ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

A Study for Spatial Approximate String Queries in both the Euclidean Space and Road Networks

R.Karthikeyan[1], V.Khanna[2]

[1]Assitant Professor, Dept. of CSE, Bharath University, Chennai, Tamil Nadu, India

[2]Dean, Dept. of IT, Bharath University, Chennai, Tamil Nadu, India

ABSTRACT: This work deals with the approximate string search in large spatial databases. To investigate range queries augmented with a string similarity search predicate in both Euclidean space and road networks. This query is spatial approximate string (SAS) query. A query is used to access only certain information at a time. a database query is a function of database software. In Euclidean space, to propose an approximate solution, the MHR-tree, which embeds min-wise signatures into an R-tree. The min-wise signature for an index node u keeps a concise representation of the union of q-grams from strings under the sub-tree of u. to analyze the pruning functionality of such signatures based on the query string and q-grams from the sub trees of ides nodes.

KEYWORDS: spatial databases, approximate string search, range query, road network.

I. INTRODUCTION

Keyword search over a large amount of data is an and textual information. To present the solution important operation in a wide range of domains. Approximate string search is necessary when users have a fuzzy search condition, or a spelling error. In the context of spatial databases, approximate string search could be combined with any type of spatial queries. This work complements the standard shortest path search with multiple incorporated. GIS data consist of both spatial support to fuzzy keyword search on spatial data. May Tobsites based on geographical information nowadays support Keyword search on their data such as busieness listings and photo. To also experimented with the Flicker Maps photo search engine and saw similar limitations. An interesting observations that during the development of our work, the results of these systems kept changing .

The rationale is that higher levels can have many keywords and it is computationally expensive to do pruning based on the condition Qt by finding similar keywords. This Date is cleaning Information from multiple data sources of ten have numerous in consistencies. the Errors can also be introduced due to irregular cities in the data collection process, from human mistakes, and many other causes. For RSAS queries, the baseline spatial solution is base on the Dijkstra's algorithm. To denote this approach as the dijkstra solution, Its performance degrades quickly when the query motivates use to find a novel method to avoid the unnecessary and /or the data on the networks increases. ESAS and RSAS queries is to build a string matching index and evaluate only the string predicate, completely ignoring the spatial component of the query. The string Solution First, the string solution suffers the same scakabutky search as the spatial solution. To want to enable the neighbor queries. The RSASSol method partitions method partitions the road network, adaptively searches relevant sub graphs, and prunes candidate points using both the string matching index and the spatial reference nodes. Multipoints ALT algorithm (MPALT) is applied, together with the exact edit distances, to verify the final set of candidates.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

II. PROBLEM FORMULATION

2.1 The Eases Queries

Computing edit distance exactly is a costly operation. Several techniques have been proposed for indentifying candidate strings with in a small edit distance from a query string fast. To handle duplicates in the q-gram to indicate the number of times it appears in the string. When all point in p are assigned into MBRs. The resulting leaf Ned MBRs are then further grouped together recursively till there is only one MBR left. The Min-wise signature with linear Hashing R-tree (MHR-tree).The rest of this section explains its construction and query algorithms.

2.2 Query algorithms for the MHR-tree

The query algorithms for the MHR-tree generally follow on the same principles as the corresponding algorithms for on the spatial query component. To can easily compute ($g_{o)}$ and s(g_s) from the query string once, using the same hash functions that tore used for constructing the MHR-tree. A State of the art technique based on q-grams and min-wise signatures is Viol .It builds inverted lists with q-gram in input strings. The general principle behind accurate spatial selectivity estimation is to partition the spatial date into a collection of buckets so that data within each bucked is as close as possible to a uniform distribution. Every bucket is defined by the MBR In the Pre-processing step to find the number of neighborhoods for each R-tree node. Next, to keep adding nodes, one at a time until the overall value for the neighborhood and uniformity quality for b_1 and the remaining nodes connote be reduced.

The key of common Index can be made form the Index word given by the Data owner and File. The secure index and a search scheme to enable fast similarity each in the context of data. In such a context, it is very critical not to sacrifice the confidentiality of the sensitive data while providing functionality. To provided a rigorous security definition and proved the security of the proposed scheme under the provided definition to ensure the confidentiality. Edit Distance poruning is Computing edit distance exactly is a costly operation. Several techniques have been proposed for indentifying candidate strings within a small edit distance form a query string fast. produced by sliding a window of length q over the characters.

2.3 The Rsas Queries

To adopt a disk-based storage model to our setting that groups network nodes based on their connectivity and distance, The distance batten two nodes, two (network) path connecting two objects of Soncern. To also store other information associated with a point (i.e, strings) after tee offset distance. To store the points on the same edge in appoints groups.

