IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p-ISSN: 2278-8735
PP 27-30
www.iosrjournals.org

Efficient Probing and Ranking Based On Concise Range Query

Remya.R.S[1],
[1](*Computer Science and Engineering Department, Muslim Association College of Engineering, Kerala, India)*
remyars2003@gmail.com

Abstract: *Spatial databases have analyzed an increasing number of applications recently, partially due to the fast press forward in the fields of mobile computing, embedded systems and the spread of the Internet. With the progress of wireless communication technology, it is quite common for people to view maps or get related services from the handheld devices, such as mobile phones and PDAs. Range queries, as one of the most commonly used tools, are often caused by the users to retrieve needful information from a spatial database. However, due to the limits of communication bandwidth and hardware power of handheld devices, displaying all the results of a range query on a handheld device is neither communication- efficient nor informative to the users. This is simply because that there are often too many results returned from a range query. In view of this problem, I present a narrative idea that a concise representation of a specified size for the range query results, while incurring minimal information loss, shall be computed and returned to the user. Some heuristic algorithms are used here. This method projected on real world data in particular range on spatial area. IR-Tree(Inverted Relevance checking tree) used for proficient document searching and ranking based on the concise range query.*
Keywords: *Spatial databases, range queries, IR-tree, inverse document frequency.*

I. INTRODUCTION

It is quite widespread these days that people want to decipher the driving or walking directions from their handheld devices (mobile phones or PDAs).However, facing the huge amount of spatial data collected by various devices, such as sensors and satellites is a tedious task. With limited bandwidth and computing power of handheld devices, how to distribute light but usable results to the clients is a very interesting, and of course demanding task. This exertion has the same inspiration as several recent works on finding good representatives for large query answers Moreover, such requirements are not specific to spatial databases. General query processing for large relational databases and OLAP data warehouses has posed similar challenges In fact, this thesis returns concise representations of the final query results in every possible stage of a long-running query evaluation. However, the focus on join queries in the relational database and the approximate representation is a random sample of the final query results. See the goal of this work is different and random sampling is not a good solution for our problem.

II. PROBLEM STATEMENT

R*-trees are a variant of R-trees used for indexing spatial information. R*-trees support point and spatial data at the same time with a slightly higher cost than other R-trees. R*Tree built by repeated insertion. There is little overlap in this tree, resulting in good query performance. Minimization of both coverage and overlap is crucial to the performance of R-trees. The performance of R-trees depends on the quality of the algorithm that clusters the data rectangles on a node. Hilbert R-trees use space-filling curves, and specifically the Hilbert curve, to impose a linear ordering on the data rectangles. Improved split heuristic produces pages that are more rectangular and thus better for many applications. Efficiently supports point and spatial data at the same time. . This indicates that the nodes will likely have small area and small perimeters. Small area values result in good concert for point queries; small area and small perimeter values lead to good performance for larger queries.

2..1 Algorithm Hilbert-Pack

To calculate Hilbert value algorithm Hilbert-Pack is used. The headings and subheadings, starting with(packs minimum bounded rectangles into an R-tree)

Step 1. Calculate the Hilbert value for each data rectangle

Step 2. Sort data rectangles on ascending Hilbert values

Step 3. Create leaf nodes (level l=0) */

While (there are more rectangles) generate a new R-tree node assign the next C rectangles to this node

Step 4. /* Create nodes at higher level (l + 1) */ While (there are > 1 nodes at level l)sort nodes at level l ≥ 0 on ascending creation time repeat Step 3

System Architecture

This model attempts to reveal user backdrop knowledge and understands how the document searching and ranking process going on. After the geographical searching of a spatial location from the loaded document, R tree created. From that concise range query specified using different range queries. Searching the document and calculate the spatial and textual relevancies.

2..2 Geographic Document Search and Ranking

Assume each document t in a given document set D is composed of a set of words W, and is associated with a location L. Given a query q that specifies a set of query keywords Wq and a query spatial scope Sq, the textual relevance and spatial relevance of a document d to q are formalized in Definitions 1 and 2, respectively. The document searching and ranking based on idf values.
Wq ∩ Sq ≠ Φ .
To quantify the relevance of d to q, a weighting function denoted by Φq(d1) is adopted. Thus, for a given q, Φq(d1)> Φq(d2) means document d1 is more textually relevant to q than document d2.

