IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

Best Keyword Search on Spatial Database with Privacy of User

Ms. R. Padmapriya, MCA., M.Phil.[1], Ms. M. Gokila Devi, MCA., M.Phil., B.Ed.,[2]

Assistant Professor, Department of BCA, Kammavar Sangam College of Arts & Science, Theni[1]

Assistant Professor, Department of BCA, Kammavar Sangam College of Arts & Science, Theni[2]

Abstract: Databases nowadays contain large volumes of data, and they are accessed by numerous users on a daily basis. The large volume of data poses challenges to both users accessing the databases and the companies or organizations managing them. Enterprises on the other hand, need to make their content visible and accessible to the users and identify which objects in their database (e.g. products) have a significant impact on the user basis and use this information for promoting their products. The original target of increasing the visibility of the available products is thus hindered by the abundance of products contained in the database. It is therefore necessary to develop data exploration techniques that will enable users to explore large databases and provide them with a wide, yet coherent overview of objects that fit their preferences. In this paper, we propose exploratory algorithms that return to the user a small number of results, which at the same time provide a wide overview of the available content. We also propose analysis techniques are KNN algorithm used for best keyword search over spatial database(location based service), AES algorithm is implemented for secure accessing of users privacy data's i.e..location information and Hilbert Curve algorithm used for cache maintenance, caching scheme that aims to reduce query-response times and network traffic between the clients and the server by attempting to answer queries locally from the cached tuples using associated predicate descriptions. Identifying frequent search objects that are attractive to the users. This algorithm more efficient algorithm that achieves results of comparable quality, but with significantly lower processing cost.

Keywords: Keyword Search, KNN, AES, HilbertCurve.

1. INTRODUCTION

Most companies today invest significant resources on making their content visible on the Web and enabling users to browse the offered products and services. Often, companies provide a plethora of different alternatives, which overwhelm the user and make it extremely difficult for her to find the products she is interested in.

When users are searching in a database, they are usually unaware of the exact database content. Quite commonly, they do not have a concrete idea of the objects' properties they are searching for but only certain preferences about them. Consequently, they need to explore the database contents to find the objects that best fit their preferences. For instance, if someone wishes to buy a laptop, one may have a general idea about the desired characteristics, but an exact description of the laptop is difficult to be strictly determined. Traditional database queries are hard constraint queries, which return either exact matches or nothing. In addition, hard constraint queries are in general quite complex, and in order to produce useful results, they require the user to be aware of the database content. They also often require the knowledge of a specific query language and the structure of the queried database. Moreover, hard constraints are quite likely to produce very small or extremely large result sets that provide little insight of the available data. As a result, users are led to pose repeatedly new queries until they retrieve a satisfying result set. Therefore, they are inappropriate for exploratory search as they pose significant difficulties to users searching the database.

Users experience frustration when they are not able to easily find the information they need. In an attempt to make database content easily accessible, several approaches have been proposed, which allow users to express their needs by posing preference queries using either sets of keywords or by indicating their interest on the objects' attributes they are searching for. The query result is typically a list of objects, usually ranked according to a function that measures the relevance or the performance of each object with respect to the query.

A key aspect that preference queries fail to capture in its entirety is the fact that users performing exploratory search are generally unfamiliar with the domain of the data they are searching, and they are possibly unclear about their wishes. A flat list of results provides little insight to the user about the available information. In addition, the relaxation of constraints induced by preference queries introduces ambiguity to the search, as each keyword query could be associated with a large

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

number of database queries. As a result queries can produce a large number of redundant results, which the user has to filter out. In this paper, we propose exploratory algorithms that return to the user a small number of results, which at the same time provide a wide overview of the available content. In addition, we present algorithms that identify items that are appealing to users and can be exploited for offering users an insight of the available items and motivating them to explore the database. We also propose analysis techniques are KNN algorithm used for best keyword search over spatial database(location based service), AES algorithm is implemented for secure accessing of users privacy data's i.e..location information and Hilbert Curve algorithm used for cache maintenance, caching scheme that aims to reduce query-response times and network traffic between the clients and the server by attempting to answer queries locally from the cached tuples using associated predicate descriptions. Identifying frequent search objects that are attractive to the users.

2. RELATED WORKS

The existing works focus on retrieving individual objects by specifying a query consisting of a query location and a set of query keywords (or known as document in some context). Each retrieved object is associated with keywords relevant to the query keywords and is close to the query location. The similarity between documents are applied to measure the relevance between two sets of keywords.

