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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Keyword Search and Geographic aspect Spatial Top k Search

Harsh Anand, Monika Kumari, Ashwini Vidyadhar Hiremath, Ashwini Yuvraj Mane

B.E. Student, Dept. of Computer Science, Sinhgad Institute of technology, Lonavala , Pune, India

B.E. Student, Dept. of Computer Science, Sinhgad Institute of technology, Lonavala , Pune, India

B.E. Student, Dept. of Computer Science, Sinhgad Institute of technology, Lonavala , Pune, India

B.E. Student, Dept. of Computer Science, Sinhgad Institute of technology, Lonavala , Pune, India

ABSTRACT:with the increasing pervasiveness of the geo-positioning technologies and geo-location services, there area large amount of spatio-textual objects offered in several applications. That type of data provides the placement connected data relating to the amount of hotels, Hospitals. This data obtained by the keyword search. This explores the helpful data relating to the actual locations or places. K-nearest neighbor (k-NN) queries Most of the present approaches to the current drawback are processing of sets of TOPK-SK queries. Supported the inverted index and also the linear quad tree, we propose a completely unique index structure. To address this drawback, we propose a set of solutions which will support inverted linear quad tree process of k-NN queries. We propose initial new index structure referred to as Dynamic Strip Index (DSI), which may higher adapt to completely different information distributions than exiting grid indexes. We proposed investigate the matter of conducting prime k spatial keyword search. We proposed to additional propose a distributed k-NN search (DKNN) rule supported DSI. The advances of GPS technology and wide-ranging usage of wireless communication devices have expedited the gathering of huge quantity of spatiotemporal information. Of interest group is information related to the location keyword.

KEYWORDS: Spatial, Keyword, Batch, k nearest neighbour query, spatial keyword query, scalability

I. INTRODUCTION

Thus, the investigation of spatial keyword search look which investigates both area and textual information of the locations has pulled in consideration from the organizations associations and research groups. Spatio-temporal location collected in numerous applications, for example, location based administrations and informal organizations, in which a search is retrieved by its spatial keyword and an arrangement of spatial keyword. Numerous location applications (e.g., area based), the nearest of two spatial items is measured by the street distance. Also, the outcome is becoming into a general practice to upgrade the nature of the query items. We research the issue of spatial top k spatial keyword search (TOPK-SK) that is, given an group of spatio-temporal items, an inquiry area question and answer set of keywords, we expect to find the k nearest questions each of which contains all keyword in the search. In the paper, we to investigate the matter of leading top k keyword search (TOPK SK) that's, given a gathering of spatio textual objects, Query a question area question and answer gathering of keywords, we to plan to retrieve the k closest objects everything about contains all keywords inside the question. The most related k location based keyword search is enter in abstract keywords searches and incorporates a wide range of location based application.

II. BACKGROUND OF USE

Yinghua Zhou et.al. States that the number of locations available online, for the location based services. There are search schemes which find the location details by indexing or searching related information on the net. Location search in the multidimensional and in Euclidean distance space. Described the problem of the current system in which how they can effectively retrieve the data which is on the indexes. Proposed hybrid indexing structure which contains 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)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

inverse indexing and the R* trees for the handling the data related to the places. Studied the different approaches in that inversed file and R*tree double index. Second inversed file the R* tree and lastly first R* tree then inverted to the file. This scheme has proven efficient based on the indexed grid structure [1].

Spatio-temporal places and their related keywords gets the by accessing their information and location details. To handle large amount of data such as e.g twitter and flicker. To retrieve the information by using the type of service requires long processing time. Rocha-Junior et.al proposed the spatial inverted index. For the top k query search. In this approach each and every object mapped with the objects and stored manually. And these objects are retrieved efficient by using this approach. This approach experimented on the low query cost [2].

There are number of location based services available on the sites. In which the number of keywords are used to retrieve the related information from those sites. Chengyuan Zhang et.al. The nearest neighbor search techniques used to find the nearest places from the database. Nearest neighbor finds based on the road distance. To enhance the data retrieval the problem of diversity of the data retrieval studied. Proposed the signature based indexing technique t0 search the spatial keyword by query retrieval system [3].

GIS dataset contains lots of information regarding the location based information. The location has the lots of information such as the textual information as well as the geo spatial information. Eg. Find the medical store from the some city. Data retrieved the information by the spatial keyword queries. The Geographical information retrieval system proposed in which searching scheme build on the number of the record dataset. Ramaswamy Hariharan et.al. Proposed the model for the spatial queries called GIR frameworks [4].

