International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 17
November 2016

A NOVEL KEYWORD COVER SEARCH USING APRIORI ALGORITHM

SIRMARIGARI SHIVANANDAM[1], DR M SREENIVASA RAO[2]& DR. G VENKATA RAMI REDDY[3]

[1]M-Tech Dept of CSE School of Information Technology, JNTU Hyderabad, TS, INDIA, Mail id: - shivanandam597@gmail.com

[2]Professor Dept of CSE School of Information Technology, JNTU Hyderabad, TS, INDIA, Mail id: - srmeda@jntuh.ac.in

[3]Associate Professor Dept of CSE School of Information Technology, JNTU Hyderabad, TS, INDIA, and Mail id: - gvr_reddi@jntuh.ac.in

Abstract

Information in spatial information bases are associated with spatial coordinates. Text pertinency and spatial proximity are considered for retrieval. It is mundane that the objects in a spatial information base (e.g., hotels/restaurants/parks) are associated with keyword(s) to represent their businesses/accommodations/features. Recent works on spatial information is predicated on a set of objects. An approach is to retrieve m most proximate objects gratifying certain utilizer designated conditions. Most proximate keyword search (Keyword cover search) is to query objects, which together cover a set of query keywords and the inter-objects distance is minimum. In which keyword rating given by different users is additionally considered. A formidable number of spatial indicators have been aimed to help spatial information retrieval. In keyword cover search user's current location is not designated. Location vigilant most proximate keyword search in spatial information can surmount this quandary by providing user's current location as input and provide a path to such a location. Travelling distance is considered to engender a path, rather than Euclidean distance. Additionally the results retrieved in keyword cover search may not be always desirable to user's cull. This critical drawback can be overcome by most proximate keyword search predicated on current location of utilizer.

Keywords: Spatial information base, point of interests, keywords, keyword rating, keyword cover.

1. Introduction

One of the most paramount kinds of queries in spatial information bases is location-predicated. Location predicated accommodations are mainly applied for most proximate Neighbors query. Given a particular location, and a set of points of intrigues, e.g. hotels, temples, parks etc., the most proximate Neighbor query returns the points of interest most proximate to the query point. Due to location-predicated accommodations and the availability of extensive digital maps plus satellite imagination (e.g., Google Maps), ye spatial keywords search quandary has gained much attention recently. Sundry information mining methods are applied to extract such kinds of points from astronomically immense spatial information bases.

A. Keyword search in spatial information

Spatial objects are associated with geo-location and a text description. Spatial keyword search is performed on spatial information. Keyword search is performed by applying queries on spatial information. In order to make expeditious retrieval sundry indexing methods are applied. Techniques that enable the indexing of information contains both text descriptions plus geo-locations which changes effective processing of spatial keyword queries that assume a geo-location plus a set of words as input and return pertinent content that matches the keywords. Spatial keyword queries are being fortified in authentic-life applications like Google Maps where points of interest can be retrieved. Spatial keyword search mainly conducts with textual information and spatial information . Spatial information includes our points of concern like hotels, restaurants, parks etc. And the keywords cognate to these objects are considered as textual information. Mainly there are two types of spatial keyword querying.

- Single object oriented
- Object set oriented

Single object oriented queries are predicated on the retrieval of a single type of object. Utilizer gives a single type of object and the retrieved result are predicated on user's location or given descriptions about keyword. Single object oriented includes following types of queries.

- kNN Query: recover the k objects closest to the user's current location (represented by a point).
- Top-k nearest neighbor Query: Top k objects with the highest ranking scores are retrieved, assessed as a compounding text verbal description to the query keywords and distance to the query location (a point).
- Range Query: Objects whose text verbal description contains userassigned keywords plus location is amongst a range specified by the user is retrieved.

A number of geo-textual indices are being. These indicators combine a spatial index and atext index structure.

Object set oriented is based on retrieval of group of different types of objects. m Closest Keyword (mCK), closest neighbor search and keyword cover search plus corporate spatial keyword query are admitted under this category. It includes retrieval of multiple types of objects satisfying certain query keywords.

- m Closest Keyword query: Retrieves m closest objects which can cover all the query keywords.
- Collective spatial keyword query: recalls collection of objectives in a particular location. User's location and keywords are given as input parameters.

The existing indices include R-Tree based indices, grid grounded indicators, space filling curve based indices etc. Along on the R-Tree established spatial indexing following text indexing can be used. Text indices are reversed file based, bitmap established and key signature file founded indices. In addition, some of the powers roughly combine a spatial and a text index while other indices resulting in hybrid index structures by tightly combining them.

