International Journal of Science Engineering and Advance Technology, IJSEAT, Vol 3, Issue 1
ISSN 2321-6905
January-2015

International Journal of Science Engineering and Advance Technology

Fast Nearest Neighbour Search Aims At Optimizing Different Objective Functions Using Si- Index

Ravikiran Borugadda[1] K.Chandrasekhar Reddy[2]
[1] Pursuing M.Tech, Asst. Professor[2]
Dept of CSE from Vikas College of Engineering and Technology, Nunna, Vijayawada.
Affiliated to JNTU-Kakinada, A.P., India

Abstract:
Several modern applications call for novel forms of queries that aspire to find objects pleasing both a spatial predicate and a predicate on their associated texts. The significance of spatial databases is reflected by the convenience of modelling entities of reality in a geometric manner. For instance locations of restaurants, hotels, hospitals and so on are often represented as points in a map at the same time as 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 case in point 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. 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.

Keywords: Nearest neighbour search, keyword search, spatial index.

Introduction:
The widespread use of search engines has made it realistic to write spatial queries in a brand new way. Conservatively 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 aptitude 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 but the nearest restaurant among only those providing all the demanded foods and drinks. The best solution to such queries is based on the IR2-tree has a few deficiencies that seriously impact its efficiency. As verified by experiments the proposed techniques outperform the IR2-tree in query response time significantly often by a factor of orders of magnitude. 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 schemes.

Related Work:
Signature file in general refers to a hashing-based framework, whose instantiation is known as superimposed coding (SC) which is shown to be more effectual than other instantiations. It is designed to perform membership tests decide 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. SC works in the same way as the classic technique of bloom filter. In pre-processing it builds a bit signature of length l from W by hashing each word in W to a string of l bits and then taking the disjunction of all bit strings.

Exisitng Method:
Spatial queries with keywords have not been expansively explored. The community has sparked enthusiasm in studying keyword search in relational databases. It is until recently that attention was diverted to multidimensional data. The best method to date for nearest neighbour search with keywords is due to Felipe et al. They nicely integrate two well-known concepts: R-tree, a popular spatial index and signature file an effective method for keyword-based document retrieval. By doing so they develop a structure called the IR2 - tree 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.

Disadvantages:
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.

Proposed System:

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 subtle compact storage scheme. In the meantime 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. We can successively merge multiple lists very much like merging traditional inverted lists by ids. Otherwise 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. The SI-index significantly outperforms the IR2 -tree in query efficiency often by a factor of orders of magnitude.

Advantages:

Distance browsing is easy with R-trees. In fact the best-first algorithm is precisely designed to output data points in ascending order of their distances. It is straight forward to extend our compression scheme to any dimensional space.

Registration:

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

Login:

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

Hotel Registration:

Admin registers the hotel along with its famous dish. Also this method is the distance of the corresponding hotel from the corresponding source place by using spatial distance of Google map.

Search Techniques:

We use Restaurant Search and Key Search for searching the document. In Key Search the user can give the key in which dish that the restaurant is famous for .This results in the list of menu items displayed whereas the Restaurant Search the user can have the list of restaurants which are located very near. List comes from the database.

Map View:

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

Distance Search:

The User can compute the distance and work out 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.

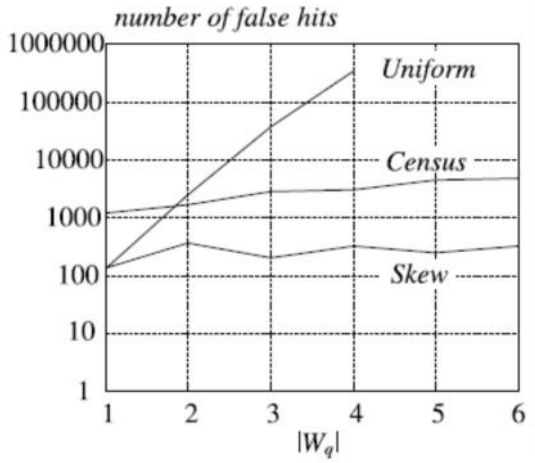
Merging And Distance Browsing:

There is a simple way to do so in an I-index. One only requirement to stock up the coordinates of each point mutually with each of its appearances in the inverted lists. The presence of coordinates in the upturned lists obviously motivates the creation of an R-tree on each list indexing the points therein. The R-trees consent to us to remedy gracelessness in the way NN queries are processed with an I-index. Distance browsing is easy with R-trees. In fact the best-first algorithm is faithfully considered to output data points in ascending order of their distances to q.

