International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

IJIEMR Transactions, online available on 16th Febraury 2018. Link :
http://www.ijiemr.org/downloads.php?vol=Volume-7&issue=ISSUE-02

Title: Predicting Neighbour set Keyword search in Multidimensional Dataset.

Volume 07, Issue 02, Page No: 420 – 425.

Paper Authors

*MUNDLAMURI RENUKA, B.JYOTHI.

* Dept of CSE, Krishnaveni Engineering College for Women.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

PREDICTING NEIGHBOUR SET KEYWORD SEARCH IN MULTIDIMENSIONAL DATASET

*MUNDLAMURI RENUKA, **B.JYOTHI
PG Scholar, Dept of CSE, Krishnaveni Engineering College for Women ,Kesanupalli, narasaraopet
Asst Professor, Dept of CSE, Krishnaveni Engineering College for Women ,Kesanupalli, narasaraopet

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. A novel method is introduced in this paper in order to increase the efficiency of the search called as Spatial Inverter Index. This new SI index method enhances the conventional inverted index scheme to cope up with high multidimensional data and along with algorithms is compatible with the real time keyword search. We develop a new access method Preference Tree with user preference. In Preference Tree time and distance is using as parameters.We develop another method for without user preferences by Z curve method. The proposed techniques outperform in query response time significantly. Spatial databases are mainly focus on multi dimensional database. In our approach we are handling the spatial queries jointly and returns the only user specified number of optimal results, we implemented a cache based approach for efficient results.

Index Terms: Nearest Neighbor Search, Keyword Search, and Spatial Index, Querying, multi-dimensional data, IR2 Trees, similarity search

1. INTRODUCTION

Multi-dimensional objects such as points, rectangles managed by spatial databases provide fast access to those objects based on different selection criteria. Location of hospitals, hotels and theatres are represented as points whereas parks, lakes and shopping malls are represented as rectangles [2]. For instance, GIS range search gives all the cafes in certain area and nearest neighbor gives location of café near to our geometrical location. Some of may have few applications which finds the objects in a huge multidimensional data along with its geometrical locations and associated texts [1]. 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 [3].We propose Preference Tree to enable fast processing for NKS queries. In particular we develop an exact user preference and without user preference that always retrieves the optimal top-k results. Preference Tree uses a set of hash tables and inverted indexes to perform a localized search. The hashing technique is inspired by Locality Sensitive Hashing, which is a state-of-the-art method for nearest neighbor search in high-dimensional spaces. Unlike LSH-based methods that allow only approximate search with probabilistic guarantees, the index structure in Preference Tree supports accurate search [4]. We evaluate the performance of Preference Tree and Z curve method on both real and synthetic datasets. The empirical

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

results reveal that Preference Tree will efficiently work than Z curve method in the basis of user preference otherwise considering time time as parametrer Z curve method will work more faster than preference tree[5].

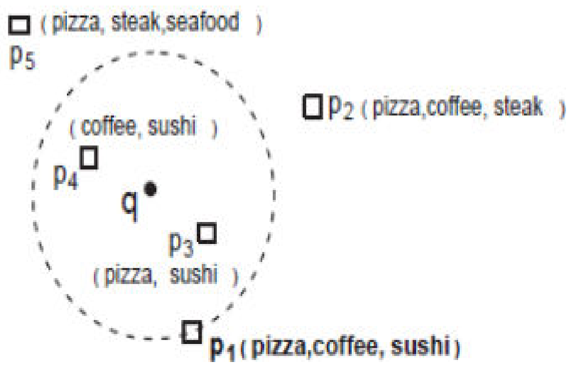

Figure 1: Online Yellow Pages Example

2. RELATED WORK

Inverted indexes (I-index) have proved to be an effective access method for keyword-based document retrieval. In the spatial context, nothing prevents us from treating the text description Wp of a point p as a document andthen building an I-index [6]. Note that the list of each word maintains a sorted order of point ids, which provides considerable convenience in query processing by allowing an efficient merge step. With the ever-increasing popularity of services such as Google Earth and Yahoo Maps, as well as other geographic applications, queries in spatial databases have become increasingly important in recent years. Novel type of query called the mclosest keywords (mCK) query: given m keywords provided by the user, the mCK query aims at finding the closest tuples (in space) that match these keywords. While such a query has various applications, our main interest lies in that of a search by document. The bR*- tree [7] is an extension of the R*-tree. Besides the node MBR, each node is augmented with additional information. A straightforward extension is to summarize the keywords in the node. With this information, it becomes easy to decide whether m query keywords can be found in this node. There are many applications which are quality sensitive and need to efficiently and accurately support near neighbor queries for all query ranges. propose a novel indexing and querying scheme called Spatial Intersection and Metric Pruning (SIMP)[8].