B. Indexing methods

The main purport of storing an index is to ameliorate performance and speed in finding pertinent documents for a search query. Indexing enables expeditious retrieval of information . Signature files are word-oriented index structures and are predicated on hashing. And it is not felicitous for profoundly and astronomically immense texts. An inverted file is a word-oriented mechanism which sanctions indexing a text amassment in order to expedite the probing task. In hybrid indexing method all these information are augmented along with the nodes of a tree. Spatial information indexing methods are as follows.

- R-Tree
- R*-Treespacial

Hybrid indexing methods combines the spatial and textual indexing plus thus furnishes effective pruning power to the searching method.

C. R-Tree and R*-Tree

An R-Tree is akin to B-Tree and it is height-balanced [9]. B-Tree is a self-balancing information structure; information are sorted and can be utilized for searches, sequential access, insertions, and effacements in logarithmic time. It is is a generalized form of binary search tree in which a node can have more than two children. R-Tree is composed by aggregating Minimum Bounding Rectangles (MBRs) and the aggregates are stored in a tree structure. The aggregation is predicated on proximity of the objects or minimum bounding rectangles. The root node in an R-Tree has at least two ingresses u less it is a leaf node. R-Tree is predicated mainly on the minimization of area of each MBR.

In R*-Tree examines not only area but additionally several others such as minimization of perimeter, area, overlap between MBR and maximize storage utilization. Hybrid indexing methods amalgamates the spatial and textual indexing and thus provides efficient pruning power to the probing method. Figure 1, below shows an R-Tree example, where the rectangles in red, green and ebony color shows MBRs.

2. Related Work

2.1 SPATIAL KEYWORD SEARCHING METHODS

Keyword search on spatial information mainly based on various indexing methods and querying methods. Here briefly presents some of such spatial keyword searching approaches.

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 17
November 2016

A. Top-k spatial keyword search using IR2-Tree:

Felipe et al. [1] proposed an efficient search method to recover top-k spatial keyword. An index structure IR2-Tree (Information Retrieval R-Tree) that cumulates R-Tree and signature files to sanction keyword search on spatial information objects. A spatial keyword query consists of a set of keywords plus query field. The effect is a list of objects ranked according to their text pertinency and distance to the query area. Signature file is utilized for text indexing. A signature file is to perform membership tests to determine whether a query word w subsists

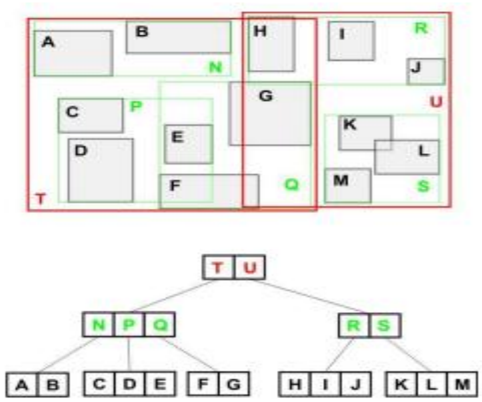

Fig. 1.R-Tree example.

In a set of words W. Represented as a sequence of bits (either 0 or 1). for each one bit represents the presence or absence of a term in a document. Bit 1 represents presence of a term and 0 otherwise. IR2-Tree is having some drawbacks, which impacted on its efficiency. The disadvantage of this method is high false convinced place. Figure 2, below shows signature file representation for a sample information set.

Algorithm for Distance Measure and Sorting

Input: user u, user_loc u($lat_{u,}$ $lon_{u,}$), result set loc { $lat_{i,}$ lon_j },

Output: Sorted result r.

Initialize distance d, collection c, double rad.

Let user u locations $lat_{u,}$ $lon_{u,}$

for($lat_{i,}$ lon_j ϵ(i,j) where i,j search results

rad= Math.sin(lat_u) * Math.sin(lat_i) + Math.cos(lat_u) * Math.cos(lat_i) * Math.cos(lon_u-lon_j);

d=rad * 180 / Math.PI;

c.add(d);

end for;

while(d:c)

in ascending of d,

r=get(c),

end while;

return r;

using KNN Euclidian

B. Top-k spatial keyword search using IR-Tree

Another indexing method is for processing the location-aware top-k text retrieval (LkT) query

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 17
November 2016

[2]. In which query placement is so soon determined and top-k results nearer to query location are retrieved. All these answers satisfy considerations afforded by the user. Both of the works [2], [3] uses IR-Tree based indexing. This framework integrates R-Tree and the inverted file. Here an IR-Tree indexing is proposed that is mainly an R-Tree extended with inverted files.

