ISSN: 2321-7782 (Online)
Volume 4, Issue 1, January 2016
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Nearest Neighbor Search Technique for Spatial Database

Zarreen Anjum[1]
Department of Computer Science & Engg. & Technology
G.H. Raisoni Institute of Engineering & Technology for Women
Nagpur, India

Prof. P. S. Mohod[2]
Department of Computer Science & Engg. & Technology
G.H. Raisoni Institute of Engineering & Technology for Women
Nagpur, India

Abstract: Spatial queries with conventional methods like range search or nearest neighbor search, involve only the use of geometric properties of the objects. But nowadays many modern applications select the objects based on both properties: geometric coordinates and texts associated with it. An existing solution for searching difficult queries uses IR-tree method and we have discussed it in the paper. We propose a new method with an objective of finding the nearest neighbor of the query while reducing the delay time incurred in searching and enchasing the of the result of a query.

Key words: Nearest Neighbor Search, Spatial Queries, IR-tree, Spatial Database.

I. INTRODUCTION

Multidimensional objects such as points, rectangle, etc are managed by spatial databases. The spatial database provides access to the multidimensional objects by using various selection criteria. Real world entities can be modeled in geometric objects by using the help of spatial databases. For example, restaurants, hospitals, hotel's locations are represented by a point in a map whereas larger natural bodies like lakes, parks, etc are represented by rectangle. The functionalities of spatial databases are useful in many applications.

Now-a-days everyone is using the advantage of search engines for nearly all the problems. The use of search engines is done by giving spatial queries. Writing queries is becoming more simpler and easier. Traditional queries always focused on the geometric properties of the object. Modern applications select the object based on both properties: geometric coordinates and texts associated with it. For example: a user would want to search a restaurant that serves coffee and buns at the same time. The search engine results by providing a restaurant that may not be globally nearest but the nearest restaurant among many who servers the demanded food and beverage.

There are many easy ways to brace the queries that combine spatial and text features. For instance in the above discussed example, the search engine first searches the restaurants that have menus with keywords {coffee and buns} and then from the searched restaurants it chooses the one which is closest and nearest. The process can be reversed by first searching restaurants those are geographically near and then search a restaurant whose menu has the keywords. But the disadvantage of this straightforward method is that it fails to provide real time answers on harder inputs. There is not more research done on spatial queries with keywords. But recently the community has shown some interest on studying keyword search in rational database.

II. Proposed Methodology

The proposed work is based on R-tree and performs searching operation on it. The system architecture is as shown in figure1.

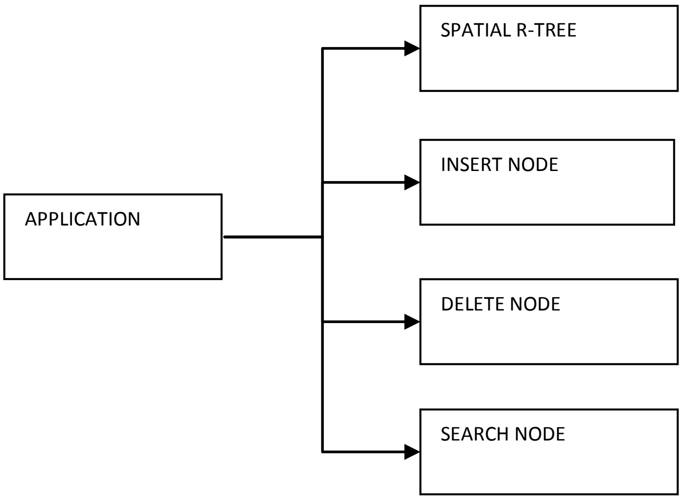

Figure 1: System Architecture

The methodologies used in proposed system are described as below and the flow of the proposed system is depicted in figure 2.