GÍSLI R compare the R* and the K nearest neighbors search. Compared the performances of the both the algorithms.[5]

There are number of devices are available which supports the location based GPS enabled services. Each location has the all textual information which defines the location in detail such as location name address and the geo spatial information latitude and the longitude. The spatial query returns the all information based on the search keyword ranked on the nearest distance.in the proposed on the R* and inverted index [6]. Formalized query type indexing way to search the locations.

Mohammad Sadoghi et.al. studied the tree structure called BE tree for indexing over the large scale dataset. BE Tree generates the expression structure to find the more relevant data [7].

Mohammad Sadoghi et.al presents the methodology to extract the data efficiently is the spatial top k search. This methodology extracts the methodology and combines with the R tree for fast processing of the data. These tree are works onthe best first basis and access the scalable data with spatial search queries [8].

Gao Cong et.al stated on the web there are large amount of data available regarding the geographical data. This has the textual and geospatial information. While searching on the net it finds the relevant data on the net and shows the results. By motivating this they proposed the indexing model for searching the location on the internet called top-k search for textual data and R tree for spatial data retrieval. Proposed the top k search for finding the text similarity and spatial proximity [9].

Dongxiang Zhang et.al studied on the daily basis there increasing the location based services. Top k search finding the most relevant places and gets the locations on the indexing basis or relevance calculated on the distance between the locations. Mainly combine the R-tree and the inverted index so that spatial pruning and textual pruning can be executed simultaneously. They proposed the scalable inverted index structure and designed the storage location methodology to effectively retrieve the locations. This in term requires less time to fetch details [10].

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

III. PROPOSED SYSTEM

Below figure shows the architecture of the of the proposed system. In the architecture it is divided in part such as database in which it contains the datasets of establishments.

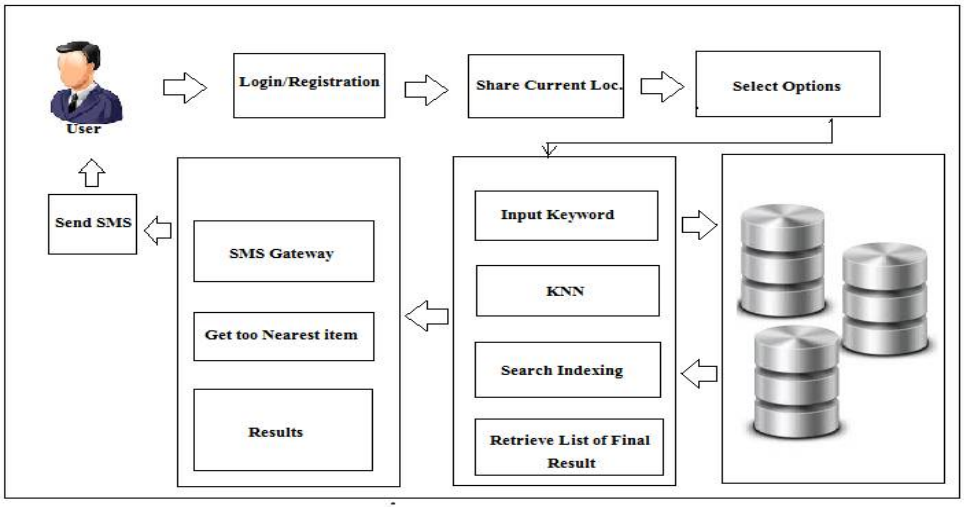

Fig. System Architecture.

Above figure shows the architecture of the of the proposed system. In the architecture it is divided in part such as database in which it contains the datasets of establishments. Other part if algorithmic part and third is the sms gateway. Inthe proposed system we process the data to find the nearest places and recommend with offer available at the particular establishment. To retrieve data efficiently, in the proposed system Search Indexing technique is used to fast data retrieval with the help of KNN Algorithm. Proposed system starts from the user registration and like real-time site. User login to the application the he will get the choices available to application. To find the nearest and better places user has to share his current location, also select the option which service he wants. Current location and the choice of use is the given input to the KNN algorithm to find the nearest places accordingly. In this process SI technique is useful to find data efficiently. Lastly we can recommend the service based and the service available at location. And finally we maintain the ratings of the service to use in future.

Algorithm: KNN

Input: Dataset and Sample
Output: Nearest neighbour (places)
Steps:
1. kNN (dataset, sample){
2. Go through each item in my dataset, and calculate the "distance" From that data item to my specific sample.
3. Classify the sample as the majority class between K samples in the dataset having minimum distance to the sample.
4.}
5.**Return NN**.