Efficient Processing of Direction Joins Using R-trees [1] presents an efficient method for processing direction joins using R-trees. The quad-tuples model is defined to represent direction relations between the minimum bounding rectangles of spatial objects. An algorithm of processing the filter step **of** joins using R-trees is given and the refinement step processing is further decomposed into three different operations. Answering Why-Not Spatial Keyword Top-k Queries via Keyword Adaption [2] A spatial keyword top-k query takes a user location and a set of keywords as arguments and retrieves the k objects that are ranked the highest according to a scoring function that considers both spatial distance and textual similarity. Efficient Collective Spatial Keyword Query Processing on Road Networks [3] We study the problem of collective spatial keyword queries on road networks (i.e., CSKQ on road networks), which retrieves a set of POIs (Point of Interests) that collectively cover the queried keywords and have the lowest cost, measured by their shortest path distances to a specified query position, and the inter-POI distances between POIs in the set. Efficient Top-k Spatial Locality Search for Co-located Spatial Web Objects [4] Locality Search, a query that returns top-k sets of spatial web objects and integrates spatial distance and textual relevance in one ranking function. Keyword Search on Spatial Databases [5] we introduce an indexing structure called IR2-Tree (Information Retrieval R-Tree) which combines an R-Tree with superimposed text signatures. We present algorithms that construct and maintain an IR2-Tree, and use it to answer top-k spatial keyword queries. Challenges in the Design and Implementation of Wireless Sensor Networks: A Holistic Approach- Development and Planning Tools, Middleware, Power Efficiency, Interoperability [6] WSN challenges by developing an integrated platform for smart environments with built-in user friendliness, practicality and efficiency. This platform will enable the user to evaluate his design by identifying critical features and application requirements. Location Aware Keyword Query Suggestion Based on Document Proximity [7] weighted keyword-document graph, which captures both the semantic relevance between keyword queries and the spatial distance between the resulting documents and the user location. The graph is browsed in a random-walk-with-restart fashion, to select the keyword queries with the highest scores as suggestions.

3. PROPOSED WORK

It is motivated by the observation of increasing availability and importance of keyword rating in decision making. We presented algorithms for the identification of objects that are constantly attractive for a large number of users over a specified period of time. The present algorithm is used for identify the frequently searched items in a spatial database through this we improve the best keyword search for the web users.

- Searching local best solution for each object in a certain query keyword.
- The number of candidate keyword covers generated is significantly reduced.
- Mining user's availability based on their interest.
- Compared to the baseline algorithm, the number of candidate keyword covers generated in this algorithm is significantly reduced.
- The in-depth analysis reveals that the number of candidate keyword covers further processed and each keyword candidate cover processing generates much less new candidate keyword covers than that in the baseline algorithm.
- Which is used for identify the frequently searched items in a spatial database through this we improve the best keyword search for the web users.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

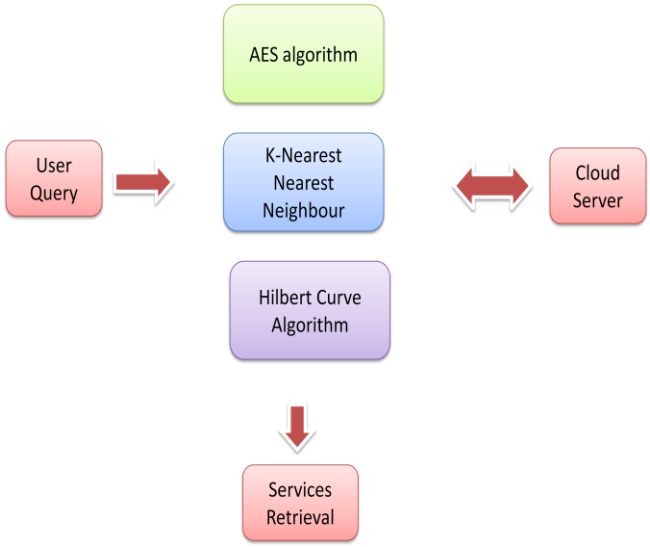


Fig 1: Architecture Diagram

3.1 QUERY PREPROCESSING

Quality of data is first and foremost step before running analysis. It is to find much irrelevant and redundant information present or noisy unreliable data during processing. Stop words are most common words found in any natural language, which carries very little or no significant semantic context in a sentence. It just carry syntactic importance which aid in formation of sentence. As a preprocessing operation it must be removed to ease further task and speedup core task in text processing.

3.2 USER LOCATION PRIVACY

The server will super imposed the user location information with the server location grid and find the location server's grid cell information. The server will also send all the information encrypted using the key sent by the client. The communication between client and server will be secure. The client will never send the GPS coordinate but will send an information of its location. AES Algorithm has also been proposed to confuse and distort the location data, which include path and position confusion.

3.3 LOCATION BASED SERVICES RETRIEVAL

A location-based instant search combines spatial search with keyword search using the AND semantics. That is, we want to retrieve records that satisfy both spatial and keyword conditions. The Location Based Service (LBS) applications can help user to find hospitals, school, gas filling station or any other facility of interest indicated by user within certain range. Just like a GPS device its location will also be updated as soon as user changes his/her position. We implement KNN algorithm for improve the efficient search over the spatial database.