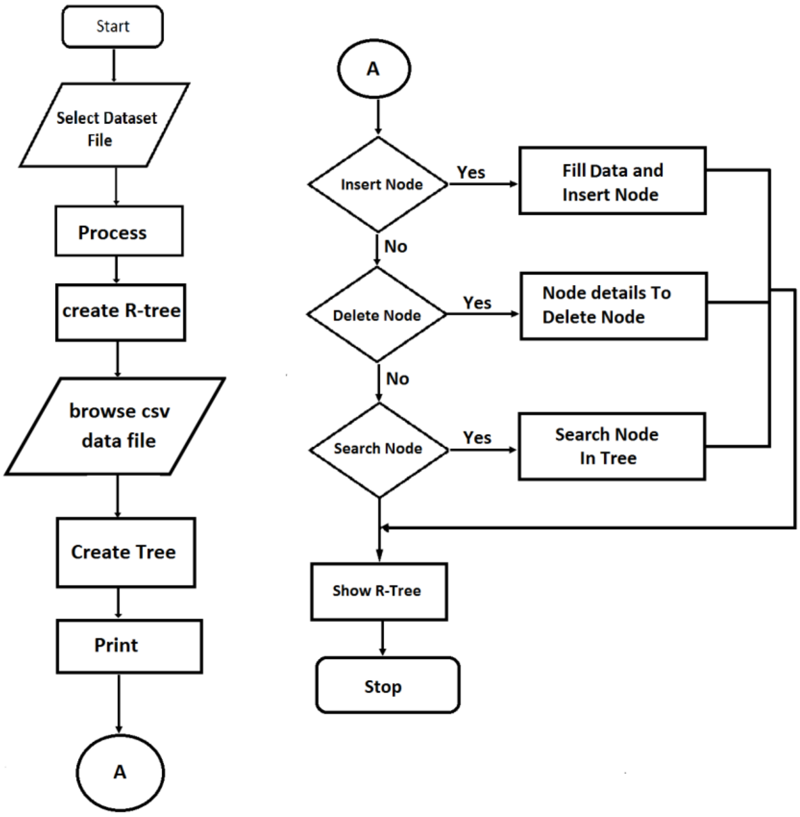

Figure 2: System Flow

A. *R-Tree*

R-trees are tree data structures used for spatial access methods, i.e., for indexing multi-dimensional information such as geographical coordinates, rectangles or polygons.The key idea of the data structure is to group nearby objects and represent them with their minimum bounding rectangle in the next higher level of the tree; the "R" in R-tree is for rectangle. Since all objects lie within this bounding rectangle, a query that does not intersect the bounding rectangle also cannot intersect any of the contained objects. At the leaf level, each rectangle describes a single object; at higher levels the aggregation of an increasing number of objects.

B. Processing spatial queries

Spatial queries will be processed with the help of group nearest neighbor query technique. In group nearest neighbor query user will give input for particular point. Given two sets of points P and Q, a group nearest neighbor (GNN) query retrieves the point(s) of P with the smallest sum of distances to all points in Q.

C. Minimum Bounding Method

The minimum bounding method (MBM) performs a single query, but uses the minimum bounding rectangle to prune the search space. Specifically, starting from the root of the R-tree for dataset, MBM visits only nodes that may contain candidate points.

III. EXPERIMENTAL SETUP

First collect datasets containing points and save them in the database. Then we choose this .txt file for processing as shown by fig. 3 by creating tables which include various fields. The points are to be inserted in those fields. It is called dataset generation. In which we make database and we can import dataset file. Figure4 shows the output after processing the datasets.

