Proceedings of the International Conference , "Computational Systems for Health & Sustainability"
17-18, April, 2015 - byR.V.College of Engineering,
Bangalore,Karnataka,PIN-560059,INDIA

Best Results for Spatial Queries Using Fast Nearest Neighbor Search with Keywords

Ranjani B. S.
M.Tech 4th sem, SJBIT, Bangalore

PavitraBai S
Asst.Prof, SJBIT, Bangalore.

Abstract: Conventional spatial queries, such as range search and nearest neighbour retrieval, involve only conditions on objects' geometric properties. Today, many modern applications call for novel forms of queries that aim to find objects satisfying both a spatial predicate, and a predicate on their associated texts. For example, instead of considering all the restaurants, a nearest neighbour query would instead ask for the restaurant that is the closest among those whose menus contain "steak, spaghetti, brandy" 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 neighbour 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.

Index Terms: Spatial queries, IR2 trees, predicate,spatial inverted index,multidimensional data

INTRODUCTION

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 neighbour retrieval can discover the restaurant closest to a given address.

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 coordinatesand 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 neighbour query), but the nearest restaurant among only those providing all the demanded foods and drinks.

There are easy ways to support queries that combine spatial and text features. For example, for the above query, we could first fetch all the restaurants whose menus contain the set of keywords {steak, spaghetti, brandy}, and then from the retrieved restaurants, find the nearest one. Similarly, one could also do it reversely by targeting first the spatial conditions – browse all the restaurants in ascending order of their distances to the query point until encountering one whose menu has allthe keywords. The major drawback of these straightforward approaches is that they will fail to provide real time answers on difficult inputs. A typical example is that the real nearest neighbour lies quite faraway from the query point, while all the closer neighbors are missing at least one of the query keywords.

LITERATURE SURVEY

2.1 DBXplorer: A System for Keyword-Based Search over Relational Databases

Internet search engines have popularized the keyword based search paradigm. While traditional database management systems offer powerful query languages, they do not allow keyword-based search. In this paper, we discuss DBXplorer, a system that enables keyword based search in relational databases. DBXplorer has been implemented using a commercial relational database and web server and allows users to interact via a browser front-end. We outline the challenges and discuss the implementation of our system including results of extensive experimental evaluation.

2.2 Efficient Query **Processing in Geographic Web Search Engines**

Geographic web search engines allow users to constrain and order search results in an intuitive manner by focusing a query on a particular geographic region. Geographic search technology, also called local search, has recently received significant interest from major search engine companies. Academic research in this area has focused primarily on techniques for extracting geographic knowledge from the web. In this paper, we study the problem of efficient query processing in scalable geographic search engines. Query processing is a major bottleneck in standard web search engines, and the main reason for the thousands of machines used by the major engines.

2.3 Retrieving Top k Prestige Based Relevant Spatial Web Objects

The location-aware keyword query returns ranked objects that are near a query location and that have textual descriptions that match query keywords. This query occurs inherently in many types of mobile and traditional web services and applications, e.g., Yellow Pages and Maps services. Previous work considers the potential results of such a query as being independent when ranking them. However, a relevant result object with nearby objects that are also relevant to the query is likely to be preferable over a relevant object without relevant nearby objects.

2.4The R-tree: An Efficient and Robust Access Method for Points and Rectangles*

The R-tree, one of the most popular access methods for rectangles, IS based on the heuristic optimization of the area of the enclosing rectangle in each inner node By running numerous experiments m a standardized test bed under highly varying data, queries and operations, we were able to design the R*-tree which incorporates a combined optimization of area, margin and overlap of each enclosing rectangle m the directory Using our standardized test bed m an exhaustive performance comparison, It turned out that the R*-tree clearly outperforms the existing R-tree variants Guttman's linear and quadratic R-tree and Greene's variant of the R-tree This superiority of the R*-tree holds for different types of queries and operations, such as map overlay.

IR2- TREES

The IR2-treecombines the Rtree with signature files. Signature file in general refers to a hashing-based framework, whose instantiation in [12] is known as superimposed coding (SC), which is shown to be more effective than other instantiations [11]. It is designed to perform membership tests: determine whether a query word w exists in a set W of words. SC is conservative, in the sense that if it says "no", then w is definitely not in W. If, on the other hand, SC returns "yes", the true answer can be either way, in which case the whole W must be scanned to avoid a false hit

3.1Building R-trees

The goal is to let each block of an inverted list be directly a leaf node in the R-tree. This is in contrast to the alternative approach of building an R-tree that shares nothing with the inverted list, which wastes space by duplicating each point in the inverted list. Further more, our goal is to offer two search strategies simultaneously.