3. PROPOSED SYSTEM

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 [9]. 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[10]. Compression is already wide used technology to reduce the space of an inverted index where each inverted list contains only ids

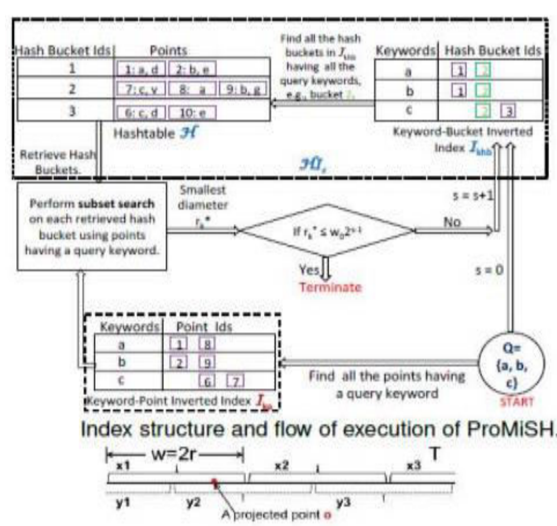

Fig No 2 Proposed System Architecture

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. Spatial databases manage multidimensional objects and provide quick access to those objects [11]. The importance of spatial databases is mirrored by the convenience of modeling entities of reality in an exceedingly geometric manner[12].

4. ALGORITHMS

Spatial inverted index Algorithm:

Inverted indexes (I-index) have proved to be an effective access method for keyword-based document retrieval. In the spatial context nothing prevents us from treating the text description Wp of a point p as a document, and then, building an I-index [13]. The index for the dataset each word in the vocabulary has an inverted list enumerating the ids of the points that have the word in their documents.

Input: Query, Cache Queries

Output: Result set generated for query

Procedure:

If Query available in cache

Result related to query: =

ForwardToTreeprocess (Query) Else

Result related to query: =

GeocodingtreeProcess (Query)

Geocoding process(Query):

Parameters

Qi—Input Spatial Query

Qj (j=1…n) ---Set of Queries contains same Location

Dist[j] (j=1…..n) -----Array for set of distances

5. PREFERENCE TREE

In this section hash-based index structures using distance and time based pruning and achieves high scalability and speedup. We present an exact and an approximate version of the algorithm and architecture [14].

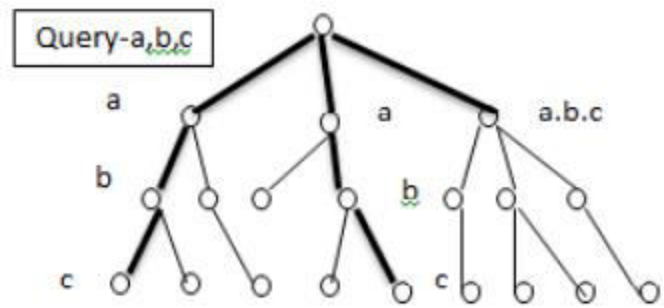

Figure 3 Preference Tree

Architecture

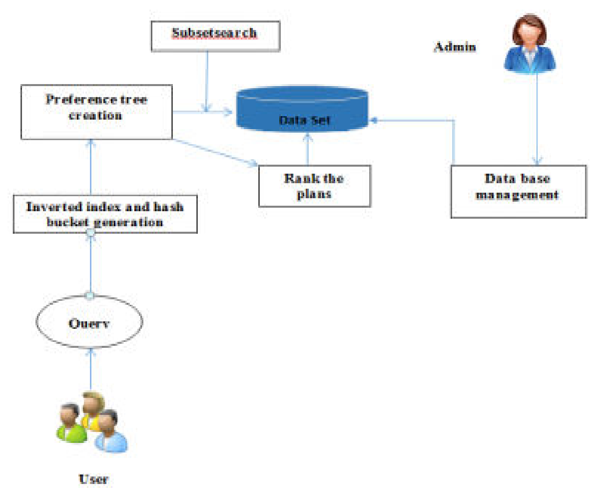

Figure 4 Architecture of Preference Tree

Preference tree creation: nodes are created in the basis of user preference and also checking the distance is less than threshold [15].

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Rank the plan: in the basis of sum of distance and sum of time which path getting shortest distance and time

Algorithm

In: location ,point of interest ,place name

Index creation Get all points containing the features

First Colum keyword

Second colum point id

Hash table generation

Create hash vector for each point which selected in previous step [16]

1.User to select the origin Either users current location or user can pick the point from Google map.