For RSAS queries, the baseline spatial solution is based on the Dijkstra's algorithm. Given a query point q, the query range radius r, and a string predicate, to expand from q on the road network using the Dijkstra algorithm until to reach the points distance r away from q and verify the string predicate either in a post-processing step or on the intermediate results of the expansion. To denote this approach as the Dijkstra solution. Its performance degrades quickly when the query range enlarges and/or the data on the network increases. This motivates us to find a novel method to avoid the unnecessary road network expansions, by combining the pruning's from both the spatial and the string predicates simultaneously. To demonstrate the efficiency and effectiveness of our proposed methods for SAS queries using a comprehensive experimental evaluation.For ESAS queries, our experimental evaluation covers both synthetic and real data sets of up to 10 million points and 6 dimensions. For RSAS queries, our evaluation is based on two large, real road network datasets[1].

2.4 The RSASSol algorithm

keyword keyword query takes a user location and
Recall that to apply the MPALT algorithm using q as
RSAS query framework consists of five steps and the comments in Algorithm. Given a query, To first find all sub graph that intersect with the query range. Next, to use the Filter Trees if tees sub graphs it retune the third step. to prune away some of these candidate points by calculating the lower and upper bounds of their distances to the query point, using VR. The fourth step is to further prune away some candidate points using the exact edit distance between the query string and strings of remaining candidates. After this step, the string predicate has been fully explored. In the final step, for the remaining candidate points, to check their inexact distances to the query point and return those with distances within. The MPALT algorithm minimizes the access to the network by avoding the.[3] user-supplied

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

keywords. Is a arguments and rectums objects that are spatially and textually relevant to these arguments. Spatial networks are often used for the modeling of transportaion networks such as real road networks. The optimized planar algorithm for is always the best strategy and given the same number of reference nodes. Second, the number of reference nodes is also critical.

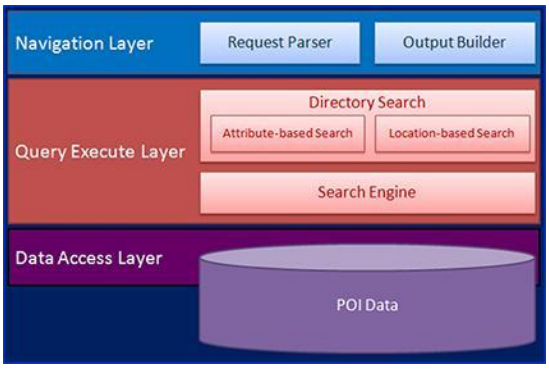
To present to solution to support fuzzy keyword search on spatial data To combine spatial index structure with inverted indexes on grams to efficient leans To fuzzy queries on maps.[4] Data cleaning Information from multiple data to irregular-ities in the data- collection process, from human mistakes, and many other causes. Approximate string search algorithms with a high real time performance. For instance, consider a spell checker such as those used by Gmail, Hot-mail, or Yahoo! Mail, which need to be invoked numerous times per second, I order to support the millions of concur-rent user using these services.[5]

III DESIGN IMPLEMENTATION

3.1 system design

Systems design is simply the design of systems. It implies a systematic and rigorous approach to design—an approach demanded by the scale and complexity of many systems problems. Systems design first appeared shortly before World War II as engineers grappled with complex communications

and control problems. They formalized their work in the new disciplines of information theory, operations research, and cybernetics. A data access layer (DAL) in computer software, is a layer of a computer is a program is which provides simplified access to the some kind, such as an entity-relational database.[6]



3.1.1 SYSTEM DESIGN ARCHITECTURE

This Applications using a data access layer can be either database server dependent or independent. If the data access layer becomes able to use whatever databases. DAL is a can talk to. the either circumstance, having a data access layer provides a centralized location for all calls into the database, and thus makes it easier to port the application to other database systems is to the mostly prevalently use in Microsoft ASP.NET environments.[7]

Navigation layer is used to provide basic navigation control over the *Qtwidget*. When this layer is attached and active, you may use the arrows to navigate the equivalent in world coordinates of 10 pixels in the desired direction. You can also use the Page Up, Page Down keys to scroll the equivalent in world coordinates of half screen height on the Y axes.[8] A parser is a piece of software that evaluates the syntax of a script when it is executed on a server. For scripting languages is use on the tob and the parser is works like to compiler might work in the other type of real application development environments. Parsers are commonly used in script development because they can evaluate code when the script is executed and do not require that the code be compiled first. The State of the art solution for the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

problem can be divided into two categories to spatial coherence based methods and vertex-importance-based approaches. The two categories of techniques, hot over have not been to compared systematically.

To under the same experimental and framework as they to redeveloped from two Independent lines of research that do no tree far to each other. [9] To focus on the fundamental application is a locating geographical resources and propose on the efficient tag centric query processing strategy.