3. Implementation

3.1 Proposed work:

In keyword-NNE algorithm one query keyword is culled as principal query keyword, and the objects retrieved are more proximate to this principal query keyword. So the query point will be the principal object. And the inter object distance from this point to other points of fascinates should be minimum. The result places are more proximate to the principal object. Principal query keyword is culled as the one in which number of objects will be minimum. Albeit the method keyword-NNE outperforms, it faces the following inhibitions.

Most of the geographic studies use distance as a simple measure of accessibility. Straight-line (Euclidean) distance is most often utilized in spatial information bases because of the facileness of its calculation. Genuine peregrinate distance over a road network is a better alternative, albeit historically an extravagant and labour intensive task. This is erroneous always, because utilizing commercial website one can directly compute time and distance, without the desideratum to own or purchase specialized GIS software or street files.

Algorithm for BKC

Input: Item set I, Item Type T, Item Location (x, y), Item Rating R

Output: Result RS.

Initialization:

α Threshold of rating.

β Threshold of distance between Items .

let Item set [T_1, T_2, $T_{3\ ....}$ T_m]

let Item set [I_1, I_2, $I_{3\ ....}$ I_i]

for each $T_m \in T$

  for each $I_i \in I$

    if $I_{i..n}(R) \geq \alpha$

      if Dist ($I_{i...n,}$ $I_{i...n}$) $\geq \beta$

        Set S= $I_{i...n,}$ $I_{i...n;}$

        RS=S;

      end if;

    end if;

  end for;

end for;

return RS;

Capitalizing on this feature, compare straight-line and peregrinate distance and peregrinate time to calculate distance between query point and other nearby locations. A major constraint of keyword-NNE is that utilizer cannot designate his current location. So that the query does not retrieve distance of the path from user's current location to principal object in GBKC. In lieu of taking euclidean distance from user's current location to the query point, travelling distance and time is calculated.

Because euclidean distance may not always give a precise result as utilizer expected. Let Ok be the set of principal objects under principal query keyword k. ok ∈ Ok be the principal object in GBKCk. Distance of ok to the user's current location L is not designated in this method. Shortest travelling distance of the path taken by utilizer from location L to the principal object in Ecumenical Best Keyword Cover can be obtained utilizing Google API [14]. Integrating this feature can make the probing more utilizer cordial and give more support for a peregrinator in good decision making. Another quandary with the keyword-NNE method is that algorithm set one query keyword with minimum number of objects as principal keyword. So that the retrieved results are circumvented by this keyword.

Utilizer cannot give principal query keyword according to his own cull. Suppose a utilizer wants to ken locations more proximate to non-principal object, such provision is not provided in this algorithm. In current location predicated most proximate keyword search utilizer can set any keyword as principal query keyword according to his cull. In lieu of culling the one with minimum number of objects, utilizer can set principal keyword as the first entered keyword. The method can retrieve identically tantamount result (GBKC) as keyword-NNE. Along with that result utilizer can cull an object in GBKC and can probe user's intrigued keyword more proximate to that culled object.

3.2 Location aware closest keyword search in spatial information:

The method is predicated on current location of utilizer. Utilizer designates his points of interest and current location. After calculating GBKC, the system returns an itinerary (an orchestrated route) covering user's current location and POIs (Points of Interest) initially designating current location of utilizer. Utilizing Geocoding API, corresponding address is converted to its latitude and longitude. From the current location most proximate object in GBKC is calculated, and the process perpetuates up to the last object. All these Points of Fascinates are represented as waypoints in map. Waypoints designate an array of points. It can alter a route by routing it through the designated location(s). A waypoint is designated as a latitude/longitude coordinate, an encoded polyline, a place ID, or an address which will be geocoded. A path covering all these waypoints is engendered. So the method engenders an itinerary (an orchestrated route) covering users current location and all objects in GBKC.