3.4 CACHE MAINTAINANCE

Caching scheme that aims to reduce query-response times and network traffic between the clients and the server by attempting to answer queries locally from the cached tuples using associated predicate descriptions. The database is assumed to be resident at the central server, with users originating transactions from client sites. Each client executes transactions sequentially, with at most one transaction active at any time. We presents an improvement to Cauchy maintenance based on Hilbert Curve Algorithm.

4. METHODOLOGY

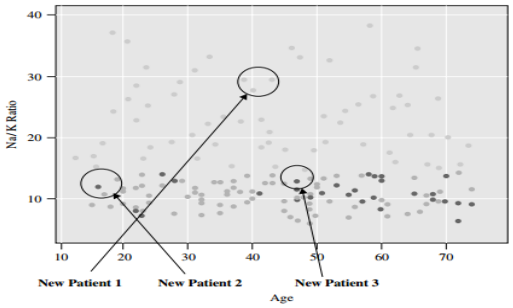
4.1 KNN ALGORITHM

K-Nearest Neighbor (KNN) search returns a specified number of data objects, sorted by their distances from a given query point. KNN has been addressed mostly in the context of spatial databases, though its applications can also be found in pattern recognition, image processing, CAD, and multimedia indexing

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017



We have seen above how, for a new record, the k-nearest neighbor algorithm assigns the classification of the most similar record or records. A distance metric or distance function is a real-valued function d, such that for any coordinates x, y, and z: 1. $d(x,y) \geq 0$, and $d(x,y) = 0$ if and only if $x = y$
2. $d(x,y) = d(y,x)$
3. $d(x,z) \leq d(x,y) + d(y,z)$

4.1 AES Algorithm

AES is a block cipher with a block length of 128 bits. It allows three different key lengths: 128, 192, or 256 bits. We propose AES with 128 bit key length. The encryption process consists of 10 rounds of processing for 128-bit keys. Except for the last round in each case, all other rounds are identical. 16 byte encryption key, in the form of 4-byte words is expanded into a key schedule consisting of 44 4-byte words. The 4 x 4 matrix of bytes made from 128-bit input block is referred to as the state array. Before any round-based processing for encryption can begin, input state is XORed with the first four words of the schedule. For encryption, each round consists of the following four steps:

- SubBytes – a non-linear substitution step where each byte is replaced with another according to a lookup table (S-box).
- ShiftRows – a transposition step where each row of the state is shifted cyclically a certain number of times
- MixColumns – a mixing operation which operates on the columns of the state, combining the four bytes in each column.
- AddRoundKey – each byte of the state is combined with the round key; each round key is derived from the cipher key using a key schedule.

4.2 Hilbert Curve Algorithm

Our cloaking algorithm consists of two steps; cloaking region generation by the Hilbert value and expansion cell selection by considering the locality. The algorithm iterates above two steps until satisfying k-anonymity.

Definition 1. Grid cells ordered by Hilbert curve

The set of cells ordered by Hilbert curve is defined as H = {C00, C01, C02, ... Cij, ..., C(N-1)(N 1)}, where i and j are the (x,y)- coordinates of a grid and N is the number of grid cells in one dimension. Next, it stores the information of adjacent cells being not connected by Hilbert curve. Because the number of objects is required to store along with the adjacent cells' information, it searches neighbor cells for each cell. If the difference of Hilbert value is greater than '1', it inserts the neighbor cell into our data structure. The definition of the adjacent cell is below.

Definition 2. Adjacent cell without sequential Hilbert curve's order

The set of adjacent cells being not connected by Hilbert curve is defined as

$$AC = \{ ac_{ij} \,||\, ac_{ij} - c_{xy} \,| > 1, {}^{¥}ac_{ij}, {}^{¥}c_{xy} \in H\}$$

Where $0 \leq i,j \leq N$ and $i = x \pm 1$, $j = y \pm 1$

By using our data structure, it can reduce a cloaking area generated and can decrease processing time for finding k-1 users.

5. EXPERIMENTAL RESULT AND DISCUSSION

In this round of evaluation, we first conduct an experiment varying the number of objects to evaluate the scalability of the indices for each type of query. In addition, to evaluate the effect of the text size of each object, we conduct an experiment varying the average number of words per object for the best search. Keywords is large or a space limitation has to be strictly satisfied. Grid based indices are not attractive for the BKS compared with the other indices.
We do not find dramatic difference in relative performance among the indexing techniques.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