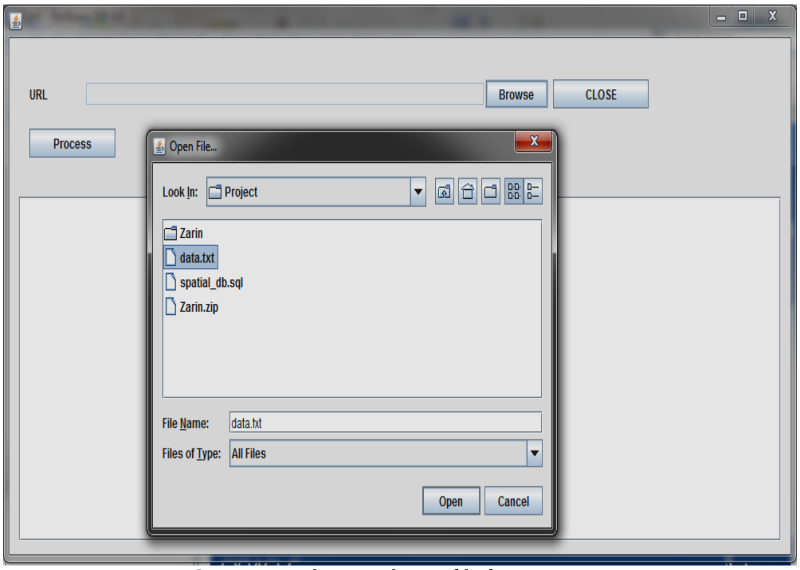

Fig 3 : Importing location dataset file for processing

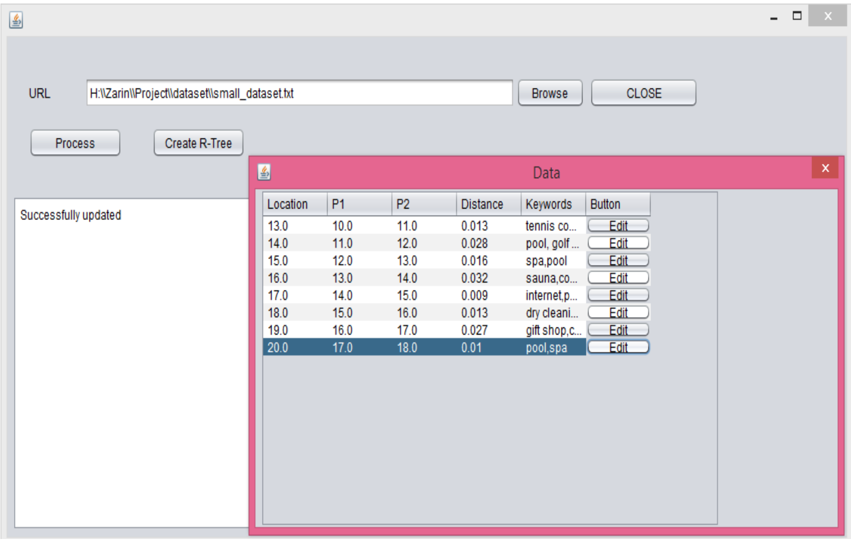

Fig4: Processing location datasets

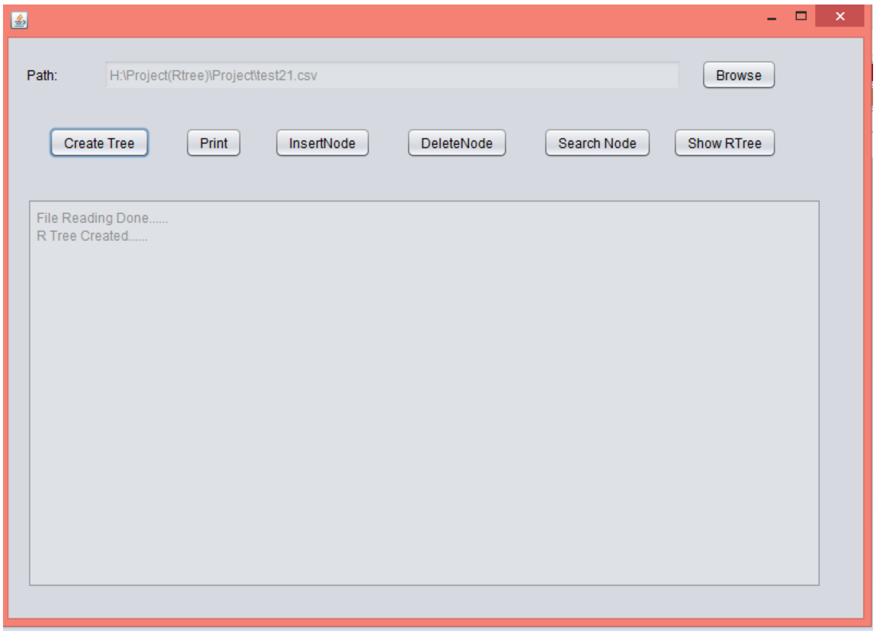

Fig5 : R-tree created on datasets

After creating the tree, we will print the R-tree. As shown by fig 6, it will display the record id , latitude, longitude of location, MBR and value of the location that describe the location.