KNN algorithms Pseudocode as show in the above algorithm 1. The dataset of the latitude and longitude of the establishments training dataset and the sample given input to the algorithm. It will go through all the data and find the similar and nearest places based on the distance. All nearest places are inserted into data structure. After that establishment who has minimum distance from the sample input retrieved and return as the output.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

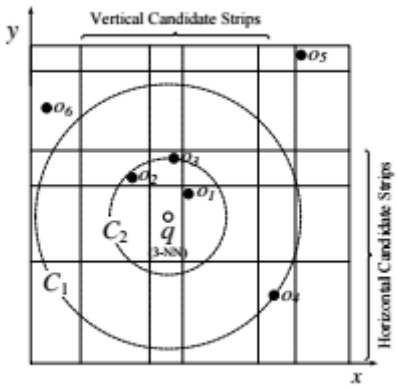

Fig. Search Indexing

Figure shows the Dynamic Strip Indexing called as the search indexing. To fast retrieval of data and efficiently to used techniques. In the proposed system for the indexing this technique is used.

IV. MATHEMATICAL EXPRESSION

S={ O = {o1, . . . , on} , R 2, loc={lat,lng}, D={lat, lng e, add, scheme} SI};
Where,
O= set of points
On= position lat and longitude of user
posi= spatial positions
Pi= the time-ordered sequence of position updates issued by oi
ptj i =position update
R 2= Euclidean Distance
qi= generic k-NN query +SI
Finds the nearest location by using the KNN and search indexing technique.
k =closest objects based on the current location of the user.
(x, y)= center
Get nearest neighbor who has minimum distance.
Scheme which are at the e establishment which is sent user through sms.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

V. SIMULATION RESULTS

Below pictures shows the final working and output of our project.