SPATIAL INVERTED LIST:

The spatial inverted list (SI-index) is fundamentally a packed in version of an I-index with embedded coordinates. Query processing with an SI-index can be done either by merging or together with R-trees in a distance browsing manner. Besides the compression do away with the defect of a conventional I-index such that an SI-index gets through much fewer space.

Experimental Results:



The shortage of IR2-tree is chiefly caused by the need to authenticate a vast number of false hits. The graph contrives the average false hit number per query as a function of jWqj. We see an exponential escalation of the number on Uniform and Census which elucidate the extreme explosion of the query cost on those data sets. Interesting is that the number of false hits fluctuates6 a little on Skew which explains the fluctuation in the cost of IR2- tree.

Conclusion:

We have answered the circumstances by initial 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 capability to carry out keyword-augmented nearest neighbour search in time that is at the order of dozens of mille-seconds. Additionally as the SI-index is based on the conventional technology of inverted index. It is willingly incorporable in a commercial search engine that applies massive parallelism, implying its immediate industrial merits. The existing solutions to such queries either acquire prohibitive space consumption or are powerless to give real time answers.

References:
[1] S. Agrawal, S. Chaudhuri, and G. Das, "Dbxplorer: A System for Keyword-Based Search over Relational Databases," Proc. Int'l Conf. Data Eng. (ICDE), pp. 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," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 322-331, 1990.
[3] G. Bhalotia, A. Hulgeri, C. Nakhe, S. Chakrabarti, and S. Sudarshan, "Keyword Searching and Browsing in Databases Using Banks," Proc. Int'l Conf. Data Eng. (ICDE), pp. 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," Proc. 31st Int'l Conf. Conceptual Modeling (ER), pp. 16-29, 2012.
[5] X. Cao, G. Cong, and C.S. Jensen, "Retrieving Top-k Prestige- Based Relevant Spatial Web Objects," Proc. VLDB Endowment, vol. 3, no. 1, pp. 373-384, 2010.
[6] X. Cao, G. Cong, C.S. Jensen, and B.C. Ooi, "Collective Spatial Keyword Querying," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 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," Proc. Ann. ACM-SIAM Symp. Discrete Algorithms (SODA), pp. 30- 39, 2004.
[8] Y.-Y. Chen, T. Suel, and A. Markowetz, "Efficient Query Processing in Geographic Web Search Engines," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 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," Proc. ACM SIGMOD Int'l Conf. Management of Data, 2009.
[10] G. Cong, C.S. Jensen, and D. Wu, "Efficient Retrieval of the Top-k Most Relevant Spatial Web Objects," PVLDB, vol. 2, no. 1, pp. 337- 348, 2009.
[11] C. Faloutsos and S. Christodoulakis, "Signature Files: An Access Method for Documents and Its Analytical Performance Evaluation," ACM Trans. Information Systems, vol. 2, no. 4, pp. 267-288, 1984.
[12] I.D. Felipe, V. Hristidis, and N. Rishe, "Keyword Search on Spatial Databases," Proc. Int'l Conf. Data Eng. (ICDE), pp. 656-665, 2008.
[13] R. Hariharan, B. Hore, C. Li, and S. Mehrotra, "Processing Spatial- Keyword (SK) Queries in Geographic Information Retrieval (GIR) Systems," Proc. Scientific and Statistical Database Management (SSDBM), 2007.
[14] G.R. Hjaltason and H. Samet, "Distance Browsing in Spatial Databases," ACM Trans. Database Systems, vol. 24, no. 2, pp. 265-318, 1999.
[15] V. Hristidis and Y. Papakonstantinou, "Discover: Keyword Search in Relational Databases," Proc. Very Large Data Bases (VLDB), pp. 670-681, 2002.

Author Profile

RAVIKIRAN BORUGADDA Pursuing M.Tech (CSE)
Vikas College of Engineering and Technology Nunna, Vijayawada. Affiliated to JNTU-Kakinada, A.P., India

K.Chandra sekhar Reddy Is Working as Assistant Professor Vikas College of Engineering And Technology Nunna, Vijayawada. Affiliated To JNTU-Kakinada, A.P., India