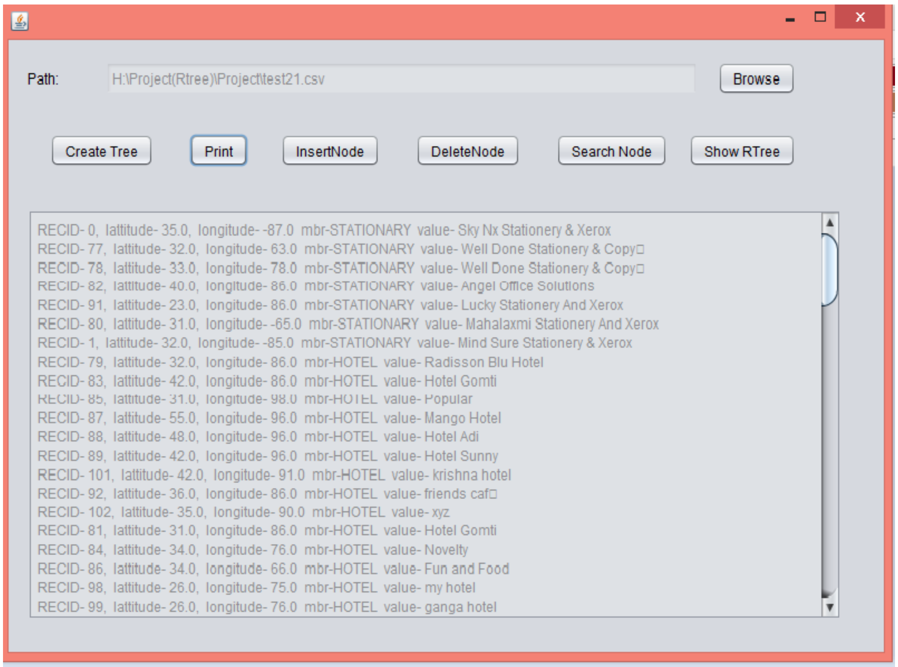

Fig 6: Printing R-tree

The below fig 7 shows the structure of R-tree. In this the root node is MBR i.e Stationary and Hotel, and in the child node we have put the locations. All the nearby objects are group together and they are represented with their minimum bounding rectangle.