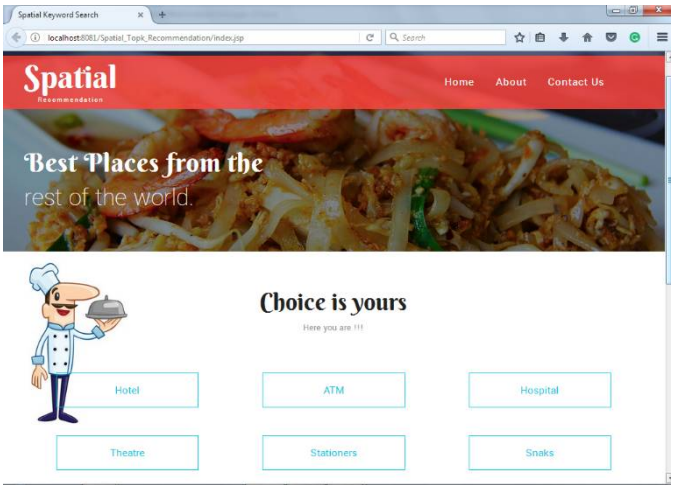

Fig. Result 1

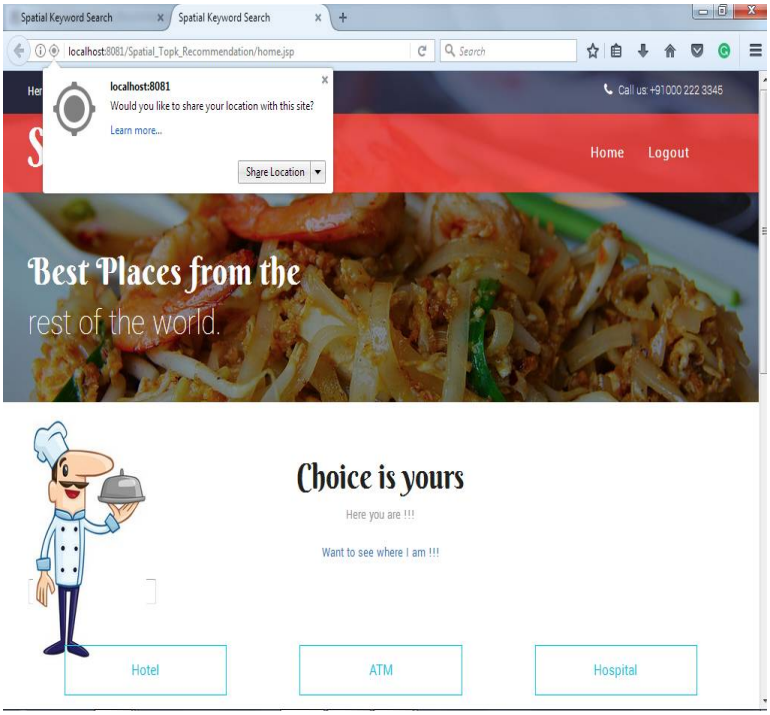

Fig. Result 2

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

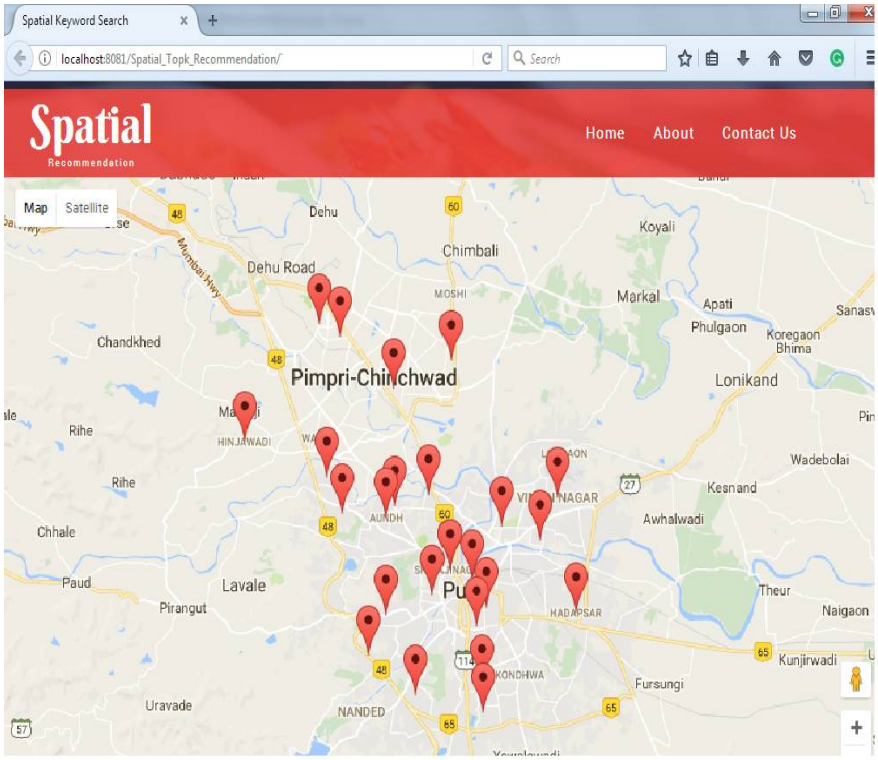

Fig. Result 3

VI. CONCLUSION AND FUTURE WORK

The problem of the K- Nearest queries in the large amount of growable database in the large applications. In this paper we propose Dynamic Search Indexing over the KNN queries over the distributed processing. In the proposed system we used the SI data partition technique to improve the efficiency of the data retrieval. Proposed system determines a nearest.

R 2= Euclidean Distance

Establishment that is exactly o contains the k -NNs for a given query.

REFERENCES

1. Y.Zhou, X. Xie, C. Wang, Y. Gong, and W.-Y. Ma, "Hybrid index structures for location-based web search," in *CIKM*, 2005.
2. JJ. B. Rocha-Junior, O. Gkorgkas, S. Jonassen, and K. Nørvag, "Efficient processing of top-k spatial keyword queries," in *SSTD*, 2011
3. C. Zhang, Y. Zhang, W. Zhang, X. Lin, M. A. Cheema, and X. Wang, "Diversified spatial keyword search on road networks," pp. 367–378, 2014.
4. R. Hariharan, B. Hore, C. Li, and S. Mehrotra, "Processing spatialkeyword (sk) queries in geographic information retrieval (gir) systems," in *SSDBM*, 2007
5. G. R. Hjaltason and H. Samet, "Distance browsing in spatial databases." *TODS*, vol. 24, no. 2, pp. 265–318, 1999
6. J. B. Rocha-Junior and K. Nørvag, "Top-k spatial keyword queries on ° road networks," in *EDBT*, 2012
7. M. Sadoghi and H. Jacobsen, "Be-tree: an index structure to efficiently match boolean expressions over high-dimensional discrete space," in *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2011, Athens, Greece, June 12-16, 2011*, 2011, pp. 637–648. [Online]. Available: http://doi.acm.org/10.1145/1989323.1989390.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

8. A. Cary, O. Wolfson, and N. Rishe, "Efficient and scalable method for processing top-k spatial boolean queries," in *SSDBM*, 2010, pp. 87–95.
9. G. Cong, C. S. Jensen, and D. Wu, "Efficient retrieval of the top-k most relevant spatial web objects," *PVLDB*, vol. 2, no. 1, 2009
10. D. Zhang, K.-L. Tan, and A. K. H. Tung, "Scalable top-k spatial keyword search," in *EDBT*, 2013, pp. 359–370
11. Z Yu, X Yu,Ken Q. Pu,Y Liu,"Scalable Distributed Processing of KNearest Neighbor Queries over Moving Objects" 2013 IEEE

BIOGRAPHY

Harsh Anand (*1*), Monika Kumari (*2*), Ashwini Vidyadhar Hiremath (*3*), Ashwini Yuvraj Mane (*4*)is a student in the computer engineering Department, sinhgad institute of technology, Pune university.