The Mapped resources is essentially to a spatial tag matching problem. Based on this data model, the problem of locating mapped resources is essentially To aim to find a spatial location that best matches the associated tags. Such a query also has great potential for location service providers.[10] To this motivates us to find a novel method to avoid the unnecessary road network expansions, by combining the pruning's from both the spatial and the string predicates simultaneously. The Spatial databases is a where keyword search becomes a fundamental building block for an increasing number of real-world applications, and proposed the IrTree. A main limitation of the IR-Tree is that it only supports exact keyword search. Its performance degrades quickly when the query range enlarges and or the data on the network increases. The secure index and a search to enable fast similarity search in the context of data. In such a context, it is very critic

To demonstrate the efficiency and effectiveness of our is proposed methods for SAS queries using to a comprehensive experimental evaluation. The query layer is a layer or stand-alone table that is defined by a SQL query[11]. Query layers allow both spatial and non spatial information stored in a DBMS to be easily integrated into GIS projects within Arc Map. Since query layers are using SQL to directly query database tables and views, spatial information used by a query layer is not required to be in a geo database.[12]

IV. CONCLUSION

A comprehensive study for spatial approximate string queries in both the Euclidean space and Road networks. To use the edit distance as the similarity measurement for the string predicate and focus on the range queries as the spatial predicate. To also address the problem of query selectivity estimation for queries in the Euclidean space Future work include examining spatial approximate sub-string queries, designing methods that are more update-friendly, and solving the selectivity estimation problem for RSAS queries.

REFERENCES

[1] S. Acharya, V. Poosala, and S. Ramaswamy. Selectivity estimation in spatial databases. In SIGMOD, pages 13–24, 1999.
[2] Kaliyamurthie K.P., Parameswari D., Udayakumar R., "QOS aware privacy preserving location monitoring in wireless sensor network", Indian Journal of Science and Technology, ISSN : 0974-6846, 6(S5) (2013) pp.4648-4652.
[3] Sharmila D., Muthusamy P., "Removal of heavy metal from industrial effluent using bio adsorbents (Camellia sinensis)", Journal of Chemical and Pharmaceutical Research, ISSN : 0975 – 7384, 5(2) (2013) pp.10-13.
[4] S. Alsubaiee, A. Behm, and C. Li. Supporting location-based approximate keyword queries. In GIS, pages 61–70, 2010.
[5] Udayakumar R., Khanaa V., Saravanan T., Saritha G., "Retinal image analysis using curvelet transform and multistructure elements morphology by reconstruction", Middle - East Journal of Scientific Research, ISSN : 1990-9233, 16(12) (2013) pp.1781-1785.
[6] A. Arasu, S. Chaudhuri, K. Ganjam, and R. Kaushik Incorporating string transformations in record matching. In SIGMOD, pages 1231–1234, 2008.
[7] A. Arasu, V. Ganti, and R. Kaushik. Efficient exact set-similarity joins.In VLDB, pages 918–929, 2006.
[8] Kalaiselvi V.S., Prabhu K., Ramesh M., Venkatesan V., "The association of serum osteocalcin with the bone mineral density in post menopausal women",Journal of Clinical and Diagnostic Research, ISSN : 0973 - 709X, 7(5) (2013) pp.814-816.
[9] N. Beckmann, H. P. Kriegel, R. Schneider, and B. Seeger. The R_-tree: an efficient and robust access method for points and rectangles. InSIGMOD, pages 322–331, 1990.
[10] Jayalakshmi T., Krishnamoorthy P., Kumar G.R., Sivamani P., "The microbiological quality of fruit containing soft drinks from Chennai", Journal of Chemical and Pharmaceutical Research, ISSN : 0975 – 7384, 3(6) (2011) pp. 626-630.
[11] A. Z. Broder, M. Charikar, A. M. Frieze, and M. Mitzenmacher. Minwise independent permutations (extended abstract). In STOC, pages327–336, 1998.
[12] X. Cao, G. Cong, and C. S. Jensen Retrieving top-k prestige based relevant spatial Tob objects.Proc. VLDB Endow., 3:373–384, 2010.
[13].Dr.A.Muthu Kumaravel, KNOWLEDGE BASED WEB SERVICE, International Journal of Innovative Research in Computer and Communication Engineering, ISSN(Online): 2320-9801,pp 5881-5888, Vol. 2, Issue 9, September 2014
[14].Dr.A.Muthu Kumaravel, Data Representation in web portals, International Journal of Innovative Research in Computerand Communication Engineering, ISSN(Online): 2320-9801,pp 5693-5699, Vol. 2, Issue 9, September 2014
[15].Dr.athir.Viswalingam, Mr.G.Ayyappan,A Victimization Optical Back Propagation Technique in Content Based Mostly Spam Filtering

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

,International Journal of Innovative Research in Computer
and Communication Engineering ,ISSN(Online): 2320-9801 , pp 7279-7283, Vol. 2, Issue 12, December 2014
[16].KannanSubramanian,FACE ECOGNITION USINGEIGENFACE AND SUPPORT ECTORMACHINE,International Journal of Innovative Research in Computerand Communication Engineering,ISSN(Online): 2320-9801,pp 4974-4980, Vol. 2, Issue 7, July 2014.
[17].Vinothlakshmi.S,To Provide Security & Integrity for StorageServices in Cloud Computing ,International Journal of Innovative Research in Computer and Communication Engineering ,ISSN(Online): 2320-9801 , pp 2381-2385 ,Volume 1, Issue 10, December 2013