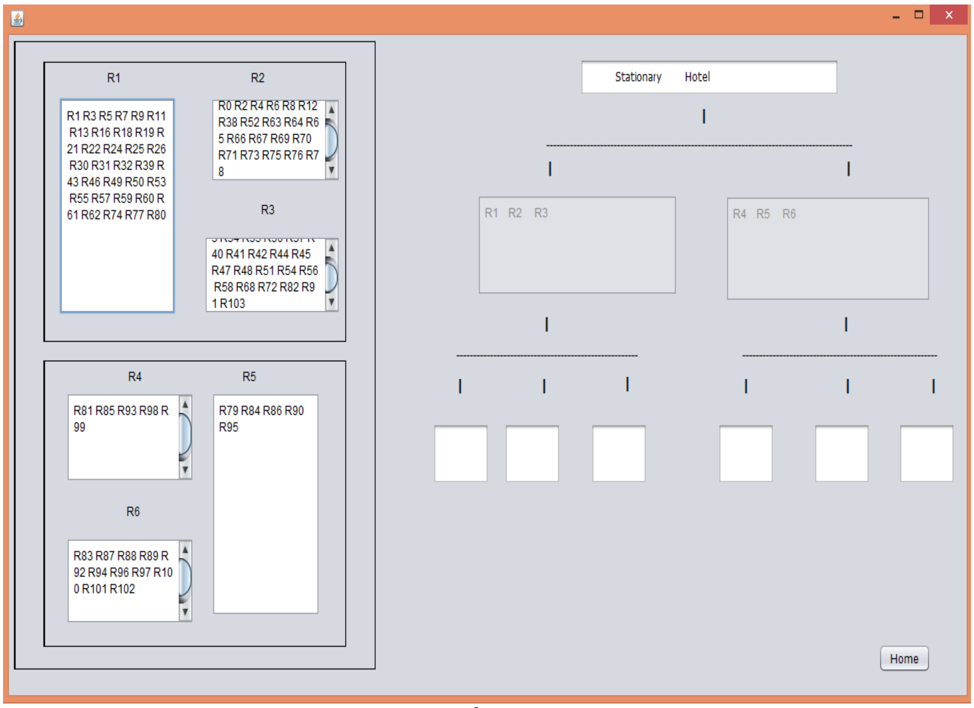

Fig 7: Displaying R-Tree

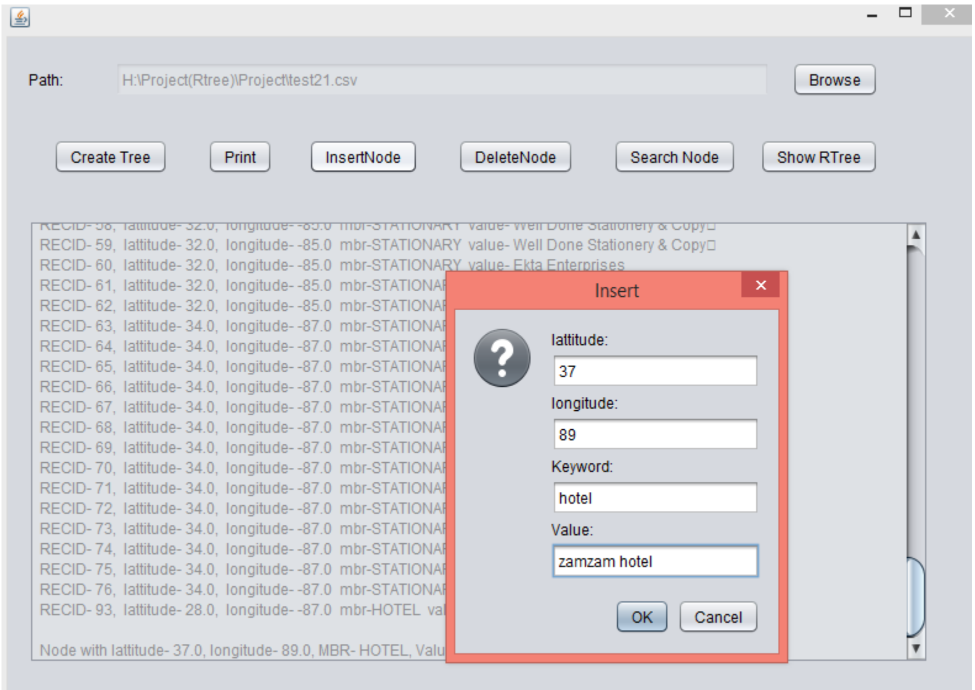

fig 8: Insert node in database

For inserting any node in the database we have to enter location's latitude, longitude, keyword and value that describe the location as shown by fig 8. After this, these information are updated into the database.