2..3 Indentations and Equations

There are various models (e.g., vector space model, probabilistic model, language model, etc.) to measure the relevance of documents to a given query. Among all those, idf, inverse document frequency is the most widely used. For simplicity, we consider the generic one hereafter. TF-IDF weighs a term in a document based on term frequency (tf) and (idf) . A term frequency tfw; d measures the number of times a word w appears in a document d, which indicates the importance of the word within the document. In IR-Tree, we use methods term frequency and inverse document frequency to search and rank query as text and spatial relevance. Term frequency means relevant words are retrieved from the database based on user query. The context of geographic document search, the idf of a word w, denoted by idf w, D,S, is defined corresponding to a candidate document set Ds instead. The documents in the D s have their locations fully covered by the query spatial scope S.
Geographic Search Engine:-

Currently, two types of approaches are used by existing geographic search engines, namely, Approach I that uses separated indexes for spatial information and textual information, and Approach II that uses a combined index' However, they both are not efficient. Approach I logically extends conventional textual search engines with spatial filtering capability of Quad-tree, R-tree, respectively. As an example, in [5], the most recent work of Approach I, an inverted file is created to index words of documents and a grid index is created to index locations of documents. Based on two indexes, a search generally follows a three step process.
Step1: Retrieving textually relevant documents with respect to query keywords via a conventional textual index.
Step 2: Filtering out the documents obtained from Step 1 that are not covered by the query spatial scope.
Step 3: Ranking the documents from Step 2 based onthe joint textual and spatial relevance in order to return the ranked results to the user.
Here, I reviewed existing works in textual index, spatial index, and geographic document search engines.

III. SYSTEM METHODOLOGY

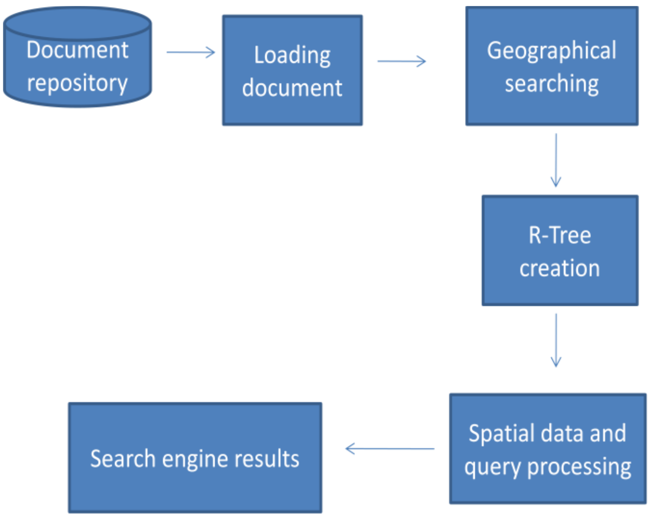


Fig 1: Architecture of efficient probing and ranking based on concise range query

Table 1: Landmark Data

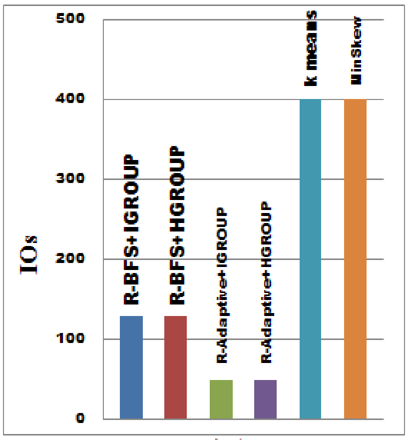
Field Name	Data Type	Size	Description
Landmark	Varchar	20	Spatial information
X coordinate	Double	5	Latitude
Y coordinate	Double	5	Longitude
Kilometer	Varchar	15	Distance from the current location.

IV. EXPERIMENTAL RESULTS

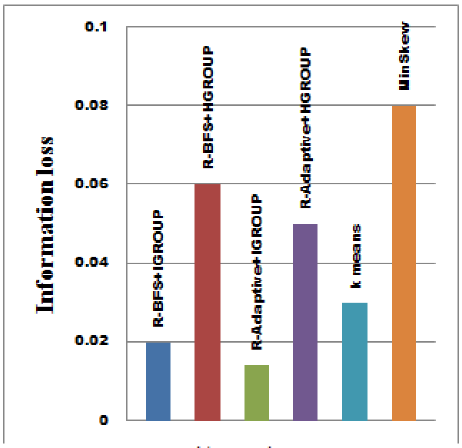
Location Scope Value	Spatial Relevance		Geographic Document Search	
	Text Value	Spatial Relevance Value	Rank	Idf Value
1.4303	S3	1.10	R1	1.99
1.6364	S4	1.40	R2	1.82
1.8373	S5	1.63	R3	1.63
1.9928	S6	1.82	R4	1.40
2.1364	S7	1.99	R5	1.10