2. Find all preference point from index

3.Get all points with i th preference

4. If (distance< threshold)

5.Get all node of levels

6.Generate hash vector using preference

7.Checking the node with hash vector Until full fill the hash vector

8.Rank the plan

9. Stop

6. SIMULATION RESULTS

The experimental evaluation of practical efficiency of our solutions with proposed and existing methods which are based on synthetic and real dataFor the datasets, the vocabulary has 200 words, each word appears in 50k data points. The deficiency of IR2 -tree is mainly caused by the need to verify a vast number of false hits. To illustrate this, the figure below plots the average false hit number per query. We see an exponential escalation of the number on Uniform and Census, which explains the drastic explosion of the query cost on those datasets. Interesting is that the number of false hits fluctuates a little on Skew, which explains the fluctuation in the cost of IR2 -tree. The space consumption of IR2 –tree, SI-Index on the datasets of uniform, skew, Census are explained in the figure below. IR2 Tree has much more space efficiency than any other technique but doesn't compensate with the expensive query time. The SI-Index accompanied by the proposed query algorithms, has presented itself as an excellent tradeoff between space and query efficiency. Compared to IR2 Tree, its superiority is very high as the factors of order magnitude is typically high than its query time.

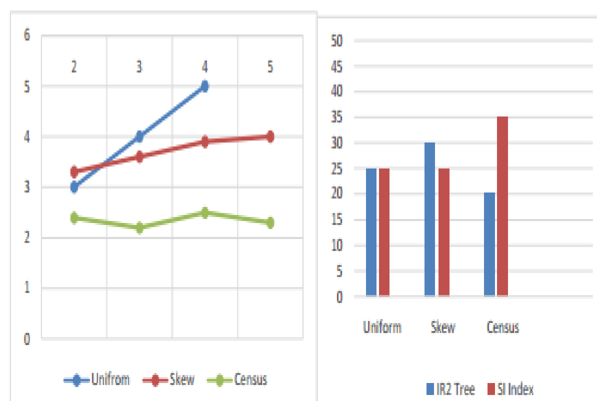

Fig.5.No. of False hits of IR2-Tree
Fig. 6. Comparison of Space Efficiency

7. CONCLUSION

We conclude that nearest neighbor search in multimedia databases needs more support from similarity search in query processing. 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 neighbor search in time that is at the order of dozens of milliseconds. 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. We plan to explore other scoring schemes for ranking the result sets Then, each group of points can be scored based on distance between points and weights of keywords. Furthermore the criteria of a result containing all the keywords can be relaxed to generate results having only a subset of the query keywords using satellite imaging to find the nearest locations

8. REFERENCES

[1] W. Li and C. X. Chen, "Efficient data modeling and querying system for multi-dimensional spatial data," in GIS, 2008, pp. 58:1–58:4.

[2] 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.

[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] I. De Felipe, V. Hristidis, and N. Rishe, "Keyword search on spatial databases," in ICDE, 2008, pp. 656–665

[5] Ninh Pham, IT University of Copenhagen Denmark ndap@itu.d, Hybrid LSH: Faster Near Neighbors Reporting in High-dimensional Space, in ICDE, 2010.

[6] 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.

[7] Vishwakarma Singh, Bo Zong, Ambuj K. Singh, Nearest Keyword Set Search in Multi-dimensional Datasets, IEEE Transactions on Knowledge and Data Engineering in 2016.

[8] D. Zhang, Y. M. Chee, A. Mondal, A. K. H. Tung, and M. Kitsuregawa, "Keyword search in spatial databases: Towards searching by document," in ICDE, 2009, pp. 688–699.

[9]. Rosslin John Robles, "Fast Nearest-Neighbor Search Algorithms Based on High-Multidimensional Data," Asia-pacific Journal of Multimedia Services Convergence with Art, Humanities and Sociology Vol.3, pp. 17-24, 20123

[10] I. D. Felipe, V. Hristidis, and N. Rishe, "Keyword Search on Spatial Databases", Proc. of International Conference on Data Engineering (ICDE), pp. 656– 665.s., 2008

[11] ShilpaThakare" Novel Method for NKS Search in Multidimensional Dataset Using Advance Promish& Ranking Function" International Journal for Research in Engineering Application & Management (IJREAM) ISSN : 2494-9150 Vol-02, Issue 08, Nov 2016

[12] K.Pujitha " A SURVEY ON SEMANTIC BASED SEARCH ENGINE FOR REAL IMAGES AND WEB URL'S USING HYPERGRAPH DISTANCE MEASURE ALGORITHM" ijptonline ISSN: 0975-766X.

[13] 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.

[14] Vishwakarma Singh, Santa Barbar, Ambuj,K. Singh, SIMP: Accurate and Efficient Near Neighbor Search in High Dimensional Spaces, EDBT 2012, March 26–30, 2012.

[15] Yufei Tao, Cheng Sheng, Fast Nearest Neighbor Search with Keywords, IEEE Transactions on Knowledge and Data Engineering , in 2012.

[16] Dongxiang Zhang, Beng Chin Ooi, Anthony K. H. Tung, Locating Mapped Resources in Web 2.0 , SIGMOD, 2013.

Name of the Student: **Mundlamuri Renuka**

cse

Renuka.mudlamuri535@gmail.com

Name of the Guide: B.Jyothi

Designation: Asst Professor

Mail Id:Jyothi.nalajala@gmail.com

College Name and Address: Krishnaveni Engineering College for Women ,Kesanupalli, narasaraopet