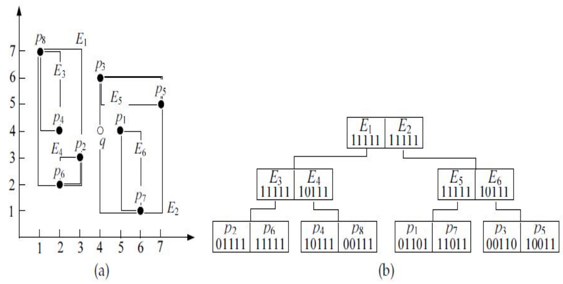

Fig b: (a) shows the R-tree and (b) gives the signatures of the entries

SPATIAL INVERTED INDEX

The spatial inverted list (SI-index) is essentially a compressed version of an I-index with embedded coordinates. Query processing with an SI-index can be done either by merging, or together withR-trees in a distance browsing manner. Furthermore, the compression eliminates the defect of a conventional that an SI-index consumes much less space.

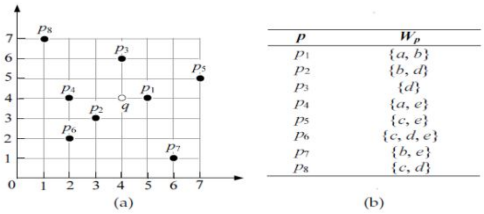

Fig. a: (a) shows the locations of points and (b) gives their associated texts.

DESIGN

The design of the project consists of the data flow between the functionalities, the main objects of the system functions.

5.1 Data 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.

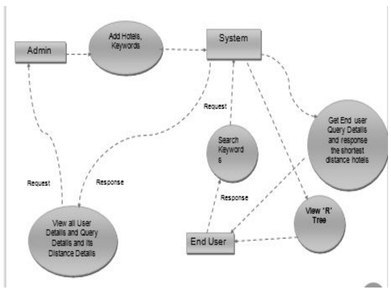

Fig:c data flow diagram

5.2 Use case Diagram

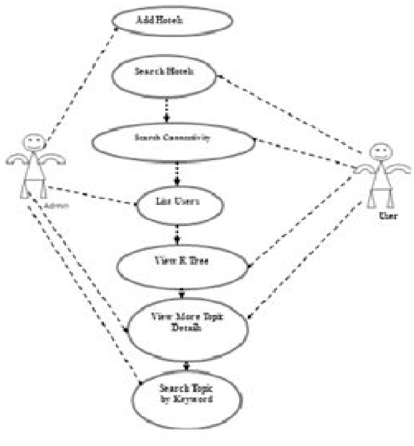

Fig d: use-case diagram

It is a type of behavioural 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.

ISSN 2320 –5547
IJITR International Journal of Innovative Technology and Research

5.3 Class Diagram

The class diagram represents the main objects interactions in the main objects, interactions in the applications and the class to be programmed. In the design of a system a number of class identified and grouped together in a class diagram which helps to determine the static relations between those objects with detailed modelling.In the diagram, classes are represented with boxes which contain three partsThe upper part holds the name of the classThe middle part contains the attributes of the classThe bottom part gives the methods oroperations the class can take or undertake.

In the design of a system, a number of classes are identified and grouped together in a class diagram which helps to determine the static relations between those objects. With detailed modelling, the classes of the conceptual design are often split into a number of subclasses.

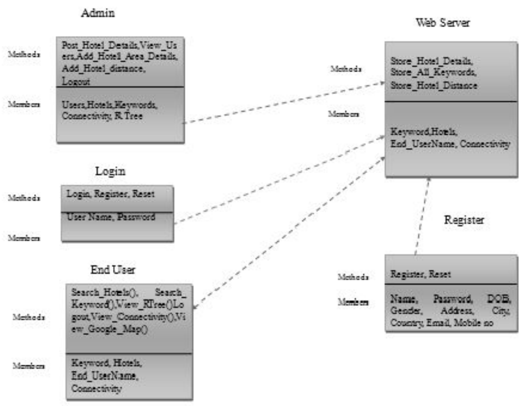

Fig e : class diagram

IMPLEMENTATION

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

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

6.3 Hotel 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.

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

6.5 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.

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

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

6.8 Distance Search: The User can measure the distance and calculate time that takes them to reach the destination by giving speed. Chart will be prepared by using these values. These are done by the use of Google Maps.

CONCLUSION

We have seen plenty of applications calling for a search engine that is able to efficiently support novel forms of spatial queries that are integrated with keyword search. The existing solutions to such queries either incur prohibitive space consumption or are unable to give real time answers. In this paper, we have remedied the situation by developing an access method called the spatial inverted index (SI-index). Not only that the SI-index is fairly space economical, but also it has the ability to perform keyword-augmented nearest neighbour search in time that is at the order of dozens of milli-seconds. Furthermore, as the SI-index is based on the conventional technology of inverted index, it is readily incorporable in a commercial search engine that applies massive parallelism, implying its immediate industrial merits.

REFERENCES

[1]. S. Agrawal, S. Chaudhuri, and G. Das.Dbxplorer: A system for keyword-based search over relational databases. In Proc. OfInternational 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 Managementof Data (SIGMOD), pages 277–288, 2006.