Table 2: Ranking of geographic location based on Idf values

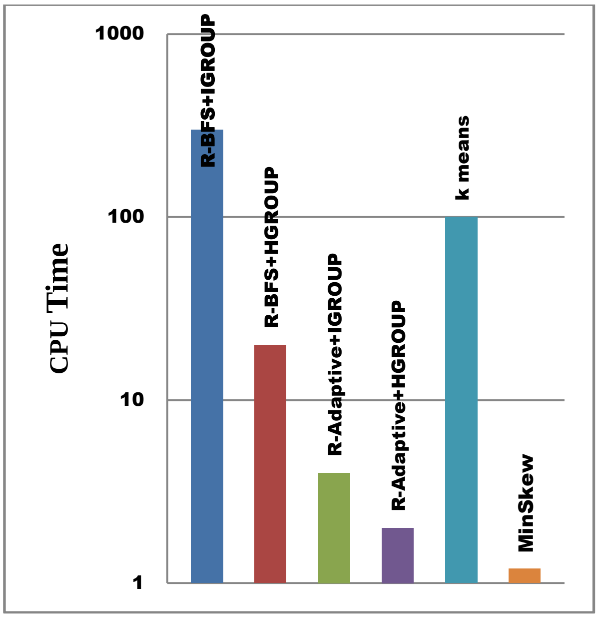
The output of special relevance checking illustrated in the graphical representation(table2)



a) I/O cost



b)CPU time



(c) Information loss

Fig. 2. Experimental results with varying query range on the data set.

V. CONCLUSION

In this paper, I focused on the efficiency issue of geographic document search and proposed an efficient indexing structure, namely, IR-tree. The contributions of this paper are summarized as follows:

- IR-tree which indexes both the textual and spatial contents of documents to support document retrievals based on their combined textual and spatial relevance's, which, in turn, can be adjusted with different relative weights.
- It propose a rank-based search algorithm based on IR-tree to effectively combine the search process and ranking process to minimize I/O costs for high search efficiency.
- A cost analysis for IR-tree and conduct a comprehensive set of experiments over a wide range of parameter settings to examine the efficiency of idf calculation.

From an extensive experimentation, IR-tree is demonstrated to outperform the state-of-the-art approaches. At present, I am prototyping a geographic search engine with IR-tree as the score and building a test bed based on idf calculation for future research. I also plan to further enhance the IR-tree index based on various access patterns. .

REFERENCES

Journal Papers :

[1] Ken C.K. Lee, Student Member, IEEE, Wang-Chien Lee, Member, IEEE Computer Society, and Hong Va Leong, Member, IEEE Computer Society" Nearest Surrounder Queries"Ieee Transactions On Knowledge And Data Engineering, Vol. 22, No. 10, October 2010

[2] Man Lung Yiu, Nikos Mamoulis, Xiangyuan Dai, Yufei Tao, and Michail VaitisIeee, "Efficient Evaluation of Probabilistic

[3] Guido Proietti and Christos Faloutsos, Member, Ieee "Analysis of Range Queries and Self-Spatial Join Queries on Real Region Datasets Stored Using an R-Tree"Transactions On Knowledge And Data Engineering, Vol. 12, No. 5, September/October 2000

[4] Xiang Lian, Student Member, IEEE, and Lei Chen, Member, IEEE " Ranked Query Processing in Uncertain Databases" Ieee Transactions On Knowledge And Data Engineering, Vol 22, No. 3, March 2010

[5] BongkiMoon, H.V.Jagadish, ChristosFaloutsos, Member, IEEE and JoelH Saltz,Member, "Analysis of the Clustering Properties of the Hilbert Space-Filling Curve" Ieee Transactions On Knowledge And Data Engineering, Vol. 13, No. 1, January/February 2001)

[6] " Efficient Threshold Monitoring for Distributed Probabilistic Data" Mingwang Tang, Feifei Li, Jeff M. Phillips, Jeffrey Jestes School of Computing, University of Utah, Salt Lake City, USA {tang, lifeifei, jeffp,}

[7] Ranking under temporal constraints-published in 2012(Authors: Lidan Wang University of Maryland, College Park, MD, USA)