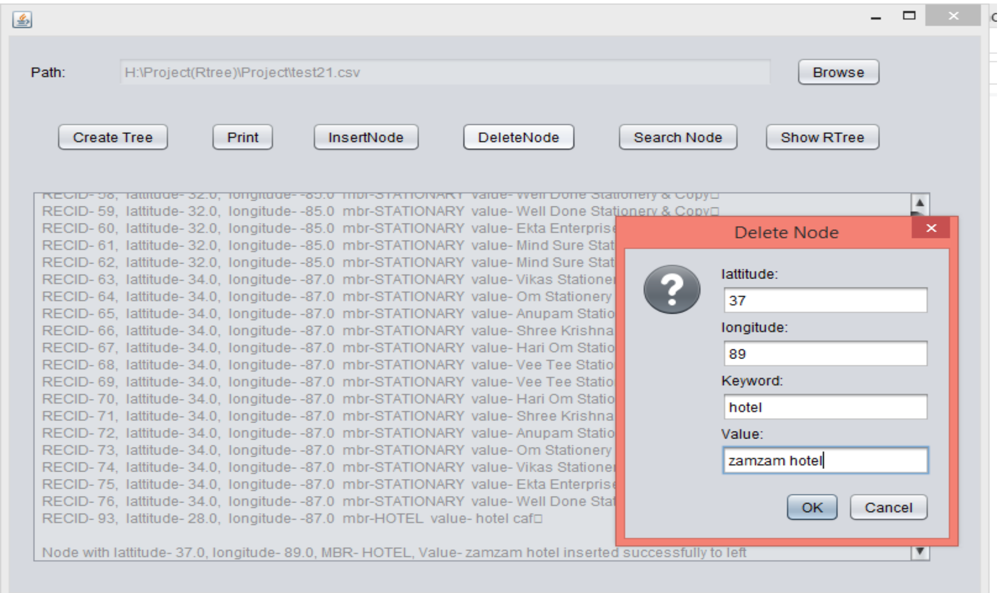

Fig9: Delete Node in database

For deleting the node we well enter the values for latitude, longitude, keyword and value of the location, then it will search the node in the database and if the node is found then the node will be deleted from the database as shown by fig 9.

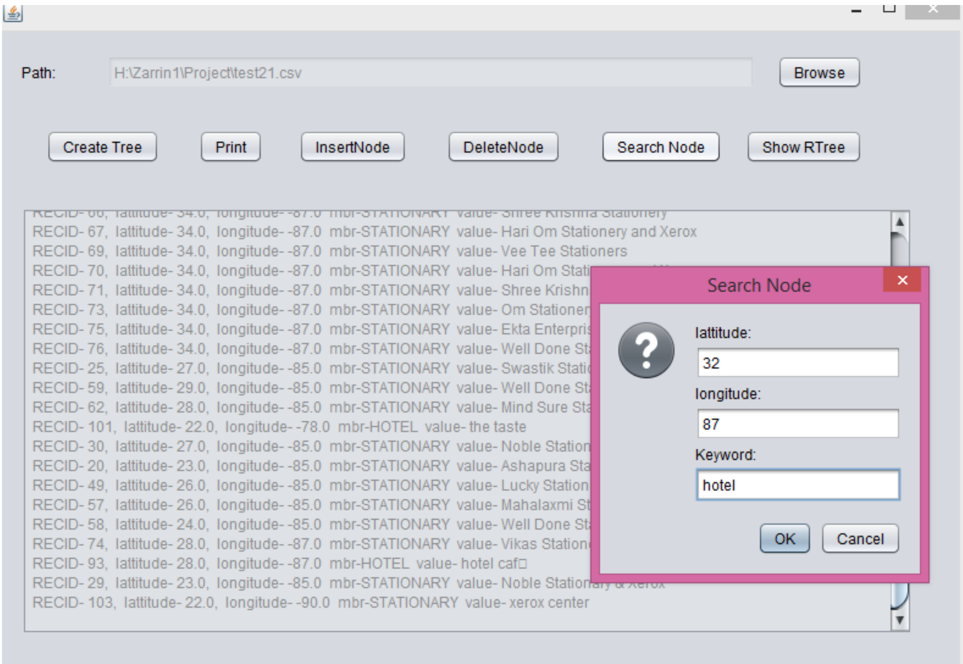

Fig10: Searching Location in database

For analyzing the query particular location will be searched, for this latitude longitude and keyword of the location will be entered and in the output nearest location of that query will be displayed as shown by fig 10.