4.Experimental Work

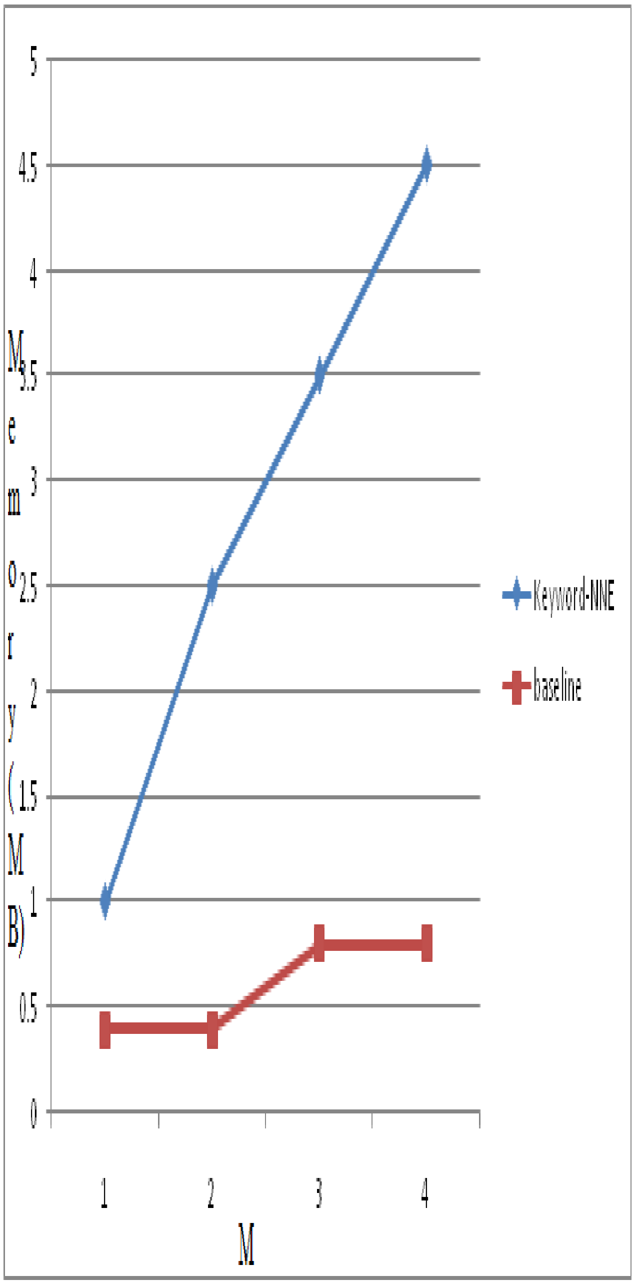

Fig 1Maximum memory consumed versus m (a ¼ 0.4).

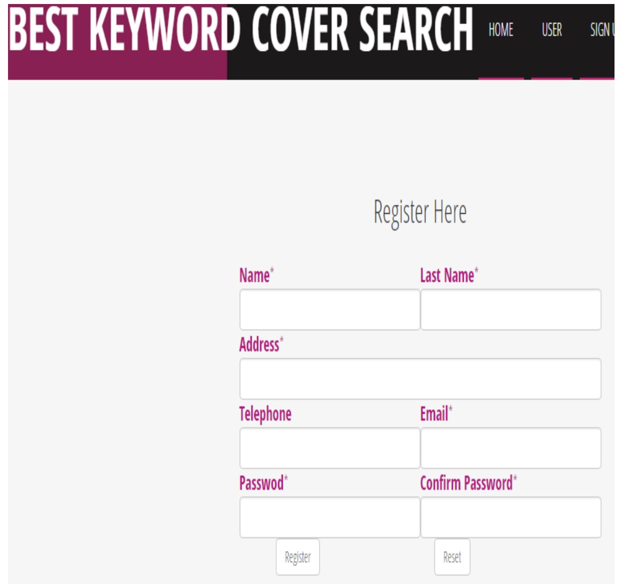

Fig 2: User Authentication Page.

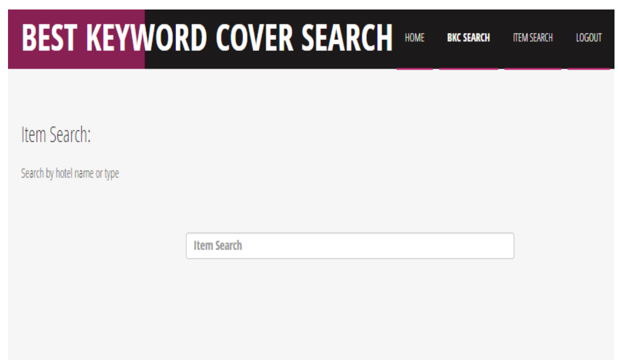

Fig 3: User Keyword Search Page.

