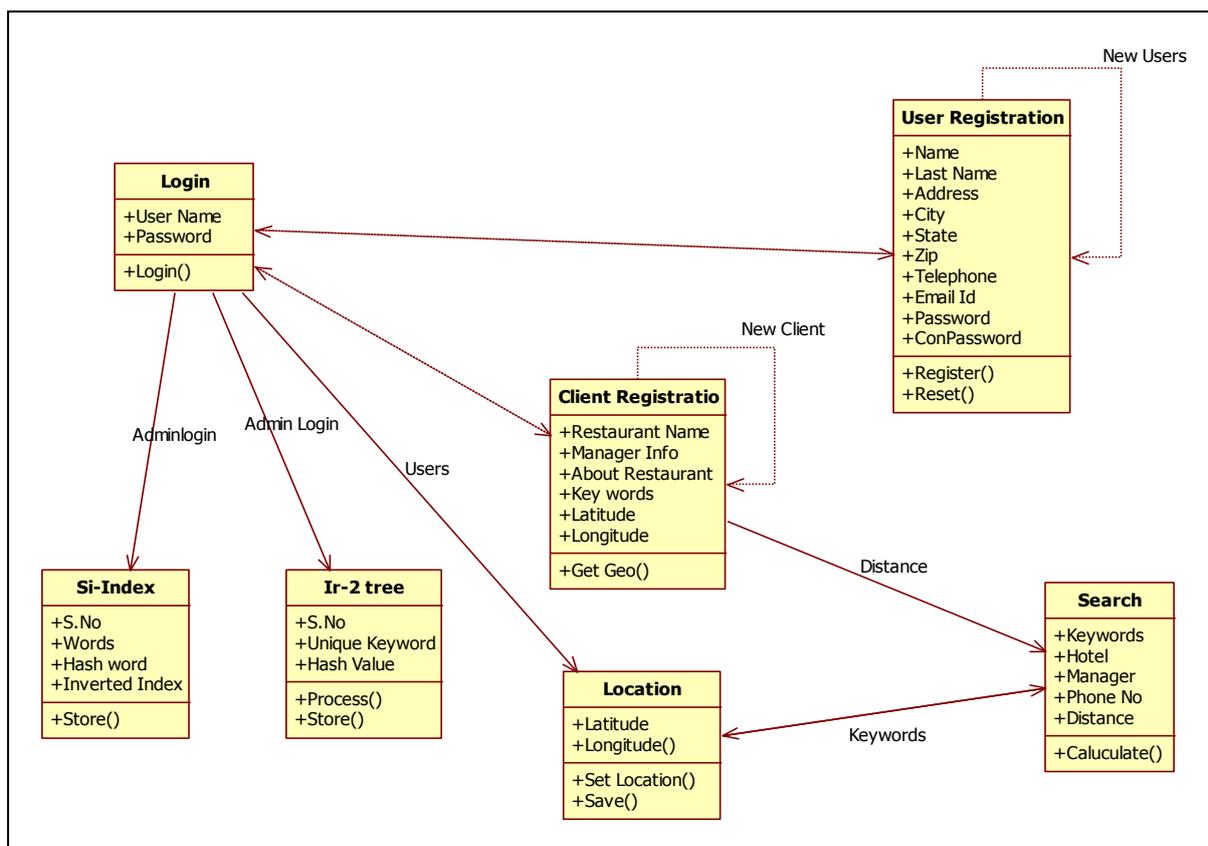
Fig c: class diagram

IV.IMPLEMENTATION

4.1Registration: In this module a User have to register first, then only he/she has to access the data base.

4.2Login: In this module, any of the above mentioned person have to login, they should login by giving their email id and password.

4.3Hotel Registration: In this module Admin registers the hotel along with its famous dish. Also he measures the distance of the corresponding hotel from the corresponding source place by using spatial distance of Google map.

4.4Search Techniques: Here we are using two techniques for searching the document 1) Restaurant Search, 2) Key Search.

Key Search: It means that the user can give the key in which dish that the restaurant is famous for .This results in the list of menu items displayed.

4.4.2Restaurant Search: It means that the user can have the list of restaurants which are located very near. List came from the database.

4.4.3Map View: The User can see the view of their locality by Google Map (such as map view, satellite view)

4.5Distance Search: The User can measure the distance and calculate time that takes them to reach the destination by giving speed.

CONCLUSIONS

Fast Nearest Neighbor Search with Keywords using spatial inverted index is extremely effective for searching nearest restaurant from user with expected menus. It does this by IR2 tree algorithm- Compression, Merging and Distance Browsing, and GPS System. In this we can add features like After selecting Hotel it will display menu card of that Hotel Implement this application for PC's and Desktops.References. We have seen plenty of applications calling for a search engine which is able to efficiently support novel forms of spatial queries that are integrated with keyword search. In this project we have developed an access method called the Spatial Inverted Index (SI-Index). This method is fairly space economical and it has ability to perform keyword augmented nearest neighbor search in real time. This method is based on conventional technology of Inverted Index. It is readily incorporable in a commercial search engine.

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), pages 5–16, 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), pages 322–331, 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), pages 431–440, 2002.

[4] X. Cao, L. Chen, G. Cong, C. S. Jensen, Q. Qu, A. Skovsgaard, D. Wu, and M. L. Yiu. Spatial keyword querying. In ER, pages 16–29, 2012.

[5] X. Cao, G. Cong, and C. S. Jensen. Retrieving top-k prestige-based relevant spatial web objects. PVLDB, 3(1):373– 384, 2010.

[6] X. Cao, G. Cong, C. S. Jensen, and B. C. Ooi. Collective spatial keyword querying. In Proc. of ACM Management of Data (SIG- MOD), pages 373–384, 2011.

[7] 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), pages 30–39, 2004.

[8] Y.-Y. Chen, T. Suel, and A. Markowetz. Efficient query processing in geographic web search engines. In Proc. Of ACM Management of Data (SIGMOD), pages 277–288, 2006.

[9] 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.

[10] G. Cong, C. S. Jensen, and D. Wu. Efficient retrieval of the top-k most relevant spatial web objects. PVLDB, 2(1):337–348, 2009.

[11] C. Faloutsos and S. Christodoulakis. Signature files: An access method for documents and its analytical performance evaluation. ACM Trans- actions on Information Systems (TOIS), 2(4):267–288, 1984.

[12] I. D. Felipe, V. Hristidis, and N. Rishe. Keyword search on spatial databases. In Proc. of International Conference on Data Engineering (ICDE), pages 656–665, 2008.

[13] R. Hariharan, B. Hore, C. Li, and S. Mehrotra. Processing spatial keyword (SK) queries in geographic information retrieval (GIR) systems. In Proc. of Scientific and Statistical Database Management (SSDBM), 2007.

[14] G. R. Hjaltason and H. Samet. Distance browsing in spatial databases. ACM Transactions on Database Systems (TODS), 24(2):265–318, 1999.

[15] V. Hristidis and Y. Papakonstantinou. Discover: Keyword search in relational databases. In Proc. of Very Large Data Bases (VLDB), pages 670–681, 2002.

[16] I. Kamel and C. Faloutsos. Hilbert R-tree: An improved r-tree using fractals. In Proc. of Very Large Data Bases (VLDB), pages 500–509, 1994.

[17] Yufei Tao and Cheng Sheng, "Fast Nearest Neighbor Search with Keywords", IEEE transactions on knowledge and data engineering, VOL. 26, NO. 4, APRIL 2014.

[18] 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.

[19] 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.