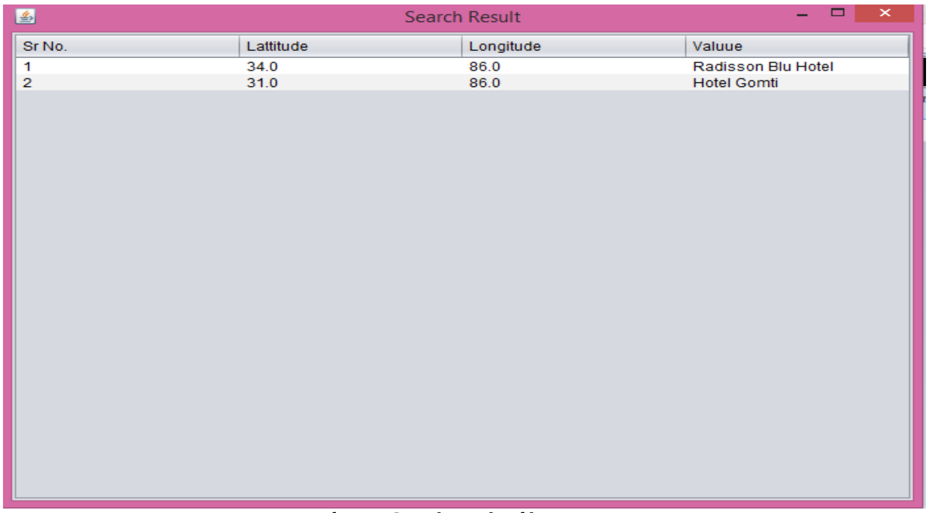

fig 11: Search result of location

IV. Result Analysis

Experimentally evaluate the practicalefficiency for solutions to NN search with keywords. Real spatial static datasets are used that catalogs vast amount of spatial data objects. We employ a number of pre-processing steps to arrive at a more structured database. The downloaded database is heterogeneous in nature. For each database, we identified and discarded the real valued fields and retain the spatial and text field. Now our database is more structured and homogenous containing only spatial and textual data.

We use different size of datasets in our experiment and calculate its execution time (in millisecond), standard deviationand as shown in fig 12.