Result Data:

Here, We collect rating and distance of 20 items which include restaurants, hotel, and bar in hyderabad city. We can see datasets having minimum 6 rating out of 10, and distance between items not more than 5 kilometers. We set T-value for rating is 6 and set T-value for distance is 5 km.

r1	r2	dist_	set_	city
10	10	0.034445468741273244	r1,h3	Hyderabad
6	10	2.125423063607505	r5,h3	Hyderabad
10	6	1.5345161156773186	r4,h3	Hyderabad
10	10	0.034445468741273244	r3,h3	Hyderabad
10	6	0.0	r1,b1	Hyderabad
10	10	0.04097215169298834	r1,b3	Hyderabad
6	10	2.0910343303392485	r5,b1	Hyderabad
6	10	2.099178483409685	r5,b3	Hyderabad
10	10	1.5553497400044018	r4,b1	Hyderabad
10	10	1.5200520176180228	r4,b3	Hyderabad
10	10	0.0	r3,b1	Hyderabad
10	8	0.04097215169298834	r3,b3	Hyderabad
10	10	0.034445468741273244	h3,b1	Hyderabad
10	9	0.049921470865877234	h3,b3	Hyderabad
10	7	0.034445468741273244	r1,h3,b1	Hyderabad
6	10	0.034445468741273244	r5,h3,b1	Hyderabad
10	10	0.034445468741273244	r4,h3,b1	Hyderabad
10	10	0.034445468741273244	r3,h3,b1	Hyderabad
10	10	0.049921470865877234	r1,h3,b3	Hyderabad
7	10	0.049921470865877234	r5,h3,b3	Hyderabad
10	9	0.049921470865877234	r4,h3,b3	Hyderabad
10	10	0.049921470865877234	r3,h3,b3	Hyderabad

5. Conclusion

Keyword Cover query provides an incipient amended spatial keyword probing method, in which keyword evaluation is con-sidered. The method is connoted on keyword- NNE algorithm. In location cognizant most proximate keyword search in spatial information user's current location is additionally taken. So the method is predicated on locationcognizant most proximate spatial keyword querying. Along with the results of keyword-NNE, user's current location and distance of the path is retrieved. To get a path to users points of interest Google waypoints are utilized. In lieu of Euclidean distance, travelling distance is calculated to get precise results. So that an itinerary covering all the POIs is engendered and the peregrinator can make sensible decision to arrange a peregrination to his fascinated location. The scope of future work relies on the following cases. The default mode of conveyance is driving; all other modes (ambulating, bicycling, and transit) have to be considered in future. Additionally multiple routes covering users POIs have to be considered.

6. References

[1] D. Felipe, V. Hristidis, and N. Rishe,"Keyword search on spatial information bases," IEEE 24th Int. Conf. Information Eng., pp.656-665, 2008.

[2] G. Cong, C. Jensen, and D. Wu, "Efficient retrieval of the top-k most relevant spatial web objects," Proc. VLDBEndowment, vol. 2, no. 1, pp. 337348, Aug. 2009.

[3] Z. Li, K. C. Lee, B. Zheng, W.-C. Lee, D. Lee, and X. Wang,"IRTree: An efficient index for geographic documentsearch," IEEE Trans. Knowl. Information Eng., vol. 99, no. 4, pp. 585-599, Apr. 2010.

[4] R. Hariharan, B. Hore, C. Li, and S. Mehrotra, "Spatial-keyword (SK) queries in geographic informationretrieval(GIR) systems," IEEE 19th Int. Conf. on SSDBM, 2007.

[5] D. Zhang, Y. Chee, A. Mondal, A. Tung, and M. Kitsuregawa.,"Keyword searches in spatial information bases: Towardssearching by document," IEEE Int. Conf. Information Eng., pp. 688-699, 2009.

[6] D. Zhang, B. Ooi, and A. Tung,"Locating mapped resources in web 2.0," IEEE 26th Int. Conf. Information Eng., pp. 521-532, 2010.

[7] Ke Deng, Xin Li, Jiaheng Lu, and XiaofangZhou,"Best Keyword Cover Search," IEEE Trans. Knowl.Information Eng.,vol. 27, January 2015.

[8] Zhicheng Li, Hu Xu, Yansheng Lu, Ailing Qian,"Aggregate Nearest Keyword Search in Spatial Information bases," IEEE12th International Asia-Pacific Web Conference, 2010.

[9] Gnttman A,"R-trees: A Dynamic Index Structure for Spatial Searching," Proc. ACM SIG MOD Int. Coaf. OnManagement, pp. 47-57, 1984.