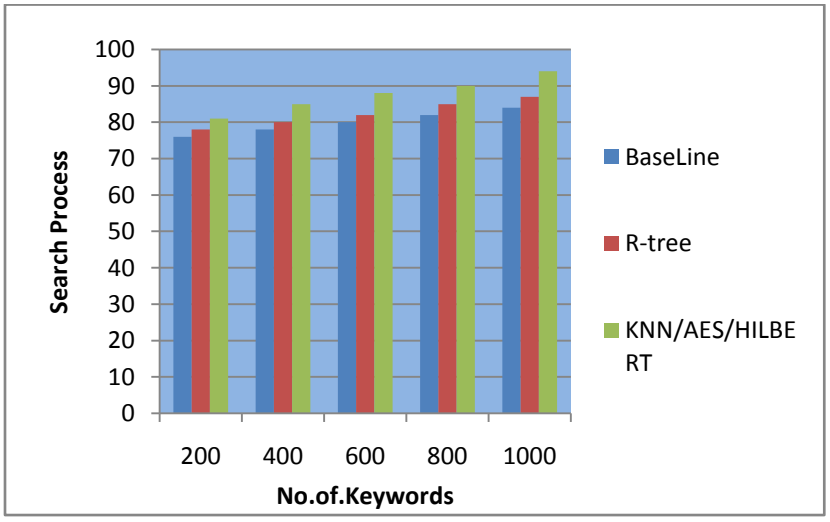


Fig 2: Experimental Chart

The query processing of the indices scales linearly with the number of objects and text length per object. Text-first indices are more sensitive to text length than the other types of indices. The experimentally shown to be a factor that makes a great difference on query performance. In this paper, the system explored the performance of the popular hash operate function of Hilbert curve's Algorithm. Then it checks the algorithm's quality in two aspects: procedure and area complexity. Then it checks the protection aspects of Hilbert curve's Algorithm.

6. CONCLUSION

The major problem with frequent set mining methods presented previews is the explosion of the number of results, it is difficult to find the most interesting frequent item sets. In this paper we propose exploratory algorithms that return to the user a small number of results, which at the same time provide a wide overview of the available content. In addition, we present algorithms that identify items that are appealing to users and can be exploited for offering users an insight of the available items and motivating them to explore the database. We also propose analysis techniques using KNN algorithm for identifying frequent search objects that are attractive to the users.

REFERENCES

[1] Rakesh Agrawal and Ramakrishnan Srikant. "Fast algorithms for mining association rules in large databases". In: VLDB. 1994, pp. 487–499.
[2] T. Brinkhoff, H. Kriegel, and B. Seeger. "Efficient processing of spatial joins using R-trees". In: SIGMOD (1993), pp. 237–246.
[3] Xin Cao, Gao Cong, and Christian S. Jensen. "Retrieving top-k prestige-based relevant spatial web objects". In: Proc. VLDB Endow. 3.1-2 (2010), pp. 373–384.
[4] Xin Cao et al. "Collective spatial keyword querying".In: ACM SIGMOD. 2011.
[5] G. Cong, C. Jensen, and D. Wu. "Efficient retrievalof the top-k most relevant spatial web objects". In: Proc. VLDB Endow. 2.1 (2009), pp. 337–348.
[6] Ian De Felipe, Vagelis Hristidis, and Naphtali Rishe. "Keyword Search on Spatial Databases". In: ICDE. 2008, pp. 656–665.
[7] R. Fagin, A. Lotem, and M. Naor. "Optimal Aggregation Algorithms for Middleware". In: Journal of Computer and System Sciences 66 (2003), pp. 614–656.
[8] Ramaswamy Hariharan et al. "Processing Spatial-Keyword (SK) Queries in Geographic Information Retrieval (GIR) Systems". In: Proceedings of the 19th International Conference on Scientific and Statistical Database Management. 2007, pp. 16–23.
[9] G. R. Hjaltason and H. Samet. "Distance browsing in spatial databases". In: TODS 2 (1999), pp. 256–318.
[10] Z. Li et al. "IR-tree: An efficient index for geographic document search". In: TKDE 99.4 (2010), pp. 585–599.
[11] N. Mamoulis and D. Papadias. "Multiway spatial joins". In: TODS 26.4 (2001), pp. 424–475.
[12] D. Papadias, N. Mamoulis, and B. Delis. "Algorithms for querying by spatial structure". In: VLDB (1998), p. 546.
[13] D. Papadias, N. Mamoulis, and Y. Theodoridis. "Processing and optimization of multiway spatial joins using R-trees". In: PODS (1999), pp. 44–55.
[14] J. M. Ponte and W. B. Croft. "A language modeling approach to information retrieval". In: SIGIR (1998), pp. 275–281.
[15] Jo˜ao B. Rocha-Junior et al. "Efficient processing of top-k spatial keyword queries". In: Proceedings of the 12th international conference on Advances in spatial and temporal databases. 2011, pp. 205–222.
[16] S. B. Roy and K. Chakrabarti. "Location-Aware Type Ahead Search on Spatial Databases: Semantics and Efficiency". In: SIGMOD (2011).