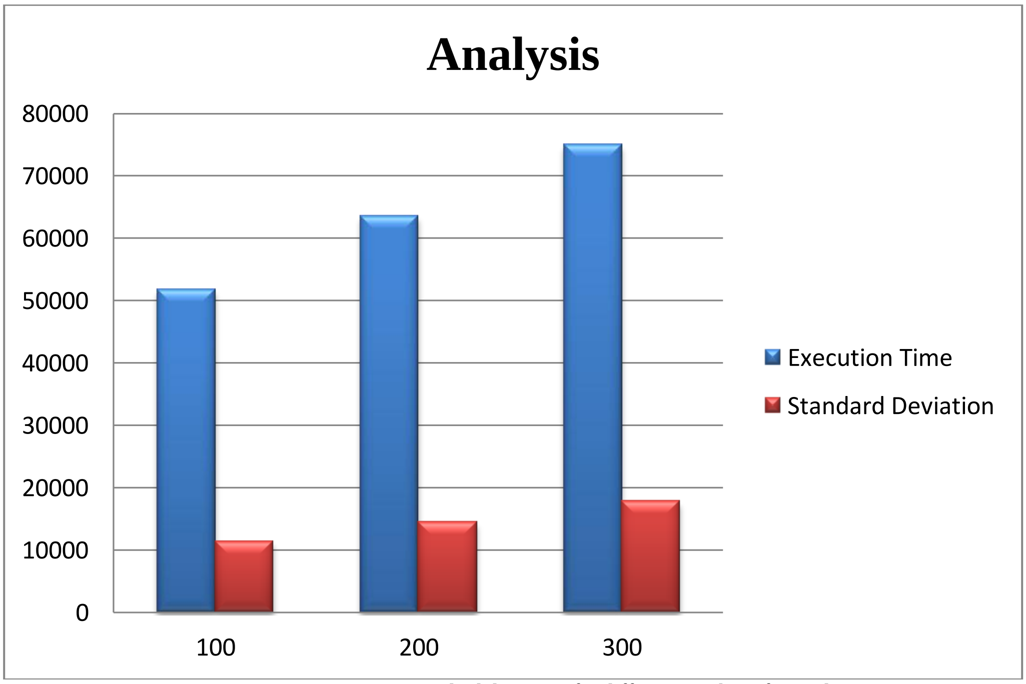

Fig12: Execution time & Standard deviation for different number of records

V. Conclusion

An indexing structure R-tree is designed for spatial access method and to answer the nearest neighbor query with keyword. As spatial database consist of large spatial objects, the time required for searching objects is more. By combining R-tree technique with minimum bounding method the performance of system for retrieving a data from database is improved, also access time is minimized.

References

1. Ke Deng, Xin Li, Jiaheng Lu, and Xiaofang Zhou, "Best Keyword Cover Search", IEEE Transactions on Knowledge and Data Engineering, VOL. 27, NO. 1, January 2015
2. Yufei Tao and Cheng Sheng, "Fast Nearest Neighbor Search with Keywords", IEEE Transactions on Knowledge and Data Engineering, VOL. 26, NO. 4, APRIL 2014.
3. Ying Zhang, Wenjie Zhang, Qianlu Lin, Xuemin Lin, Heng Tao Shen, " Effectively Indexing the Multidimensional Uncertain Objects", IEEE Transactions on Knowledge and Data Engineeringvol 26 Issue No.03 March 2014
4. X. Cao, G. Cong, C.S. Jensen, and B.C. Ooi, "Collective Spatial Keyword Querying," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 373-384, 2011.
5. J. Lu, Y. Lu, and G. Cong, "Reverse Spatial and Textual k Nearest Neighbor Search," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 349-360, 2011.
6. Feifei Li, Member, IEEE, Bin Yao, Student Member, IEEE, and Piyush Kumar, "Group Enclosing Queries", Transactions on Knowledge and Data Engineering, VOL. 23, NO. 10, October 2011.
7. X. Cao, G. Cong, and C. Jensen, "Retrieving top-k prestige-basedrelevant spatial webobjects," Proc. VLDB Endowment, vol. 3, nos.1/2, pp. 373–384, Sep. 2010.
8. Ken C.K. Lee, Wang-Chien Lee, Member, and Hong Va Leong," Nearest Surrounder Queries", IEEE Transactions on Knowledge and Data Engineering, VOL. 22, NO. 10, OCTOBER 2010
9. Kao, B. Lee, S. Lee, F. Cheung, D." Clustering Uncertain Data Using Voronoi Diagrams and R-Tree Index,"IEEE Transactions on Knowledge and data engineering, Vol.22 No.9, 2010.
10. D. Zhang, Y.M. Chee, A. Mondal, A.K.H. Tung, and M. Kitsuregawa, "Keyword Search in Spatial Databases: Towards Searching by Document," Proc. Int'l Conf. Data Eng. (ICDE), pp. 688-699, 2009.
11. G. Cong, C.S. Jensen, and D. Wu, "Efficient Retrieval of the Top-k Most Relevant Spatial Web Objects," PVLDB, vol. 2, no. 1, pp. 337- 348, 2009.
12. Elias Frentzos, Kostas Gratsias, and YannisTheodoridis, " On the Effect of Location Uncertainty in Spatial Querying", IEEE Transactions on Knowledge and Data Engineering, VOL. 21, NO. 3, March 2009.
13. Sze Man Yuen, Yufei Tao, Xiaokui Xiao, Jian Pei, " Superseding Nearest Neighbor Search on Uncertain Spatial Databses", IEEE Transaction on Knowledge and Data Engineering,vol. 22, no. 7 July 2010.
14. I.D. Felipe, V. Hristidis, and N. Rishe, "Keyword Search on Spatial Databases," Proc. Int'l Conf. Data Eng. (ICDE), pp. 656-665, 2008.
15. R. Hariharan, B. Hore, C. Li, and S. Mehrotra, "Processing Spatial-Keyword (SK) Queries in Geographic Information Retrieval (GIR) Systems," Proc. Scientific and Statistical Database Management (SSDBM), 2007.